GAIA

G J H McCall, Cirencester, Gloucester, UK

Introduction

The Gaia concept has evolved in the 30 years since it was first introduced by James Lovelock, an independent scientist and inventor. It was initially a rather vague model relating to the climate and diversity of the planet Earth, though living organisms were critical to it. The workings of the model were initially unspecified. The concept, however, was one of a 'superorganism' operating to 'regulate' the planet, especially its surface temperature, yet lacking the 'foresight' possessed by intelligent animals. Lovelock updated his work in 2000, publishing *The Ages of Gaia*. The discussion here is based on this later book, and it is quite unavoidable to echo much of what Lovelock has said, because he is the only and definitive source.

Gaia in the Twenty-First Century

Gaia is essentially about life, because life is seen to combine with inanimate processes on Earth, affecting and even regulating the physical state of the biosphere. Lovelock found it surprisingly difficult to find a good definition of 'life'. Of the definitions found in *Webster* – "the property of plants and animals (ending in death and distinguishing them from organic matter) which makes it possible for them to take in food, get energy from it, grow etc." – and in *Oxford* – "the property which differentiates a living animal or plant or a living portion of organic tissue, from dead or non-living matter; the assemblage of the functional activities by which this property is manifested" – neither is satisfactory, and the second is tautological in the extreme. To the first might be added, before "etc.", the words "and move, in the case of animals". Lovelock added to his definition of life that "living things use energy directly from the Sun and indirectly from food" (*see* **Origin of Life**).

There is no difficulty in accepting that advanced living animal organisms, such as humans, are made up of intricate communities of connecting cells, and, as Lynn Margolis has shown, that cells are derived from micro-organisms that once lived free (*see* **Precambrian:** Prokaryote Fossils; Eukaryote Fossils). Larger entities, such as ecosystems, are also accepted, and space exploration has contributed to this understanding by allowing the entire planet Earth to be viewed from space. The Gaia concept likewise involves envisaging the entire globe as an integrated system, with the atmosphere, the seas, the rivers, and the rocks interacting to modulate the planet's physical state and thus the environment in which life can exist, with the presence of life contributing significantly to the interactions. Gaia is thus not a synonym of 'biosphere' or 'biota': it is a much larger entity. When the Gaia model was originally proposed in the 1970s, it was considered that the atmosphere, oceans, climate, and crust of Earth were regulated to maintain a comfortable state for life to exist, by and for the biota. Temperature, oxidation state, acidity, and certain aspects of the rocks and waters were kept, at any time, constant, maintained in homeostasis, by the organisms at Earth's surface. This concept is now seen to have been incorrect, because both life, which is continually evolving, and the geological environment are in a state of constant change, and the conditions needed to maintain life change very rapidly, with the changing needs of the biota, so homeostasis cannot be maintained for more than very brief periods in Earth history. The Earth is thus seen as being like a helicopter, which is, unlike a fixed-wing plane, never in stable flight. The changing and evolving needs of the biota require that the brief periods of homeostasis are quickly overtaken by new requirements. The concept now is of a superorganism in which the active feedback processes operate automatically, so that solar energy maintains comfortable conditions for life.

Molecular Biology: The Laws of the Universe

Lovelock regarded the emergence of the science of molecular biology – embodied in the information-processing chemicals that underpin the genetic basis of most life on Earth – as having taken life science out of a routine classificatory and descriptive pursuit into a new and exciting study of how all the components in life are related. Equally important are physiology, the study of organisms seen holistically, and thermodynamics, a branch of physics dealing with time and energy, connecting living processes with universal laws. Two fundamental universal laws of physics are that (1) energy is conserved, however much it is dispersed, and (2) energy is always abating. Hot objects cool, but cool objects do not heat up spontaneously; water flows downhill, but not uphill. Once used, energy cannot be recovered. Natural processes always move towards an increase in disorder, which is measured by entropy; entropy expresses the tendency to burn out. Looking at the relationship between life and entropy, Lovelock

referred to Erwin Schrödinger's conclusion that life has the ability to move upstream against the flow of time, apparently paradoxically and contrary to the second universal law. In fact, what is operating is a tightly coupled system to favour survival; energy is taken in (e.g., oxygen from the atmosphere is breathed), converted (e.g., stored body fats and sugars are transformed), and then excreted (e.g., waste products such as carbon dioxide are released back into the atmosphere). If the entropy of excretion is larger than the entropy of the oxygen consumed, life continues, despite the second universal law.

The Superorganism Concept

There is difficulty in envisaging an eruptive planet with a molten core and other complex inorganic processes as a living superorganism. However, the inspiring 'whole-planet' image of Earth as seen from space and the contrast between the environment on Earth and the environments on the moon, Mars, Venus, and Mercury have focused research on considering how the significant planetary differences arose, and in particular on the question of how and why the atmospheres differ. The atmospheres of the Moon, Mars, Venus, and Mercury are a good starting point for comparisons to Earth, because the atmospheres are the least complex and most accessible of the zones of all these planets; indeed, the atmospheric compositions on other planets were known before space exploration commenced (*see* **Solar System:** Mars; Moon; Mercury; Venus).

The Earth has an atmosphere of N and O, with traces of carbon dioxide, methane, and nitrous oxide, not in equilibrium, whereas the atmospheres of Mars and Venus are dominated by carbon dioxide and are in equilibrium. If the atmospheres of Mars and Venus were heated, there would be no reaction with the surface materials, whereas heating Earth's atmosphere would produce reactions leading to a carbon dioxide-dominated atmosphere. Lovelock concluded that the improbable atmosphere of Earth "reveals the invisible hand of life". The atmosphere contains oxygen and methane, which should react to form water vapour and carbon dioxide: that this does not occur, and that constant atmospheric compositions of these gases are maintained, reveal, Lovelock believes, that there is regulation by life (*see* **Atmosphere Evolution**).

Scientists as early as Eduard Suess and Vladimir Vernadsky accepted that there was continuous interaction between soils, rocks, oceans, lakes, rivers, the atmosphere, and life. Much later, Stephen Jay Gould stated that "organisms are not billiard balls, struck in a deterministic fashion and rolling to optimal positions on life's table". Living things influence their own destiny in an interesting and complex, but comprehensible, way. Thus the sum total of the physical state of a planet, with life, is a combination of the inanimate processes and the effects of life itself. JZ Young said that the entity that is maintained intact, and of which we all form part, is the whole of life on the planet. This statement really provided the link between theory and consensus, on the one hand, and Gaia concept, on the other, expressing as it does the view that the entire spectrum of life on the planet has to be considered alongside the geological and inanimate physical processes, if we are to understand how the planet works. This, of course, has led to the present preoccupation in educational circles with 'Earth System Science' (*see* **Earth System Science**). Gaia goes further than Earth System Science, which is purely a holistic educational approach, in requiring a global system that has the capacity to regulate the temperature and composition of Earth's surface, hydrosphere, and atmosphere, keeping it comfortable for living organisms.

Criticism of the Gaia Concept

Criticism of the Gaia concept, once advanced, was by no means slight, and the Gaia model was not taken seriously by scientists, at all, until the early 1970s. Fred Doolittle came out with the belief that "molecular biology could never lead to altruism on a global scale" – altruism by living organisms being apparently inherent to the concept. Richard Dawkins in 1982 supported him: "the selfish interests of living cells could not be expressed at the distance of the planet". It was also remarked that Gaia lacked a firm theoretical basis. Heinrich D Holland considered that biota simply react to change in the state of Earth's near-surface environment and processes, geologically produced, and those that adapt better survive: the rest do not. Many scientists saw Gaia as a teleological concept, requiring foresight and planning by organisms, something that the model surely never represented. However, a major step was taken at the Chapman conference of the American Geophysical Union in 1988, when numerous papers on Gaia were presented: the question of the scientific testability of the Gaia hypothesis was raised.

Holland's statement was really an oversimplification, because the environmental constraints to which an organism adapts can in no way be entirely inorganic in origin – geological processes are a combination of the inorganic and the organic. Lovelock stated this when he objected that "life cannot have adapted solely to an inert world determined by the dead hand of chemistry and physics". The two main objections to Gaia were, first, the teleological one

(forecast or clairvoyance seemed to be needed for Gaia to be true) and, second, the fact that ecological regulation by life could only be partial. What must be operating is a combination of regulation by the living and the inorganic, even though in the long term, a wholly inorganic interference, such as an asteroidal impact, would be reacted to by the biota, much of it by dying off. However, some of the biota may survive, regenerating and expanding to again affect the physical environment of the life forms continuing to exist at or near the surface.

Other objections to the Gaia concept have been a creationist argument, based on Schrödinger's conclusions on 'life', that some organisms do not need the sun, and also a suggestion that Lovelock's image of Earth as a 'spaceship' ignores the fact that 40 000 tonnes year^{-1} of extraterrestrial dust enters Earth's atmosphere and is deposited on Earth's surface.

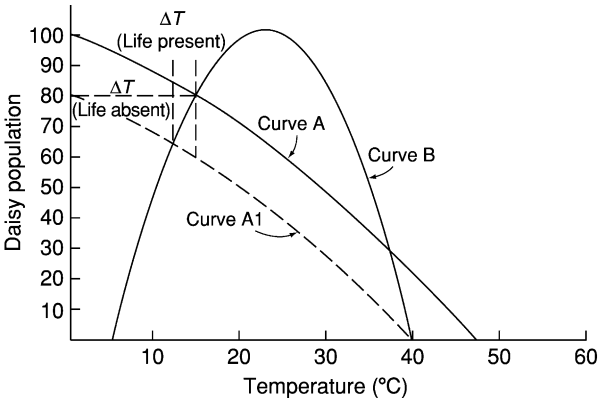


Figure 1 Regulation by white daisies. The helmet-shaped curve (B) depicts the response of daisies to temperature. Curves A and A1 depict the responses of planetary temperature to the area covered by daisies, curve A1 being for a lower heat input by the planet's star. In the absence of daisies, the change in planetary temperature (ΔT) would be nearly 15°C, whereas in their presence ΔT is only about 3°C. Reproduced with permission from Lovelock J (2000) *The Ages of GAIA: A Biography of Our Living Earth*. Oxford: Oxford University Press.

Lovelock's Hypothesis

The 'Daisyworld' Model

Lovelock saw that he needed a simple model to illustrate his point. The effect of snow cover on the ground, a purely inorganic change, had already been modelled to show how snow cover changes the albedo and thus the cooling in the atmosphere. Lovelock developed a simple model, the 'Daisyworld' model, illustrating the way life could have a similar effect.

The Daisyworld parable of 1982 proposes a planet like Earth in size, mass, and orbital distance from a star that is like Earth's sun in mass and density; like Earth's Sun, the Daisyworld Sun increases its output as it ages (the nature of the H/He reaction means that our Sun was 30% cooler at its beginning, but will eventually heat up so as to consume Mercury, Venus, and possibly Earth). The Daisyworld planet has more land than sea, compared to Earth, is well watered, and plants can grow anywhere on the land surface if the climate is right. The sole plant is a daisy, which may be dark, neutral, or light in colour shade. A single parameter, variable temperature, controls whether the daisy can grow – 5°C is the growth threshold, 20°C is the optimum, and 40°C is the upper limit. The mean temperature of the planet is a simple balance between the heat received from the sun and the heat lost to the cold depths of space in the form of long-wave radiation. The complication on the real Earth of reflection upwards and blanketing downwards of heat by clouds is avoided, by having all the rainfall on the Daisyworld planet occur at night and having no clouds present in daytime. There is just enough carbon dioxide in the atmosphere for the plants to grow. The mean temperature will be determined by the average shade of colour of the daisy coverage. If the average is the dark colour shade, the albedo is low, more heat is absorbed, and the surface is warmed; if the average is the light shade, 70–80% of heat is reflected and the surface is colder. In the scale 0–1, dark daisies will have an albedo of, say, 0.2; light daises will have an albedo of 0.7 and the bare ground, 0.4. The initial weak Sun will increase until 5°C is attained and that will favour the dark daisies, because they have greater absorption of sunlight and will warm up the temperature just beyond 5°C. The light daisies will be disadvantaged and will fade and die. Next season, there will be more seeds of the dark strain and it will become warmer and so on, until most of the planet is colonized by the dark strain. However, when the upper limit temperature is approached, the light strain, because of their high albedo, will keep the temperature in check and so will take over from the dark strain. The growth of the star's heat flux will eventually be so great that nothing can keep the temperature below 40°C and all the daisies will die off. Flower power will no longer be enough. The planet will be barren and there will be no way back. Figure 1 reproduces Lovelock's most important Daisyworld diagram. The Daisyworld parable was devised to counter the 'teleological' criticism.

The Four Components of Gaia

Lovelock described Gaia as an automatic, nonpurposeful, goal-seeking system, and considered it to be made up of four components:

1. Living organisms that grow freely, exploiting any environmental opportunities that open up.
2. Organisms that are subject to the rules of Darwinian natural selection: the species that leave the most progeny survive.
3. Organisms that affect their physical and chemical environment. Thus, animals change the atmosphere by breathing, taking in oxygen and breathing out carbon dioxide. Plants and algae do the reverse. In numerous ways, all forms of life incessantly modify the physical and chemical environment.
4. The existence of constraints or bounds that establish the limits of life. It can be too hot or too cold; there is a comfortable warmth in between, the preferred state. It can be too acid or alkaline; neutrality is preferred. For almost all chemicals, there is a range tolerated or needed by life. For many elements, such as iodine, selenium, or iron, too much is a poison and too little causes starvation. Pure, uncontaminated water will support little, but neither will the saturated brine of the Dead Sea.

Definitions and Criticisms

Lovelock advanced a set of definitions:

1. Life is a planetary phenomenon with a cosmological lifespan. On this scale, it is nearly immortal and has no need to reproduce.
2. There can be no partial occupation of the planet by living organisms. Such a condition would be as impermanent as half an animal. Ineluctable physical and chemical forces would soon render (the planet) uninhabitable. The presence of a sufficient number of animals on the planet is needed for the regulation of the environment.
3. Our interpretation of Darwin's great vision is altered. Gaia draws attention to the fallacy of the concept of adaptation. It is no longer sufficient to say that "organisms better adapted than others are likely to leave offspring". It is necessary to add that the growth of the organism affects its physical and chemical environment; the evolution of the species and, therefore, the evolution of the rocks are tightly coupled in a single, indivisible process.
4. Theoretical ecology is enlarged. By taking the species and the physical environment together as a single system, we can, for the first time, build ecological models that are mathematically stable, and yet include large numbers of competing species. In these models, increased diversity among the species leads to better regulation.

There are criticisms of Lovelock's definitions. Of definition 1, it is perhaps better to admit that there remain possibilities of abrupt extinction of life on Earth through physical disasters. For example, the entire planetary biota might have been extinguished in the latest Proterozoic; if glaciation, a complex phenomenon attributed to Milankovitch cycles in the Solar System, had covered the globe in the 'snowball Earth' episode, it might well have extinguished all life. Definition 2 begs the question, "what about the long early period in the Archaean, when only very primitive forms of unicellular life apparently existed?" The planet in fact got on very well through a vast period of time when it was only partially inhabited by life, and that being very primitive, compared with now. Regarding definition 3, surely to the word 'rocks' the words 'hydrosphere and biosphere' should be added. Finally, in definition 4, the statement that "increased diversity leads to better regulation" is questionable. The evidence for this does not seem to be to be given by Lovelock: the word 'regulation', which is fundamental to Gaia concept, seems to be a problem. The concept appears to be tantamount to accidental regulation: in the real world, as opposed to the Daisyworld, the reality would seem to be that the biota make a major contribution to the physical and chemical environment of the biosphere, and that, throughout Earth history since life appeared on the planet, the bounds of livability of the planetary biota as whole have never been exceeded, despite several mass extinctions – yet total extinction could happen at any time. Also, like the inorganic physical and chemical (geological) changes, changes induced by the biota can be benign or adverse. The biota can react to change to counter adverse changes; initially benign changes can lead to adverse changes in the long run (e.g., overpopulation). There is no law of the sum total planetary biota adjusting to regulate the environment, keeping within bounds, but so far (luck?), the bounds have not been overridden in the case of the overall biota, not at any time in geological history, since life first appeared.

Gaia and the Geological Record

Lovelock, in his second book, considered Gaia in terms of the geological record, discussing the Archaean, the Middle Ages, and the contemporary environment. Before concluding, it seems apposite to refer to a recent contribution to the discussion by Euan Nisbet.

Nisbet's Essay

Euan Nisbet, in a wide-ranging Fermor Lecture to the Geological Society in 2002, explored the question whether the presence of life or inorganic processes

had constituted the dominant factor in shaping the physical development and in controlling the physical conditions at or near Earth's surface since life appeared on the planet more than 3500 million years ago. He found that life had a dominant role in controlling the condition of the atmosphere through this immensely long period, and that, in turn, the surface temperature had a very significant control on the tectonic evolution, especially plate tectonics. Even so, Nisbet could not rule as to which processes, biological or inorganic, had been the dominant factor – the two kinds of processes had operated together to control the physical conditions. That life had exerted a significant effect on the ongoing physical evolution could not be argued against.

Nisbet based his conclusions on inorganic and organic models, and in them not all the assumptions made are necessarily correct. For instance, Nisbet tended to extend plate tectonics back through the Archaean, whereas there are strong arguments against this (e.g., see the publications by Hamilton in 1998, by Bleeker in 2002, and by McCall in 2003). Considering the real world, rather than models, Nisbet could not decide which had been in the driving seat, and he believed that the question was possibly not quantifiable. The atmosphere was, however, largely the product of the existence of life at or near Earth's surface.

Beyond Gaia

Gaia, at the least, is a brilliant concept, and Lovelock has provided a stimulating basis for looking at the entire globe and for integration of the effects of the progressive development of life and purely inanimate geological development through time. This concept also encompasses the maintenance throughout time of conditions suitable for life to continue to exist throughout 3500 million or more years. However, at this minimal assessment, it is no more than Earth System Science under another name, but emphasizing the special contribution of life to the changing state of the atmosphere, hydrosphere, and geological processes, which in sum determine the physical conditions under which life of some sort can continue to 'operate'. The Gaia concept, however, is claimed to be more than this, and the critical word is 'regulate'. Lovelock never intended Gaia to have a teleological significance, but even so, to regulate something would seem to imply an element of design. It may be that when an extreme condition occurs, such as an extreme greenhouse condition (Cretaceous) or a 'snowball Earth' approximation (end Proterozoic), the reactions of the living plant and animal populations that survive may be beneficial, but equally they could actually be adverse and eliminate even more of the global biota. One cannot escape the fact that the physical state of the near-surface zones of the planet has, for more than 3500 million years, been sufficiently benign for some life to continue to exist; throughout the development of life, from protozoans to destructive humans, the physical state of Earth's surface has never led to extinction of the sum total of life on the planet. Yet the history has surely been one of reactions by life to changes, including extremely critical situations, both inorganically and organically triggered; changes have never reached the point of driving life over the edge to extinction, but even so, this is not regulation, because there is no rule involved. The irretrievable end-point situation has just not happened. In fairness to Lovelock, in his 1991 book he did include a chapter entitled "The People Plague", in which he argued that humans, having no predators, had in effect become a plague on the planet, and could well take the life on the planet to that irretrievable end-point situation. Of course, one could argue that humans do have predators, even if minute ones – viruses and microbes that could blot them out forever and leave behind an Earth smiling with other life forms, until the Sun fulfils its destiny and incinerates the planet.

The influence of the existence of life on the physics and chemistry near the surface of the planet, and the feedback to life from that – the essence of Gaia – is well understood, but has obviously become more complex as the spectrum of life on Earth has become more complex, starting with the primitive prokaryotes in the Early Archaean. For example, a plant population limited to algae must have left bare rock surfaces in the Archaean, whereas once higher plants appeared, there must have been a much more complex reaction of the land surface to the Sun's radiation. In fact, these relationships, though complex and possibly unquantifiable at the present state of the planet, are less obscure than are other processes. For example, how does an animal obtain information that another animal exists and has a character or activity to which the first animal can adapt, to obtain benefit? How does the information get into the genetic process? The present author, in an unpublished work, *The Vendian (Ediacaran) in the Geological Record*, has called this mysterious process 'cognizance'. An example of this is the case of eoforaminifera found in Uruguay, in latest Proterozoic rocks. These organisms agglutinated fine mineral particles on their surface, and it has been glibly said that they did this to make themselves less palatable to predators. But how did they know that their mates were being eaten? It would seem that this aspect of genetics is what scientists should now be concentrating on, rather than 'regulation' by the sum total biota.

See Also

Atmosphere Evolution. **Biodiversity**. **Earth Structure and Origins**. **Earth System Science**. **Evolution**. **Origin of Life**. **Palaeoclimates**. **Precambrian:** Eukaryote Fossils; Prokaryote Fossils. **Solar System:** Mars; Moon; Mercury; Venus. **Trace Fossils**.

Further Reading

Bleeker W (2002) Archaean tectonics – a review. In: Fowler CMR, Ebinger CJ, and Hawkesworth CJ (eds.) *The Early Earth: Physical, Chemical and Biological Development. Special Publications 199,* pp. 151–181. London: Geological Society.

Gaucher C and Sprechmann P (1999) Upper Vendian skeletal fauna of the Arroyo del Soldado Group, Uruguay. *Beringeria* 23: 55–91.

Hamilton WB (1998) Archaean magmatism and deformation were not products of plate tectonics. *Precambrian Research* 91: 131–175.

Lovelock J (1978) *GAIA: A New Look at Life on Earth*. Oxford: Oxford University Press.

Lovelock J (1991) *GAIA: The Practical Science of Planetary Medicine*. London: Gaia Books.

Lovelock J (2000) *The Ages of GAIA: A Biography of Our Living Earth*. Oxford: Oxford University Press.

McCall GJH (2003) A critique of the analogy between Archaean and Phanerozoic tectonics based on regional mapping of the Mesozoic–Cenozoic plate convergent zone on the Makran, Iran. *Precambrian Research* 127(1-3): 5–17.

Myers N (1984) *Gaia: An Atlas of Planet Management*. New York: Doubleday.

Nisbet E (2002) The influence of life on the face of the Earth. In: Fowler CMR, Ebinger CJ, and Hawkesworth CJ (eds.) *The Early Earth: Physical, Chemical and Biological Development. Special Publications 199*, pp. 275–307. London: Geological Society.

GEMSTONES

C Oldershaw, St. Albans, UK

Introduction

There are almost 4000 minerals known, of which only about 50 are commonly used as gemstones. Those that form crystals of sufficient size and quality to be cut and fashioned as gems are referred to as 'gem quality' or 'cuttable' pieces; other minerals or rocks with particularly attractive features (colour, texture, or pattern) may be called 'decorative' pieces. Crystals are usually faceted (cut and polished) to give a gemstone with a number of flat faces, while decorative stones are mainly tumbled or polished to produce pieces for personal adornment or *objets d'art*.

Gemstones are formed in each of the three main rock types: igneous, sedimentary, and metamorphic. Mining methods depend on the type of gemstone and its optical and physical qualities and on whether it is being mined from the rock in which it was formed or retrieved from secondary (placer) deposits produced by weathering and erosion.

A study of modern-day gemstone mining and retrieval covers every mining method, from the traditional searches in streams and rivers using a pan or sieve to the ultra-high technology and research models used in diamond mines deep underground. Ultimately, any source of gemstones will be mined and exploited only if is financially viable.

Quantifying Gemstone Mining

While mineral exploration is generally well documented and statistics are available, providing figures for production for most minerals on a country-by-country basis, this is generally not the case with gemstones (except perhaps for diamonds), which have always been difficult to quantify. Much gemstone mining is carried out in remote places, with secrecy and security to protect the interests (and sometimes the lives) of the owners (Table 1). In some areas, families or groups of villagers purchase a licence to mine; in other areas there is little regulation, and mining may be controlled by local leaders or warlords, with all the problems associated with regions of conflict and political unrest.

Over large regions it is difficult to assess the production figures for particular gemstones, although general trends in availability and value (which could be as much a product of fashion as of supply) give some indication of production figures. On a smaller scale, a specialist gemstone buyer (such as a ruby or sapphire buyer) will, by building relationships with miners, mine owners, and local traders, gain expertise and some knowledge of the volume and quality of gemstones being mined in that region. As new localities are identified and new gemstones reach the markets, previously known localities or gemstones may lose their appeal or a region may become unproductive – so the market can be quite changeable.

Table 1 The main localities of some of the best-known gemstones (diamond, ruby, sapphire, emerald, aquamarine, chrysoberyl, topaz, tourmaline, peridot, garnet, pearl, opal, spinel, zircon, turquoise, nephrite jade, and jadeite jade)

USA
- Aquamarine
- Emerald
- Jadeite jade
- Nephrite jade
- Peridot
- Ruby
- Sapphire
- Topaz
- Tourmaline
- Turquoise

Canada
- Diamond
- Nephrite jade

Mexico
- Opal
- Topaz
- Tourmaline
- Turquoise

Honduras
- Opal

Colombia
- Emerald

Brazil
- Chrysoberyl
- Diamond
- Emerald
- Opal
- Topaz
- Tourmaline

Germany
- Topaz

Italy
- Tourmaline

Former Czechoslovakia
- Garnet
- Opal

Egypt
- Emerald
- Peridot
- Turquoise

Nigeria
- Aquamarine
- Sapphire
- Spinel
- Topaz

Zaire, Angola, and Namibia
- Diamond

Zambia
- Chrysoberyl
- Emerald

Botswana
- Diamond

South Africa
- Diamond
- Emerald
- Peridot
- Ruby
- Tourmaline

East Africa
- Aquamarine
- Diamond
- Emerald
- Ruby
- Sapphire
- Tanzanite
- Tourmaline

Madagascar
- Aquamarine
- Chrysoberyl
- Topaz
- Tourmaline

Russia
- Demantoid garnet
- Diamond
- Emerald
- Nephrite jade
- Topaz
- Tourmaline

Afghanistan
- Aquamarine
- Ruby
- Spinel

Pakistan
- Aquamarine
- Emerald
- Grossular garnet
- Ruby
- Spinel
- Topaz

India
- Almandine garnet
- Aquamarine
- Chrysoberyl
- Diamond
- Emerald
- Ruby
- Sapphire

Sri Lanka
- Chrysoberyl
- Garnet
- Ruby
- Sapphire
- Spinel
- Topaz
- Tourmaline
- Zircon

Myanmar (*formerly Burma*)
- Chrysoberyl
- Jadeite jade
- Peridot
- Ruby
- Sapphire
- Spinel
- Topaz
- Tourmaline
- Zircon

Thailand
- Almandine garnet
- Ruby
- Sapphire
- Zircon

China
- Aquamarine
- Diamond
- Nephrite jade
- Peridot
- Ruby
- Sapphire
- Turquoise

Japan and Taiwan
- Jadeite jade
- Topaz

Australia
- Diamond
- Emerald
- Nephrite jade
- Opal
- Sapphire

New Zealand
- Nephrite jade

Guatemala
- Jadeite jade

Iran
- Turquoise

Gem-testing laboratories, museums, and specialist gemstone collectors are often the first to hear of a new find. Some gemstones will be mined in sufficiently large volumes to reach the marketplace worldwide, for example tanzanite (the blue variety of the mineral zoisite), which was discovered in 1967 in Tanzania and is now available through many retail outlets. Others may be mined out fairly quickly, with most of the gemstones going to a few specialist buyers.

Some particularly rare finds are found only in museums or private collections. Other gems are cut for collectors but, because of their rarity or physical properties (for example they may be too soft for everyday wear), do not reach the retail market. New localities and new gemstone finds of interest, such as a particularly fine-coloured gemstone or a mine that is producing particularly clean large specimens of a gemstone, are usually reported in gemmological journals such as the *Journal of Gemmology*, *Gems and Gemmology*, and *Australian Gemmologist*.

Alluvial and Eluvial Deposits

Gems formed in igneous or metamorphic rock may be subject to weathering and erosion, breaking down the rock and releasing the gemstones (*see* **Weathering**). The gems may remain where they are or be transported by ice, wind, or water and deposited elsewhere. Eluvial gem deposits are formed as a result of the weathering of rock that has remained in the same place. Alluvial deposits are deposited by flowing water (*see* **Sedimentary Environments:** Alluvial Fans, Alluvial Sediments and Settings).

Gemstones are generally harder and heavier than the surrounding minerals. They are not carried as far

by flowing water and tend to sink faster, becoming concentrated in pockets or areas along riverbanks or within gravels (gem gravels). Therefore, alluvial gemstone deposits contain gemstones that are sufficiently hard and durable to withstand the conditions without breaking, rather than those that are heavily included or prone to fracture or cleavage.

Because nature has already partially eliminated the weaker specimens, the percentage of gem-quality gemstones in gem gravels is usually high, with the result that more gemstones are retrieved from gem gravels than from any other type of deposit. Alluvial deposits, such as the gem gravels of Sri Lanka and Myanmar (formerly Burma), contain a wide range of gemstones including ruby, sapphire, spinel, chrysoberyl, topaz, tourmaline, and garnet.

Because gemstones associated with gem gravels are often found together, the discovery of one type of gemstone from a gem association (sometimes called a tracer gem) can be used by exploration teams and prospectors to 'trace' potential gemstone mining areas. Another technique is to map the courses of ancient river beds or present-day rivers and streams and then follow tracer gems downstream to find areas where the gemstones are present in large enough concentrations to be retrieved.

The oldest and most traditional mining methods are still practiced in areas where gemstones are near the surface and relatively easy to find and retrieve and where labour is cheap. In Indonesia, Malaysia, Sri Lanka, and India, for example, local people search rivers and streams and excavate the gravels and sediments from now-buried riverbeds.

Panning for gems works on the principle that the gems are generally heavier than the surrounding mud, pebbles, or rock fragments. As the pan of water and sediment is 'jiggled', the gemstones settle towards the bottom of the pan while the lighter constituents and water are washed over the pan's edge (Figure 1). The heavier concentrate may be sieved to separate larger gems or it may be spread out on tables or cloths to be hand-sorted.

Alluvial deposits are generally mined by similar methods wherever they are found in the world, usually by digging shafts or pits or by collecting the alluvium from rivers and streams using sieves and pans. Dams may be constructed to manage the water flow, and water may be diverted to wash the gem gravels (Figures 2 and 3). Final sorting is done by hand.

Often the mining and retrieval of gems in these traditional ways is a family project, varying only slightly from country to country. The men and boys may work in the rivers or dig the pits, while women and girls sieve or sort the gems (Figure 4).

Figure 1 A lady river panning, washing gravel in search of chrysoberyl cat's-eyes (Kerala, India). Printed with permission from Alan Jobbins.

Ruby and Sapphire Deposits

The rubies and sapphires of Thailand and Cambodia are found in alluvial and eluvial gravels, derived from highly alkaline basalt (Figure 5). The main mining area is near the border between Thailand and Cambodia, around Chantaburi.

Sri Lanka is a source of alluvial sapphires, which are derived from pegmatite and gneiss and are found in a wide range of colours, including the pinkish-orange padparadscha. Padparadscha is particularly rare and is named after the sinhalese for lotus flower, whose colour it resembles.

Rubies and sapphires are members of the corundum family, and different trace elements within them give rise to the colour range. Rubies are red (coloured by chromium and possibly vanadium); blue sapphire is coloured by iron and titanium. Other sapphires, such as green sapphire, pink sapphire, and mauve sapphire, result from the addition of various trace elements during the gem's formation. Colourless (white) sapphire has the simple chemical composition aluminium oxide (Al_2O_3).

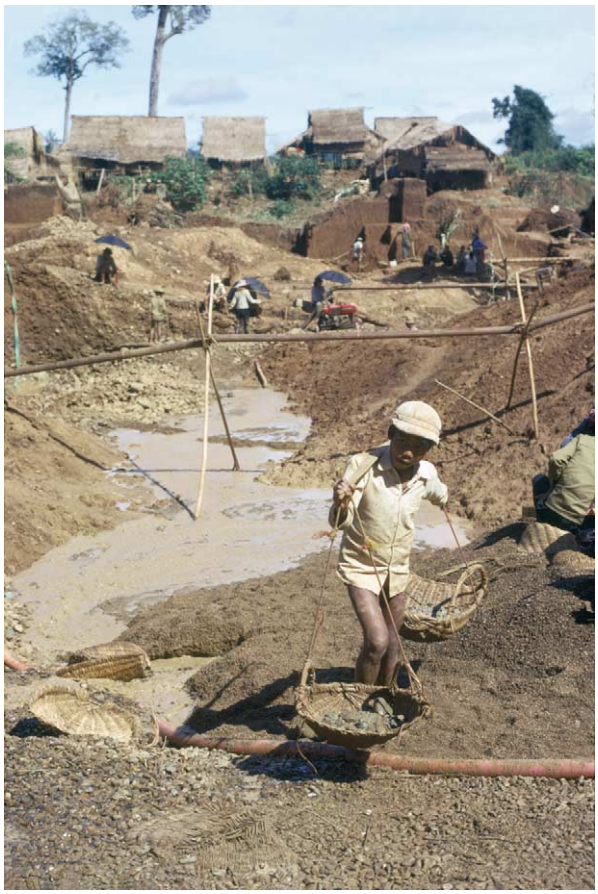


Figure 2 A young worker carries the gemstone concentrate called 'pay dirt' or 'rai' in 'Burmese-style' baskets (Pailin, Cambodia). © Alan Jobbins.

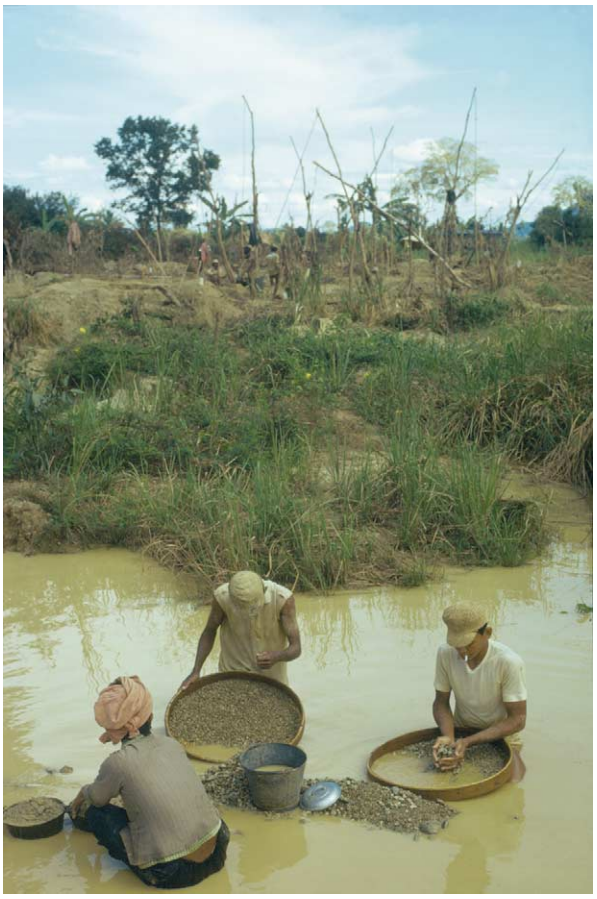


Figure 3 Washing gemstone concentrate in a small pond (Pailin, Cambodia). © Alan Jobbins.

The colour does, to some extent, affect the mining, as the more popular and valuable colours are more sought after. It should be noted, however, that some gemstones, including corundum, may have their colour changed or enhanced with heat treatment, irradiation (electrons, X-rays, etc.), or other techniques.

Kashmir sapphires, which are famous for their cornflower colour, are mined in Sanskar, an inhospitable mountainous area often covered by snow, northwest of the Himalayas. The sapphires are found in feldspar pegmatites or in gravels derived from the pegmatites. There has been intermittent mining since the 1920s; however, little is known about present mining because of the political situation.

Mogok Rubies

The most famous rubies are the 'pigeon blood'-coloured rubies from the Mogok Stone Tract area of northern Myanmar, about 110 km north of Mandalay. Documentary evidence shows that mining has been carried out since the sixteenth century, but there have been finds of prehistoric tools that suggest that mining has been carried out for many thousands of years.

Traditional methods involve digging a narrow pit or vertical shaft, just wide enough for the miner to be lowered down to dig out the layer of weathered metamorphosed crystalline limestone (marble) that contains the rubies. The miner removes rock (often 2–3 m below the surface) and places it in a basket to be hauled to the surface and sorted.

Following the establishment of the Burma Ruby Mine Company in Victorian times, some mining techniques were modernized. Water pumps were introduced to speed the washing process, and intricate channels were constructed from wooden planks, along which water was sluiced to separate the rubies from rock fragments. As a result of the collapse of the market following the introduction of synthetic rubies and sapphires in the early twentieth century, the British Company failed and traditional methods were reintroduced. In the 1990s new mines were established in the Mong Hsu area, about 250 km east of Mandalay.

Figure 4 Girls sorting gemstone concentrate (Pailin, Cambodia). © Alan Jobbins.

Figure 5 Eluvial workings in weathered (decomposed) basalt above basalt lava (Pailin, Cambodia). © Alan Jobbins.

Gemstones in Igneous Rocks

Where rock is harder, picks and drills may be sufficient to prise loose the gemstones from the parent rock (the host); otherwise the rock has to be mined, crushed, washed, and sorted to retrieve the gemstones. Diamonds are mined on a larger scale and with more highly mechanized methods than any other gemstone.

Extrusive Igneous Rocks

Alluvial gem deposits derived from basaltic lavas have been mentioned above. The basaltic lavas (extrusive igneous rock, where magma has been erupted from volcanoes as lava) may also be mined. Peridot is commonly formed in basaltic lava. The peridot generally forms as small crystals or within vesicles (small bubbles) and voids in the lava. Other gems that crystallize from lavas as they cool include zircon, ruby, and sapphire.

Intrusive Igneous Rocks

Some of the largest crystals form where the magma is not erupted but rises sufficiently in the Earth's crust to cool slowly. The slower the cooling, the larger the crystals that can form. Igneous rocks formed in this way are said to be intrusive.

Gems may form in dykes (small igneous intrusions) intruded into surrounding rocks, which may crop out at the surface as a result of weathering. Mining

of weathered rock on the surface and the retrieval of gems from these eluvial deposits may lead to mining of the gem-bearing rock beneath. Different mining methods will be used as conditions and rock types change; for example, mining of the Montana sapphires of North America began as surface mining and dyke digging. The harder rock of the dykes – the intrusive igneous rocks that held the rubies – was mined in preference to the non-gem bearing rock on either side. As the gem-bearing rocks were followed deeper, mining changed from surface and dyke digging to underground mining.

Granite is a common intrusive igneous rock. Gemstones found in granite include quartz, feldspar, and tourmaline, but often the crystals are not of sufficient quality to be cuttable or may be too difficult to remove from the rock without damage. In this instance, the rock may be mined for use with the gemstones still *in situ* and used as a decorative stone, for example as a work-surface in a kitchen, a stone floor in a shopping mall, or a façade on a building (*see* **Building Stone**).

Pegmatites Pegmatites are intrusive igneous rocks. They produce a greater range of gemstones than any other rock type and have also been the source of some of the largest gemstones ever mined. The large crystals form as the water-rich portion of a granite-like molten rock is put under increased pressure as it is squeezed into fractures in the surrounding rock. As the molten rock begins to solidify, the elements that it contains begin to crystallize. The largest gemstones and some quite rare gemstone varieties are found in gem pockets at the centre of the pegmatite, where they have formed from the hot concentrated mineral-rich fluid that was the last to crystallize (Figure 6).

Pegmatites occur throughout the world, but the largest gem-producing pegmatites are those of Minas Gerais, Brazil. Other pegmatite areas include the Pala area of San Diego County in California, the Nuristan area of Afghanistan, the Sverdlovsk region of the Urals, and the Altai region of north-western China. Gems formed in pegmatites include topaz, tourmaline, kunzite, and members of the beryl family, such as the blue aquamarine and the pink morganite.

Hydrothermal gemstones Gems are also formed from hydrothermal fluids (from 'hydro' meaning water and 'thermal' meaning hot), which escape from magmas and may contain rare elements such as fluorine and beryllium. As the hydrothermal fluids move away from the magma along fractures and fissures in the surrounding rock, they solidify and form mineral veins (*see* **Mining Geology: Hydrothermal Ores**). Close to the surface, hydrothermal veins may also include elements carried by groundwater and other near-surface waters. Amethyst, topaz, benitoite, and emerald are examples of gemstones found in hydrothermal veins.

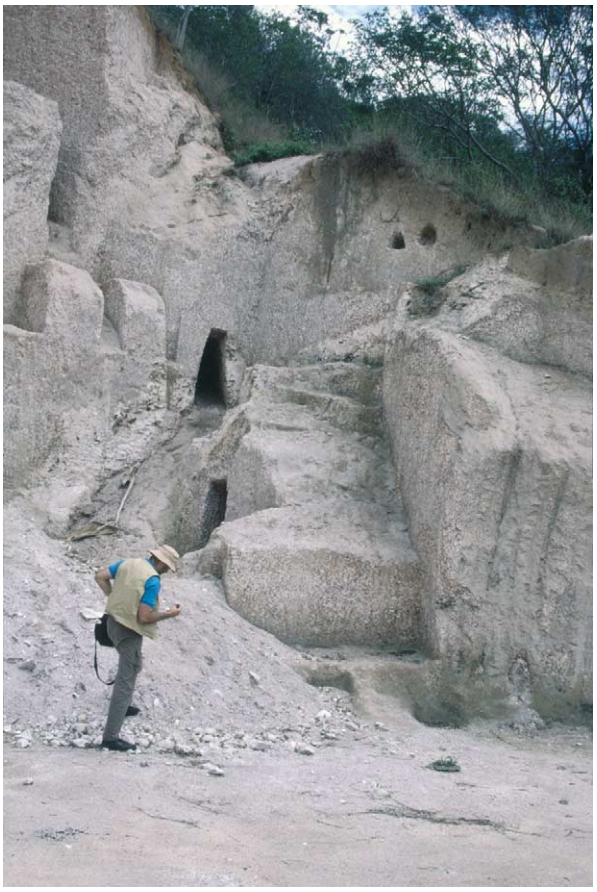


Figure 6 Workings in granite with cross-cutting pegmatite, which is a source of aquamarine and chrysoberyl (River Marambio Baia, Minas Gerais, Brazil). © Alan Jobbins.

In April 2000, while blasting for a new tunnel in the Naica silver and lead mine of central Chihuahua, Mexico, miners found two underground caverns containing crystals of selenite, some of which were more than 15 m in length. Hydrothermal fluids deposited gold, silver, lead, and zinc near the surface, but groundwater in these caves formed the huge crystals. Although the caverns were kept secret, for fear of vandalism or theft, the conditions within were sufficiently hostile to keep most people out. With temperatures of more than 50°C and humidity above 85%, visitors became uncomfortable and disorientated in less than 10 min. Mining of such large gemstones has generally been carried out only to supply museums and to satisfy collectors of spectacular specimens.

Gemstones in Metamorphic Rocks

All rocks, including sedimentary and igneous rocks, may be altered by pressure and/or temperature to form metamorphic rocks. Metamorphic rocks include marble, gneiss, and schist. The gemstones they contain will depend on the composition of the rock before metamorphism and on the conditions of pressure and temperature to which they have been exposed.

Ruby and Sapphire

The area affected by metamorphism may cover only a few metres (usually as a result of local faulting or folding) or hundreds or thousands of kilometres (owing to larger regional tectonic events such as mountain building). For example, the rubies of Myanmar are associated with the mountain-building episode that formed the Himalayas more than 65 Ma ago, when the continent of India moved north, colliding with the mainland and pushing up the highest mountains in the world. The Myanmar rubies and sapphires (Mogok), Pakistan rubies (Hunza Mountains), Tanzania rubies (Morogoro), and Vietnam rubies (Luc Yen deposit and Quy Chau deposits) occur in marble formed by the metamorphism of limestone.

Kenya rubies (Mangari area of the Tsavo West National Park) are found in micaceous metamorphic rocks; the main mines are the Penny Lane and John Saul mines. The Penny Lane ruby mine is worked as an open cast pit and, along with the John Saul mine, is a modern and highly advanced plant. Large construction vehicles follow tracks down into the pit to collect the rock fragments that have been blasted from the rock face. These are then taken to the washing and sorting plants, where the rubies are retrieved.

Mica schist, formed from metamorphosed shale and slate, may contain gems such as garnet, andalusite, kyanite, and iolite (cordierite). The best jadeite jade is from Myanmar. Formed by high-pressure metamorphism, it is found as river boulders or mined from dykes in serpentinized rocks.

Emerald

In Colombia, the Muzo emeralds are found in thin layers of white limestone in soft black carbonaceous shales. Explosives are used and enormous trucks transport the soft shales to washing plants, where the harder limestone is separated from the shale and the emeralds are retrieved. Landscapes are altered as hillsides are removed. In addition to the large-scale removal of the rocks, groups of people also work on a far smaller scale. As the remaining shales are weathered, the gems are washed out and transported downhill into the valleys below. Local people search the riverbeds and sediments for emeralds, often using spades or just their hands.

The Muzo and Chivor regions are the two main Colombian emerald-mining regions. Fine-coloured emeralds are produced from both regions, though those from the Muzo tend to have a more yellowish-green colour than the bluish-green Chivor emeralds. Private companies with government supervision run the mines at Muzo. The main Chivor mine is privately run. The government issues short-term permits, usually lasting only 5 years, then issues new ones.

The Colombian emeralds have a long mining history. They were used for decorative and ceremonial purposes before their discovery by the Spanish conquistadors, who introduced them to Europe. After this they became the source of most emerald jewellery until the end of the nineteenth century.

During Roman times emeralds were mined from sites in the Habachtal area of the Austrian Alps, about 3000 m above sea-level. The oldest known source of emeralds is the famous Cleopatra Mine of Ancient Egypt, situated on the Red Sea, where mining is thought to have been carried out as long ago as 4000 years BP. Mining was a labour-intensive practice, and it is thought that thousands of miners worked below ground in the labyrinth of tunnels.

More recently emeralds have been discovered in other countries, including Brazil, Australia, Pakistan (Swat Valley), Afghanistan (Panjshir Valley), Russia (near Sverdlovsk, Ural Mountains), South Africa, Zimbabwe (Sandawana Valley), Tanzania, and Zambia. They occur in hydrothermal veins, in schists (including mica schist and weathered talc schist), and in metamorphic limestones (including dolomitic marble). In some areas they are associated with pegmatite intrusions.

The quality and colour of the emeralds varies between countries and from mine to mine, but many compete well with the best Colombian emeralds, producing high-quality gemstones. Production has varied, often depending upon the political situation in a country rather than the potential of the mine.

Specimens may have characteristic inclusions that can be used to confirm their country of origin and even sometimes the mine from which they came (*see* **Fluid Inclusions**). For example, emeralds from Chivor can contain inclusions of small pyrite crystals, while those from Muzo may contain calcite crystals. Both the Chivor and the Muzo emeralds have distinctive three-phase inclusions, comprising a gas, a liquid, and a solid. Sandawana emerald is famous for its inclusions of the mineral tremolite, which look like a series of fine bars; they may also contain two-phase inclusions (containing a liquid and a gas bubble). Some Russian emeralds contain inclusions

of the mineral actinolite and shiny yellowish-brown mica flakes. Indian emeralds have a characteristic two-phase inclusion that resembles a comma with a jagged outline.

Other Sedimentary Gemstone Deposits

In addition to alluvial and eluvial gemstone deposits, gemstones may also form in veins or cavities within rocks or as crusts on their surfaces. Following transportation, they react with other elements to produce new minerals or may be deposited directly from solutions. New minerals may also form deposits when water cools, evaporates, or changes level, leaving a mineral residue. Examples of gemstones formed in rocks as a result of the evaporation, cooling, or transportation of mineral-rich fluids include turquoise, malachite, rhodochrosite, agate, and amethyst.

Turquoise usually forms as a thin layer or vein within the rock. The rock is mined and the turquoise removed for polishing or fashioning, often as beads or small pieces for use as inlays. Malachite can occur as large botryoidal (rounded) masses or as thin layers and is often associated with azurite. Rhodochrosite can occur as large outcrops, which are mined using heavy machinery and cut and polished as smaller decorative pieces.

Agates are fairly common and are usually found as harder geodes within rock faces of igneous origin. Underground mining, following rock layers containing the geodes, results in a labyrinth of tunnels and mineshafts. The agate mines in Idar-Oberstein (Germany), which have been mined for agates since Roman times, are now popular with tourists and groups of visiting gemmologists.

Opal is formed in sedimentary rock where organic remains have decomposed to leave minerals within the rock or cavities into which minerals have been transported by fluids. Large specimens, such as fossilized dinosaur bones, may be replaced by opal, and small irregular voids within a rock can be filled.

Australia is the best-known and most important source of opal. The main mining areas are Lightning Ridge, Coober Pedy, White Cliffs, Quilpie, and Andamooka. Tourists can obtain licenses to mine, checking the ground and riverbeds for small rock fragments containing opal and panning rivers during their stay. Mining companies work on a larger scale with mechanized vehicles and diggers. In order to escape the extreme heat, some miners live in subterranean houses, where only the rooftops show above the surface.

Conclusion

The mining method used depends on whether the gemstones are found *in situ* (in the host rock in which they formed) or as secondary placer deposits eroded from or washed out of the host rock and accumulated elsewhere. The locality, its accessibility, the availability of a local workforce, and the political situation are also factors. The specific properties of gemstones – their beauty, rarity, durability, and value, which is often a result of fashion as much as size or clarity – will also affect the mining method used. The method chosen should not damage fragile or particularly valuable specimens, as the aim is to maximize production and, ultimately, profit.

See Also

Building Stone. **Fluid Inclusions**. **Igneous Rocks:** Kimberlite. **Mining Geology:** Exploration; Mineral Reserves; Hydrothermal Ores. **Sedimentary Environments:** Alluvial Fans, Alluvial Sediments and Settings. **Weathering**.

Further Reading

Hall C (1993) *Gems and Precious Gemstones*. London: Apple Press.

Hall C (1994) *DK Eyewitness Handbook Gemstones*, 2nd edn. London: Dorling Kindersley.

Hughes RW (1997) *Rubies and Sapphires*. Boulder: R W H Publishing.

Mumme I (1988) *The World of Sapphire*. Port Hacking: Mumme Publications.

Oldershaw C (2003) *Philip's Guide to Gems*. London: Philips.

Oldershaw C, Woodward C, and Harding R (2001) *Gemstones*, 2nd edn. London: Natural History Museum Publications.

Sevdermish M and Mashiah A (1996) *The Dealer's Book of Gems and Diamonds*, 2nd edn. Gemmology (A M) Publishers Ltd.

Sinkankhas J (1981) *Emeralds and Other Beryls*. Radnor, PA: Chilton Press.

Webster R (1994) Read PG (ed.) *Gems: Their Sources, Descriptions and Identification*, 5th revised edn. London: Butterworths.

GEOARCHAEOLOGY

L Joyner, Cardiff University, Cardiff, UK

Introduction

The interface and overlap between the Earth sciences and archaeology has developed into the subject called geoarchaeology or archaeological geology, where geological concepts and methods are applied to archaeology. The definition of exactly what geoarchaeology encompasses is still open to discussion. The emergence of this new sub-discipline really took off in the mid-twentieth century when geological techniques began to be applied to archaeological sites and materials, although the origins of the use of geological principles and techniques in archaeology date back to the nineteenth century when the two disciplines were developing. The emergence of geoarchaeology coincided with the development of the 'new archaeology' which embraced a much more scientific approach to the study of antiquity. The first approach of geoarchaeology was the application of geological techniques to archaeology, but has since developed into a more integrated interdisciplinary approach. Geoarchaeology encompasses both field and laboratory techniques and includes stratigraphy for determining the succession of occupation levels, sedimentological studies related to archaeological site formation, geomorphology of archaeological landscapes, pedology, geochronology, petrology, geochemistry of sediments and artefacts to determine composition, technology and provenance, and geophysics used to detected subsurface archaeological features. Thus geoarchaeology deals with both natural materials and processes, and artificial 'manmade' materials and landforms. Some definitions of geoarchaeology only include stratigraphy, site formation processes, and geomorphology, and omit the analysis of artefacts and geophysical techniques. This overview with cover the broader definition of geoarchaeology.

Stratigraphy

The principles of geological stratigraphy are equally applicable to archaeological stratigraphy, a fact that has long been accepted in archaeology. Geoarchaeologists are able to interpret the nature of the stratigraphic units; whether they are natural deposits or are manmade. Most archaeological deposits can be treated as special kinds of sediments, and sedimentological principles and interpretations applied. A thorough knowledge of postdepositional processes is required, as these can significantly affect the original structures and textures, as well as the stratigraphy. Techniques of stratigraphy have been used to date artefacts found in specific layers, either directly through annual laminae, or indirectly through dating of, for example volcaniclastic deposits.

Soils and Sediments

The study of soils and sediments in and around archaeological sites provides much evidence for the landscape, environmental setting, as well as for human activities and natural events. An assessment is usually made of the types of deposits encountered including composition, texture, chemistry, particle sizes and shapes, sediment structures, and colour. These details can be used to identify the type of soil or sediment, and assess how the deposit accumulated through time. Soil micromorphology has been successfully used to identify areas within sites that were used for specific activities such as cooking areas, smelting of metals, and animal enclosures.

Site Formation Processes

Site formation processes often involve both natural and anthropogenic processes of accumulations or depletions of sediments and/or soils through time and space. The study of these sediments and soils involves identifying the type and composition of the deposit and its source, as well as the identification of any weathering, transportation, environmental, or postdepositional processes involved. Much experimental and ethnographic research has been conducted in order to understand the degradation processes operating at archaeological sites, and the rate and nature of sediment accumulation over time and space. For example, experimental studies have examined the way in which ditches silt up, and ethnographic studies have been concerned with the rate of degradation of mudbrick (adobe) structures.

Geomorphology

Changes occur in the geomorphology of the landscape over time, such as the position of the coastline and the courses taken by rivers. The study of these changes can reveal the shape of the landscape at a particular time in history. The shape of the landscape

would have affected the way in which ancient peoples interacted with their environment. Therefore, the study of the geomorphology of the area around archaeological sites can provide valuable evidence about the topography in antiquity and how ancient peoples used the landscape in particular periods. It can provide information on the agricultural potential, the availability of water supplies and raw materials, the communication routes available, and the defence potential of a site. Many archaeological sites that were formerly by the sea or a river are now several kilometres away from the modern shoreline or river bank. It can also reveal evidence for catastrophic events that may have affected past civilisations. For example, tephra deposits on Santorini and Crete in Greece provide evidence for the 1628 BC eruption of the volcano on Santorini (Thera) which is thought by some archaeologists to signify the decline of the Minoan civilisation on Crete. Geomorphological studies can also reveal evidence for manmade structures, for example the Iron Age Hillforts that are common in England are manmade earthworks. The geomorphological study of archaeological sites through the time periods they were occupied is very important. Past landscapes may have been very different to the pattern of the landscapes seen today, having changed through time. The interpretation of how a site operated could thus be drastically different and therefore wrongly interpreted, if the modern landscape was assumed to be identical to the ancient one. For example, the ancient Greek city of Thermopylae was located on the coast in 480 BC when the Battle of Thermopylae was fought between the Spartans and the Persians. Today the coastline is some 5 km away from this site (Figure 1).

Geomorphological studies can also provide information on how an archaeological site has changed through time and the impact this will have had on the preservation of a site, whether it has been preserved by a covering of sediment or partially destroyed by erosional processes. Geomorphological studies will provide information on environmental changes in areas around archaeological sites, and will thus aid environmental reconstructions.

The exploitation of natural resources can significantly affect the landscape. For example, the development of quarrying and mining activities leaves behind visible signs of human activity.

Geomorphological surveys involve mapping topographic features seen in the field and noting the drift geology to produce geomorphic maps. The data collected is supplemented by the use of aerial photography and satellite images, as well as by analysis of core drills

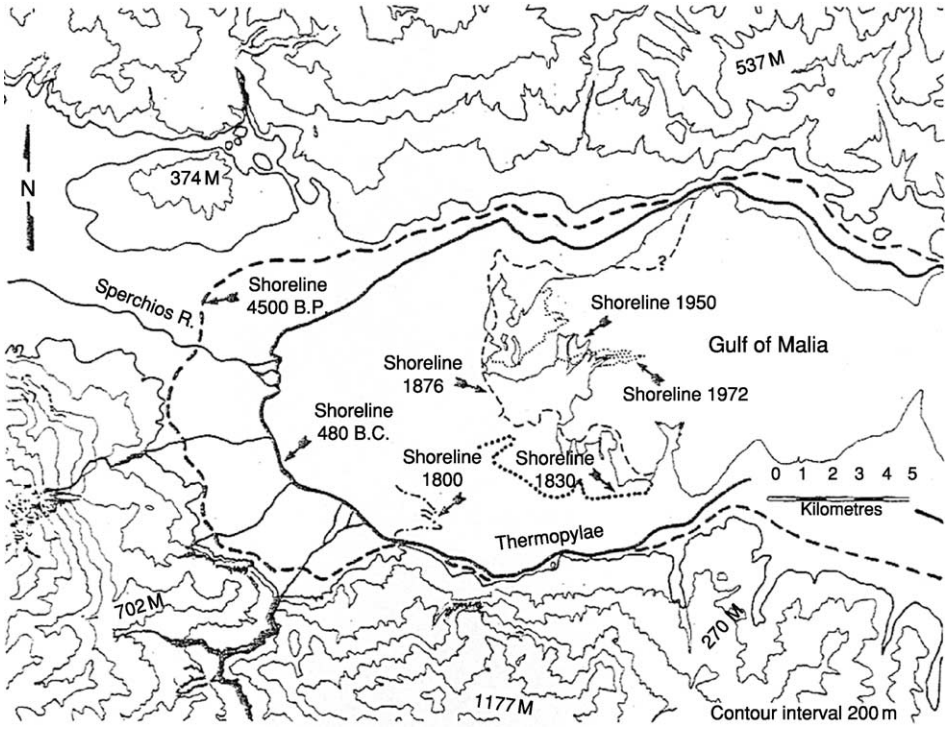


Figure 1 Map showing where the coastline has moved from 480 BC to the present day. The ancient Greek city of Thermopylae was located on the coast in 480 BC when the Battle of Thermopylae was fought between the Spartans and the Persians. Today the coastline is some 5 km away from this site. (Kraft *et al.* (1987) *Journal of Field Archaeology* 14: 181–198. Reproduced from the Journal of Field Archaeology with permission of the Trustees of Boston University. All rights reserved.)

taken from specific areas of a site. Core drilling has been used in geomorphological studies to ascertain the nature, stratigraphy, and extent of human occupation. This is a relatively rapid and cost-effective method which is potentially minimally destructive.

Aerial Photography

Aerial photography has been used since the beginning of the twentieth century to identify archaeological features not usually visible at ground level, or those that are buried. Both vertical and oblique photographs can be used, as each can accentuate different features. Stereoscopic pairs of aerial photographs are often used as they give a three-dimensional image of the topography. Some buried features will be visible through differential crop markings which result in contrasting crop growth over the buried feature. For example, ditches and pits are shown up by superior crop growth.

Geophysical Methods

A number of geophysical techniques have been adapted and applied to archaeology, since the mid-twentieth century. These include electrical resistivity, magnetometry, gravitometry, and ground penetrating radar. These techniques allow the identification of archaeological sites and their layouts thus aiding the interpretation of ancient settlements and manmade structures prior to potential excavation. These techniques are often used on their own, with no further need for excavation. They are used to locate and detect the extent of archaeological sites, as well as to map the distribution of features within known archaeological sites (Figure 2). In addition to being used to gain information about the spatial distribution of features, geophysical methods can be used to provide information on features buried at variable depths thus giving information on site development and the stratigraphic sequence of a site. The geophysical prospection of urban sites has become increasingly common with the redevelopment of 'brownfield' sites within large cities which require an archaeological survey to be carried out before development is possible. However, electrical and magnetic methods can prove problematic when used in modern urban areas.

Magnetometry identifies differences in the magnetic properties of soils which may have been caused by human activities such as burning, humic decomposition, compaction, and by building of structures. Magnetic surveying is difficult if it is carried out in an area with nearby interfering magnetic sources such as near modern buildings and power lines, which tend to obscure the weaker signals from archaeological features. Electrical resistivity detects differences in the electrical conductivity of the soil which may have resulted from the construction of ditches and mounds, and compaction. The soil moisture content is important for electrical resistivity surveying, as it cannot be undertaken if the soil is too dry.

Geohysical methods are non-destructive and more cost-effective than test excavations. A geophysical survey will provide valuable information which can be used to plan an excavation programme effectively and efficiently, as well as providing site information without the need for excavation.

Magnetic Susceptibility

Magnetic susceptibility surveys have been used to identify areas of human activity, as these can change the magnetic character of soils and sediments at the site of these activities such as cooking, metal processing, and other activities involving heating. This technique is usually used in conjunction with magnetometry for optimum results.

Archaeoseismology

Archaeoseismology is the study of the effects of earthquakes in antiquity. The devastating effects of sizeable earthquakes can be recorded in the archaeological record. The way in which buildings and other structures collapse can give information on the earthquake event, and can be used to date destruction layers at some sites. The Modified Mercalli scale of earthquake intensity is used to interpret the size of the earthquakes that caused damage at archaeological sites. Stone buildings tend to be more resistant to earthquake damage than wooden buildings, and mud-brick structures are most susceptible to earthquake damage. In addition, the type of deposit that an archaeological site is built upon will also affect the intensity of damage caused by an earthquake. Most damage is caused where buildings are constructed on unconsolidated sediments, whereas a harder underlying rock will reduce the severity of earthquake damage. The interpretation of archaeoseismic events can be confused with other events which produce similar patterns of destruction, such as landslides and poor construction techniques of buildings.

Palynology

Palynology, the study of pollen, can reveal evidence for ancient cultigens, dietary information, and the seasonal occupation of sites. Changes in the types of vegetation on a site may be significant and can be used to indicate changes of use of a site or climate

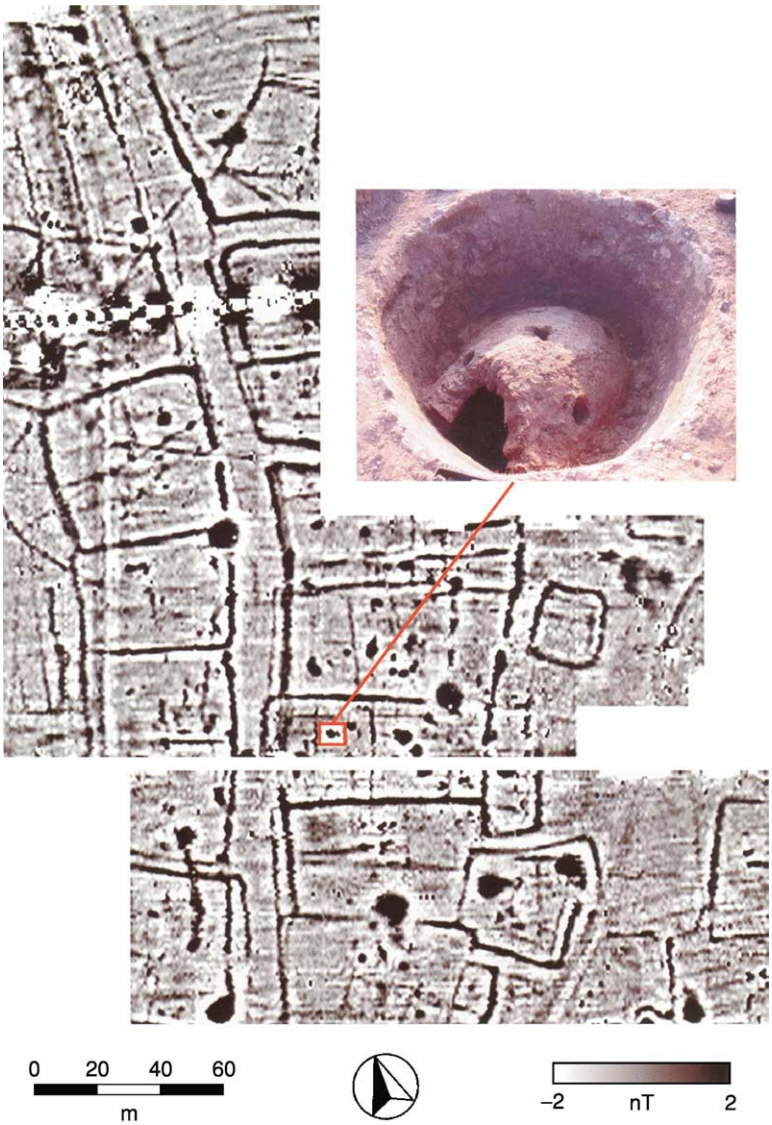


Figure 2 Sedgefield, County Durham. Fluxgate gradiometer data. 1 × 0.5 m. Romano-British settlement straddling a road. Excavation by Time Team revealed a complex of workshops and a small kiln as highlighted in the inset photograph. (Reproduced from Gaffney C and Gater J (2003) *Revealing the Buried Past: Geophysics for Archaeologists*, Tempus Publishing Ltd (Plate 19).)

change. Palynology has been extremely important in aiding environmental reconstructions from archaeological evidence, particularly with the identification of native and non-native pollen species types.

Characterization Studies

Characterization studies seek to address questions about the raw material constituents of artefacts, their provenance, ancient trade, or distribution routes, their technology of production and manufacture, and patterns of consumption. Various geochemical, mineralogical, and imaging techniques have been applied to archaeological materials to answer these questions. These techniques may involve taking samples from artefacts. However, the recent trend has been towards micro-sampling and non-destructive analysis.

Geochemistry of Archaeomaterials

The geochemistry of a range of archaeomaterials has been carried out using a number of different techniques which are also used in the Earth sciences. Major, minor, and/or trace elements are detected by a number of analytical techniques. The range of archaeomaterials include ceramics, metals, glass, lithics (particularly obsidian), and organic materials. The geochemical signature of these archaeomaterials has been used to answer a number of archaeological

questions. The chemical characteristics of particular artefacts are used to predict where the artefacts were made, provenance their source of production, as well as to source the raw materials used. This is done by comparing the geochemical signatures of unprovenanced artefacts with the geochemical signatures of locally available raw materials, or the geochemical signature of artefacts that have been produced in known localities (Figure 3). The multi-element chemical data obtained from known and unknown artefacts is usually processed by various multivariate statistical techniques. These techniques show how the samples group together, thereby suggesting a close association and possible provenance. Many artefacts of material culture are artificially produced, so cannot be directly compared with local raw materials sources. These include metal alloys, ceramics, glass, and faience. Lithics may be compared directly with rock sources. The chemistry of material artefacts has also helped to deduce what they are made of, and to suggest the technological processes involved in production. The geochemical techniques used in archaeological studies have included optical emission spectroscopy (no longer used), which was superseded by atomic absorption spectroscopy (commonly used in the 1970s to the 1990s), and then more recently by inductively coupled plasma atomic emission spectroscopy and mass spectroscopy (favoured techniques since the 1990s). Instrumental neutron activation analysis has been used for trace element determination, but this technique is declining in use as reactors are closed down. Surprisingly, wavelength dispersive X-ray fluorescence spectrometry has not been widely used in archaeological science. However, air-path energy dispersive X-ray fluorescence spectrometry has been used regularly, especially for metals, but also for other materials, as it is a non-destructive technique. Proton induced X-ray emission spectrometry and proton induced gamma ray emission spectrometry analysis have also been used, as has electron probe microanalysis. Chromatographic techniques have been applied to the study of organic archaeomaterials and these include gas chromatography and gas chromatography mass spectrometry.

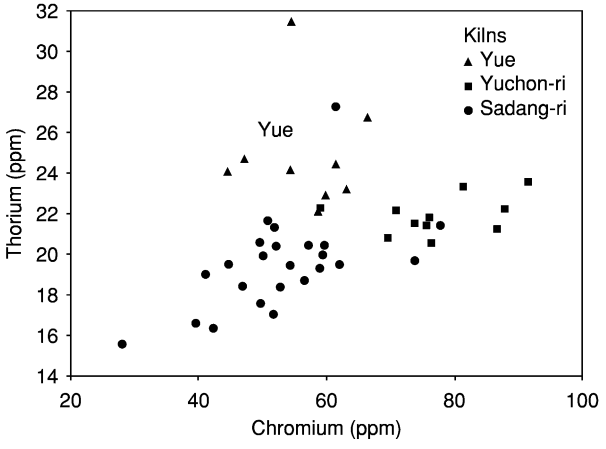


Figure 3 Geochemical grouping of trace element data from Korean Celadon ceramics dating to the Koryo dynasty (twelfth century AD). The bivariate plot shows the concentrations of Cr (ppm) plotted against Th (ppm) (determined by instrumental neutron activation analysis) for ceramic pots manufactured at different kiln sites in Korea and Chinese Yue ceramics for comparison. The trace element data shows that the geochemical signature of the clays used in these two Korean kiln sites are distinguishable from each other and the Chinese Yue ceramics. (Reproduced from Hughes M and Joyner L (2000) In: Portal J (ed.) *Korea: Art & Archaeology*, British Museum Press. Courtesy of the Trustees of the British Museum.)

Postdepositional processes can also affect the geochemistry of artefacts, substantially altering their geochemical signature. This particularly affects ceramics. Recycling, such as the addition of scrap metals in metal production or the addition of cullet in glass production, obscures the chemical signatures restricting the use of geochemistry in provenancing artefacts made of metals and glass.

Metals and Ores

Metals such as gold, silver, lead, tin, copper, iron, and arsenic were used in antiquity. Some were used in their raw state, such as gold and copper, and others were alloyed, such as copper and tin to produce bronze. Geoarchaeologists are concerned with all aspects of metal production, from ore sources to extraction and mining, to roasting of ores, smelting, melting, refining, alloying, casting, and the use and trade of metal artefacts (Figure 4).

Stable Isotopes

The use of stable isotope analyses in artefact provenance has increased rapidly in recent years. Carbon, oxygen, sulphur, strontium, and lead have all be used to provenance materials such as lithics, metals, and glass. Marble provenancing was one of the first studies to use stable isotopic data to discriminate between sources in the Aegean region. A large database of the isotopic signatures of $\delta^{13}C$ and $\delta^{18}O$ for marble quarries in Greece and Turkey has been built up which has been used for provenancing artefacts and for associating broken pieces of artefacts (Figure 5).

Mineralogy of Archaeomaterials

Many artefacts of material culture are made from geological raw materials such as rocks for building and statuary, clays for ceramics, sand for glass

production, limestone and gypsum for plaster production, gemstones for decorative purposes, minerals for pigments, and metals for tools. Mineralogical techniques that are commonly employed in the Earth sciences have been applied to the study of ancient artefacts. These include optical microscopy, polarising light microscopy, and X-ray diffraction.

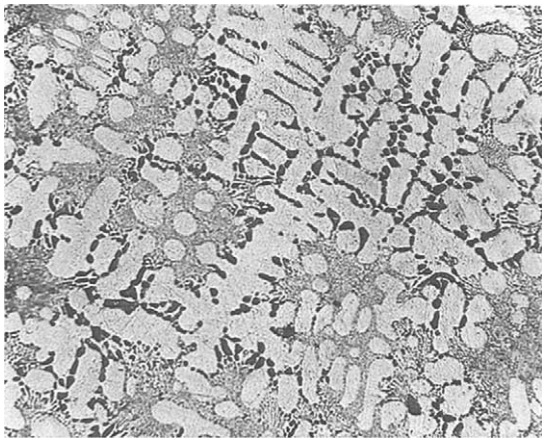


Figure 4 Structure of the cast silver-copper alloy of an Egyptian silver figure of Nefertum (EA66818). Polished and etched sample. Field of view 0.5 mm across. (Reproduced from Cowell M and La Niece S (1991) In: Bowman S (ed.) *Science and the Past*, British Museum Press. Courtesy of the Trustees of the British Museum.)

Other techniques, which are less commonly used in the Earth sciences, have also been applied to the study of ancient artefacts, and these include infrared spectroscopy and Raman microscopy.

Lithic Petrology

In archaeology, the term lithic or stone is used to describe rock that has been utilized by humans. The use of lithics in antiquity dates back to before the Paleolithic period and includes building stone, statuary, tools, vessels, sarcophagi, jewellery, and pigments. Petrological examination of lithic artefacts has allowed potential sources of these raw materials to be suggested. This allows suggestions to be made for potential trade/exchange routes that may have been taken to import the lithic materials/artefacts to a site where they were excavated, as well as highlighting contacts between past groups of peoples. For example, the identification of Preseli Mountain stone at Stonehenge has raised the possibility that the stone was transported from South Wales to Wiltshire specifically to be used in the construction of Stonehenge.

Ceramic Petrology

Petrological techniques have been applied to pottery to determine clay paste compositions, the possible

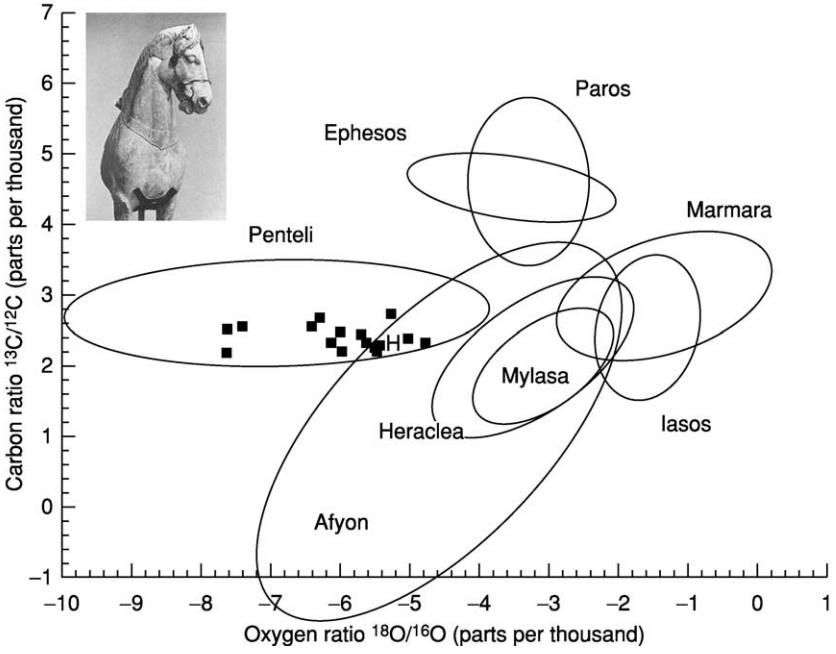


Figure 5 Stable isotope analysis for carbon and oxygen of the marble sculptures from the mausoleum of Halicarnassus including the horse, denoted by H (see inset), superimposed on the data for Classical marble quarries. The ellipses represent 90% confidence limits for the samples from the named quarries. (Reproduced from Hughes M (1991) In: Bowman S (ed.) *Science and the Past*, British Museum Press. Courtesy of the Trustees of the British Museum.)

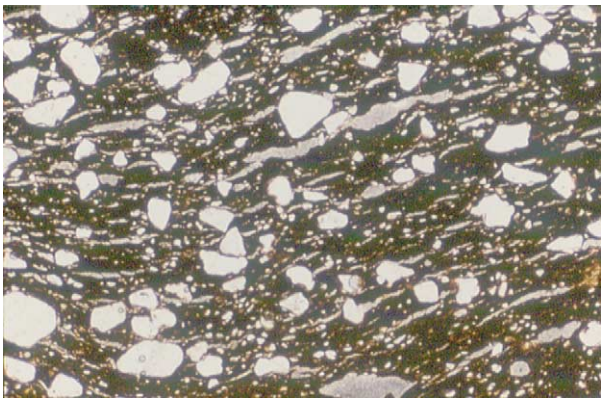


Figure 6 Thin section of an Early Byzantine pottery sherd from a cooking pot excavated from the Sanctuary of St Lot, Deir 'Ain Abata, Jordan. Plane polarized light. The fabric shows the clay micromass (brown) in which rounded quartz sand grains (white) have been intentionally added (temper). The thin elongate structures are voids which are naturally incorporated into the clay as it is worked (wedging) prior to vessel formation. This vessel has been wheel-thrown and has been well-fired in an oxidising atmosphere. Field of view is 3 mm across. (Reproduced from Joyner and Politis (2000) *Internet Archaeology* **9**. http://intarch.ac.uk/journal/issue9/daa_toc.html)

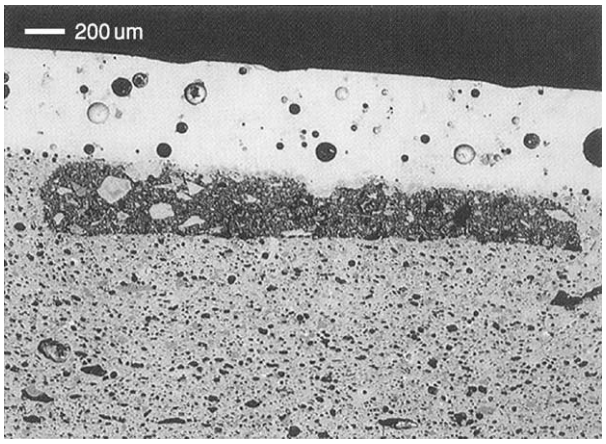


Figure 7 Back Scattered Electron Microscope image of a Korean Celadon ceramic dating to the Koryo dynasty (twelfth century AD) showing the ceramic body (grey), white inlaid decoration composed of kaolinite (rectangular dark area) and alkaline glaze (white). Note the scale bar. (Reproduced from Hughes M and Joyner L (2000) In: Portal J (ed.) *Korea: Art & Archaeology*, British Museum Press. Courtesy of the Trustees of the British Museum.)

addition of temper (material such as sand, shell, or crushed pottery or rock added to the clay to improve its workability and/or firing properties), provenance, and firing conditions (Figure 6). The use of ceramic petrology has revealed the complexity of pottery making and the long distance trade of pottery, either for itself or for the contents the vessels may contain. The study of prehistoric pottery in the American Southwest in the 1940s revealed a widespread trade which was unexpected in such an early period. Petrological techniques have also been applied to bricks, mud bricks, plasters, cements, mortars, and concretes.

Technology Studies

Various analytical techniques have been employed to deduce how artefacts were made. Some of these techniques include the chemical and mineralogical techniques already mentioned. In addition, imaging techniques are frequently used to elucidate the sequence of manufacture of artefacts of different materials. Microscopy, scanning electron microscopy, and radiography are the techniques that are frequently used (Figure 7). For example, in pottery studies, polarising light microscopy is used to deduce the clay paste recipes used, the scanning electron microscope is used to estimate the approximate temperatures achieved during the firing process, and radiography is used to look at the forming techniques used to construct pottery vessels.

Geochronology – Dating Techniques

Before the advent of modern dating techniques, archaeologists relied upon stratigraphy to determine the relative ages of cultural deposits. Since the middle of the twentieth century, a number of scientific dating methods have been applied in the archaeological sciences which have provided absolute ages, which are independent of the stratigraphy. These include radiocarbon dating (or ^{14}C dating), 40Ar-39Ar dating, Uranium series dating (which involves the radioactive decay of ^{238}U, ^{235}U & ^{232}Th to lead), electron spin resonance dating, thermoluminescence dating (releases accumulated energy by application of heat), obsidian hydration (the rate of absorption of water is dependent on glass chemistry and temperature, resulting in an hydration rim whose thickness can be used to estimate the time elapsed since the obsidian artefact was created), and dendrochronology (tree rings).

Conclusions

Geoarchaeology combines both the Earth sciences and archaeology in the quest for interpreting the material culture of past civilizations. It has become a broad but distinct subdiscipline within archaeology requiring a firm knowledge of both geological techniques and processes, and archaeology. It is hoped that this broad, though not necessarily comprehensive, review of geoarchaeology will give a feel for the range

of work being conducted in archaeology which requires a geological understanding of the environment and ancient materials. It is a dynamic discipline in which there is a continual interaction and cross-over of the two subject areas.

See Also

Analytical Methods: Geochemical Analysis (Including X-Ray); Mineral Analysis.

Further Reading

Bowman S (ed.) (1991) *Science and the Past*. London: British Museum Press.

Davidson DA and Shackley ML (1976) *Geoarchaeology: Earth Science and the Past*. London: Duckworth.

French C (2003) *Geoarchaeology in Action: Studies in Soil Micromorphology and Landscape Evolution*. London: Routledge.

Geoarchaeology. Journal published by Wiley Interscience.

Goldberg P, Holliday VT, and Ferring CR (2001) *Earth Sciences and Archaeology*. New York: Kluwer Academic/Plenum Publishers.

Herz N and Garrison E (1998) *Geological Methods for Archaeology*. Oxford: Oxford University Press.

Leute U (1987) *Archaeometry: An Introduction to Physical Methods in Archaeology and the History of Art*. Weinheim, Cambridge: VCH.

Rapp GR Jr and Hill C (1998) *Geoarchaeology: The Earth-Science Approach to Archaeological Interpretation*. New Haven: Yale University Press.

Renfrew C and Bahn P (2004) (Fourth Edition) *Archaeology: Theories, Methods and Practice*. London: Thames and Hudson Ltd.

Stiros S and Jones RE (eds.) (1996) *Archaeoseismology*, Fitch Laboratory Occasional Paper 6, British School at Athens.

Waters MR (1992) *Principles of Geoarchaeology: A North American perspective*. Tuscon: University of Arizona Press.

GEOCHEMICAL EXPLORATION

E M Cameron, Eion Cameron Geochemical Inc., Ottawa, ON, Canada

Introduction

Mineral deposits are small in size, ranging in width from several metres for a gold vein to a few kilometres for a large porphyry copper deposit. The deposits that were not found during the early decades of exploration are usually concealed beneath younger geological formations. All deposits can be discovered by grid drilling, but this would be enormously expensive. Thus, a body of exploration techniques has been developed to narrow the search to promising targets. The exploration geologist searches for rocks and geological features that are known to be associated with mineral deposits. The geophysicist looks from the air or ground for patterns that may indicate mineralization. The geochemist seeks element signatures that have dispersed outwards and upwards from a deposit. The aim is not to be constrained by the small physical size of a deposit, but to look for evidence that will enlarge the target and make it easier to locate. Much of the globe has been assessed by geochemical surveys guided by materials such as soils, stream and lake sediments, waters, and plants. Originally, the sole purpose of these surveys was to identify targets for mineral exploration (*see* **Mining Geology:** Exploration), but information on the geochemistry of the environment is becoming an increasingly important objective.

Near-Surface Dispersion of Elements

One of the most frequently used geochemical approaches for expanding target size is the sampling of stream sediments. Where a stream cuts across a partially exposed deposit, it removes mineral fragments and disperses these downstream. Several kilometres downstream, there may be minerals that are characteristic of the deposit. These may be detected by chemical analysis of the stream sediment or by physically separating and identifying minerals. Most mineral deposits contain sulphides (*see* **Minerals:** Sulphides). These are not stable when in contact with air, and they oxidize, releasing the constituent elements. Some elements, such as zinc, which is released during the oxidation of sphalerite, are soluble in water and move downstream in solution. Other elements are less soluble and form secondary minerals that are precipitated close to the source or move as solids suspended in stream waters; lead in galena, for example, forms the secondary mineral lead sulphate. Thus there may be 'clastic' dispersion, which is the movement of solids, and 'hydromorphic' dispersion, which is the movement of dissolved elements. Elements in solution can be measured by analysing the

water in which they are dissolved, but they tend to precipitate as the water moves downstream; analysis of sediments can therefore identify elements that have been dispersed in both clastic and hydromorphic form. In areas where streams are scarce and lakes are abundant, such as in the Canadian and Fennoscandian shields, both lake sediments and waters are used in analyses. Glaciers erode mineral deposits and disperse the minerals. After melting of the continental-wide glaciers of the northern hemisphere, vast areas of till and related glacial sediments were exposed. Evidence of 'dispersion trains' or 'fans' leading out from mineral deposits can be obtained by analysis of the tills or the soils developed on the tills. Also, because sulphide minerals may have been dispersed with the tills, oxidation of the tills and the included sulphides may lead to secondary hydromorphic dispersion from the fans into streams and lakes (Figure 1).

Once a target is defined by stream or lake sediments, soils are usually collected and analysed to define the target more closely. Plants can also be sampled and analysed. Plants have the advantage of extending their roots over an area and depth that is greater than the volume comprising a soil sample, and thus may give a more representative estimate of trace elements present. In arid regions, some plant roots extend to great depth to reach water, and thus may be mineralized near a deposit. A few species of plants grow in abundance on soils with a high content of a particular metal and can be used to identify the metal-rich area.

The 'mobility' of an element is a critical aspect of its usefulness in geochemical exploration. Mobility is an indication of how far an element may travel dissolved in water and thus broaden the signal derived from the mineral deposit. Some metals dissolve as cations, others dissolve as anions. Cations are most soluble at low pH. Sulphide deposits that oxidize at the surface may produce waters with a pH of 3 or even lower. In acidic waters, most metal cations are soluble. As waters derived from such a deposit flow from the deposit, they mix and become diluted by other waters and their pH increases. As the pH increases, cations are precipitated in a sequence that corresponds to their different chemical properties. In water that originally contains lead and zinc, lead will be precipitated first, at a pH lower than the pH that precipitates zinc. Elements that dissolve as anions behave differently, because these are most soluble at neutral and high pH; having a negative charge, they are not adsorbed by colloids with a negative surface charge. Because most surface waters and groundwaters are not acidic, elements dissolving as anions, such as uranium, molybdenum, arsenic, and rhenium, are most useful as wide-travelling indicators.

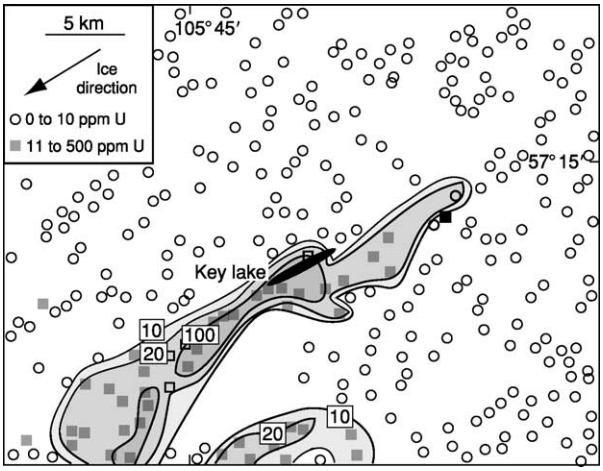


Figure 1 Uranium content of lake sediments from the uranium–nickel deposit around Key Lake, Saskatchewan. Contours at 10, 20, and 100 ppm U. Ice has dispersed fragments of the mineralization to the south-west in glacial tills. Following retreat of the ice, the tills and the uranium and nickel minerals were weathered, then uranium was dissolved as an anion and moved into the lakes. Within the lakes, the uranium precipitated in organic-rich centre-lake sediment. Anomalous contents of uranium in waters were first found in 1969. Subsequent detailed geochemistry, including lake sediment sampling and boulder tracing, led to the discovery of the deposit in 1975. Modified with permission from Maurice YT, Dyck W, and Strnad JG (1985) Secondary dispersion around the uranium–nickel deposit at Key Lake, northern Saskatchewan. In: Sibbald TII and Petruk W (eds.) *Geology of Uranium Deposits. Canadian Institute of Mining Metallurgy Special Volume,* pp. 38–47. Montreal: Canadian Institute of Mining, Metallurgy and Petroleum.

Diamond Exploration

Diamonds are a high-value commodity. Raw diamonds worth six billion dollars are sold each year; a single mine may produce several billion dollars worth of diamonds over the life of the mine, and one-half billion dollars is expended each year on exploration. Diamonds come from the upper mantle. Fragments of diamond-bearing rock are picked up by kimberlite magma (*see* **Igneous Rocks:** Kimberlite) rising from the deeper mantle en route to the surface. The magma has a high content of volatiles and moves upwards at a speed that has been likened to that of an express train. When the kimberlite reaches the surface, the sudden release of magmatic gas and interaction with groundwater causes explosions, producing pipe-shaped diatremes filled with fragmented kimberlite and country rock. Most pipes have a small surface area, usually 50 to 500 m in diameter, but occur in clusters. Kimberlite is soft and easily eroded, forming depressions that become lakes or are covered by overburden. The ease of erosion, which leads to their

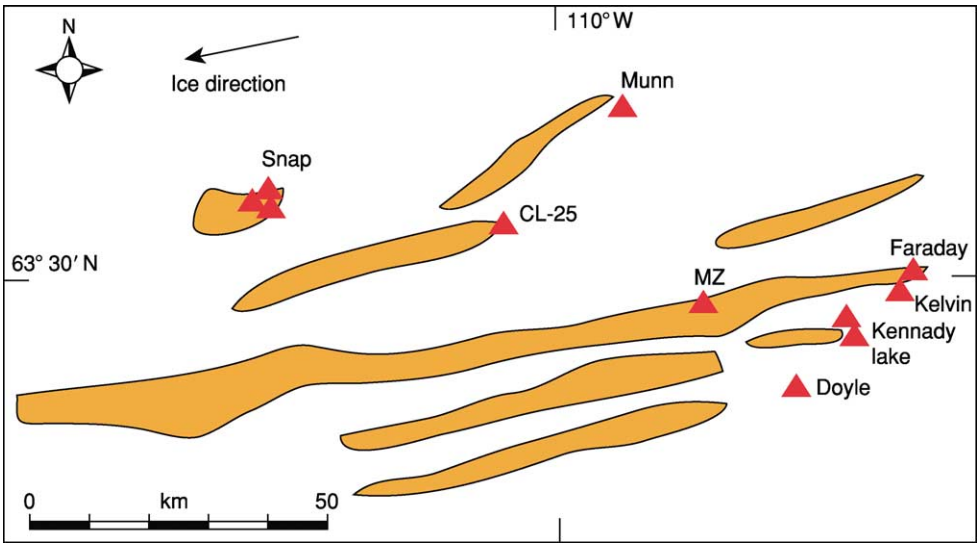


Figure 2 Dispersal trains of kimberlite indicator minerals in glacial tills, south-east Slave province, Canadian Shield. Triangles indicate locations of known kimberlites. Indicator minerals first found in the Mackenzie valley, 600 km to the west, led Charles Fipke and Stewart Blusson to the discovery of this rich diamond province. Original figure by John Armstrong.

concealment, also provides a key to their discovery: their mineral constituents may be dispersed far by streams, glaciers, or wind.

Being directly derived from the mantle, the minerals forming kimberlites are different from those of most other rocks exposed at the surface and are of higher specific gravity. When these minerals (for example, pyrope garnet, chrome diopside, chromite, ilmenite, and magnesian olivine) are identified in a heavy mineral concentrate of stream or river sediment or till, they may be followed to the source. In the Slave diamond province of northern Canada, slim, pencil-like glacial dispersal trains of kimberlite indicator minerals (KIMs) in glacial tills can extend far from their sources (Figure 2). Once the approximate sources of the KIMs are determined, other techniques, mainly geophysical, can be used to locate the kimberlites. Diamond, being so hard, is well preserved during erosion and transport, but its concentration is so low in the kimberlites – less than a part per million – that it is rarely seen in KIM concentrates (*see* **Igneous Rocks:** Kimberlite).

Exploration for diamonds would be simple if every kimberlite contained economical amounts of the mineral. But that is not the case; less than 1% of kimberlite intrusions is worth mining. So, how can KIMs that relate to diamondiferous kimberlites be identified? The clue to the answer was first provided by the electron microprobe analyser. Diamonds often contain small inclusions of other more common minerals, derived from the same source region in the mantle. Using the microprobe, these can be identified and their compositions determined. Thus, it was found that inclusions of pyrope garnet in diamonds have lower contents of calcium than do similar garnets from elsewhere, and chromites in diamonds have unusually high chromium contents. Ilmenite has high magnesium and low ferric iron contents. With the knowledge gained from inclusions, kimberlite minerals with specific compositions, derived from the part of the mantle where diamonds coexist, but much more abundantly than diamonds, could be sought in the KIM concentrate. Ilmenite compositions have an additional significance. Diamond is a reduced substance and can be destroyed if its host region in the mantle becomes more oxidized, or during transport within a kimberlite magma. Ilmenite, with high magnesium and low ferric iron contents, is thus indicative of reduced conditions that preserve diamond. Figure 3 shows four different fields containing levels of MgO and Fe_2O_3 for ilmenite from four kimberlite localities in southern Africa. In the most oxidized field (1), no diamonds are preserved. In the most reduced field (4), diamond crystals are preserved and show late-stage overgrowths.

Deep-Penetrating Geochemistry

Most undiscovered mineral deposits lie beneath some form of cover, which may range up to several hundred metres in thickness. The cover may be laterite, the product of severe modification of the underlying rocks during tropical weathering. In this case, there may be sufficient geochemical clues retained from the parent material that analysis of the lateritic soils can identify the presence of a buried mineral target.

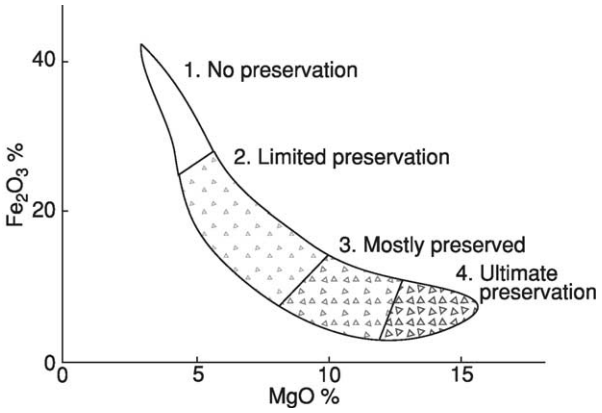


Figure 3 Composition of ilmenite as a predictor of the preservation of diamond in the host kimberlite. Figure based on samples from kimberlites in southern Africa. See text for discussion. Modified with permission from Gurney J and Zweistra P (1995) The interpretation of major element compositions of mantle minerals in diamond exploration. *Journal of Geochemical Exploration* 53(1–3): 293–309.

However, in many cases, the cover is 'exotic': it has been derived from elsewhere and has no relationship to the underlying rocks. Examples are piedmont gravels, desert sands, alluvial deposits, and most types of glacial deposits. Exotic cover makes it more difficult to locate the buried target; identification must depend on the migration of elements from the deposit, through the cover, to the surface. The indicator elements must migrate either as a gas (for example, radon or sulphur gases) or as metals dissolved in water. This involves movement through a zone that is unsaturated in water, the vadose zone, which can be tens to hundreds of metres thick in arid climates. In the vadose zone, the mineral grains are surrounded by a thin film of water, which moves slowly downward to the water table. Migration may take place by diffusion of ions through the water film or as gases diffusing through air space in the vadose zone. An alternative mechanism to diffusion is advection, which is the movement of an entire mass of air or groundwater containing the gas or dissolved ion.

The exploration geochemist uses a variety of gases in soils, including hydrocarbons and sulphur gases, to explore for buried deposits. For the migration of gases from depth, an interesting experiment was performed by CR Carrigan and colleagues at the Nevada Nuclear Test Site. Chemical explosives equivalent to a 1-kt nuclear charge were detonated at a depth of 400 m in bedded tuff. Two bottles of gas were placed near the charge, one containing ^{3}He, the other containing SF_6. Sampling sites were established at the surface to detect these gases. SF_6 was the first gas to be detected, after 50 days, along a fault, during a strong barometric depression. ^{3}He was first detected at the surface 325 days later. These results do not fit with gaseous diffusion; compared with SF_6, ^{3}He has a much higher diffusivity and should reach the surface long before SF_6 does. The diffusivity of SF_6 is such that it would require tens to hundreds of years to reach the surface, yet it has happened in days. Why? The reason is that gaseous diffusion is overridden by a much faster advective mechanism, barometric pumping. High barometric pressure forces air down fractures and into pore space in the rock, around the fractures. Gases within the rock mix with air in the pores. When the barometric pressure drops, air in the porous rock, now containing the gases, returns to the fracture and, after several cycles of high and low pressure, reaches the surface. Pumping occurs because the volume of air entering rock porosity is much greater, compared to the volume of air present in the fractures. Pumping provides the "breathing volume" that permits large vertical movements during high-pressure 'inhalation' and low-pressure 'exhalation'. Barometric pumping can withdraw gases from depth several orders of magnitude faster than is possible by molecular diffusion. But why did it take much longer for ^{3}He, compared to SF_6, to reach the surface? This is because a gas with high molecular diffusivity can more readily diffuse out of the upward-moving air–gas mixture in the fractures, into porous wall rock.

In the search for buried deposits, great attention has been given to the detection in soils of elements that have been transported to the surface as dissolved constituents in water. Initial studies focused on chemical diffusion of ions in water. But rates of diffusion are very low and the water films around grains in the vadose zone move downwards at a rate that can be several orders of magnitude faster than the rate at which ions diffuse upwards. It is analogous to trying to walk up a very fast down escalator. As with gases, the most effective way to bring dissolved ions to the surface is by advection, the upward transport of mineralized groundwater. Capillary migration is one such mechanism for advective transport, but it is effective only for depths below the surface not much greater than 10 m. A different process, earthquake-induced surface flooding, or seismic pumping, has been identified as being effective over much greater depths. This process has long been known to seismologists. Surface flows of groundwater occurred along fault lines in desert areas of Iran during earthquakes in 1903, 1923, and 1930, and more recent examples have been observed in Montana and California. During earthquakes, stress fields can become compressional, closing fractures in basement rocks and forcing groundwater to the surface, up faults.

Because many mineral deposits occur along faults, mineralized groundwaters bathing a buried deposit can become incorporated in the upward flow. The Atacama Desert, a region in northern Chile of high seismicity and hyperaridity, produces the world's greatest quantity of copper. The principal deposits were found exposed at the surface, but large areas of prospective ground are covered by piedmont gravels of Miocene age. EM Cameron and colleagues have found soils with elevated amounts of various porphyry indicator elements (copper, molybdenum, rhenium, and selenium), above porphyry deposits buried beneath gravels. These anomalies lie along fracture zones in the indurated gravels, which appear to have formed during the reactivation of older basement faults that were involved in the formation of the deposits (Figure 4). These soils are more saline than are soils away from the fracture zones. Mineralized, saline groundwaters have been pumped to the surface during earthquakes; these have evaporated, then the constituents have been redistributed by the infrequent rains. In the case of the Spence copper deposit in northern Chile, the proportions of elements found in the soils correspond to those found in groundwater 80 m below.

A very different environment is found in the Abitibi belt of Ontario, Canada. Here world-class gold and base metal deposits of Archaean age have been mined for over 100 years. The first deposits were found exposed in areas free of glacial cover. However, large parts of this highly prospective region are covered by thick, impervious lacustrine clays deposited during retreat of the ice sheet 10 000 years ago. Recent work by SM Hamilton and co-workers has provided encouraging evidence that geochemistry can 'see' through the clays. Hamilton's group has identified reduced (i.e., low-eH) columns in the clays; the columns extend from the sulphide subcrop to the water table and represent an upward flux of reduced substances (e.g., Fe^{2+}) derived from oxidation of the sulphides. On reaching the water table, these substances react with atmospheric oxygen to produce acid: $Fe^{2+} + O_2 + 2H_2O\ Fe(OH)_3 + H^+$. The acid dissolves carbonate originally contained in the clays,

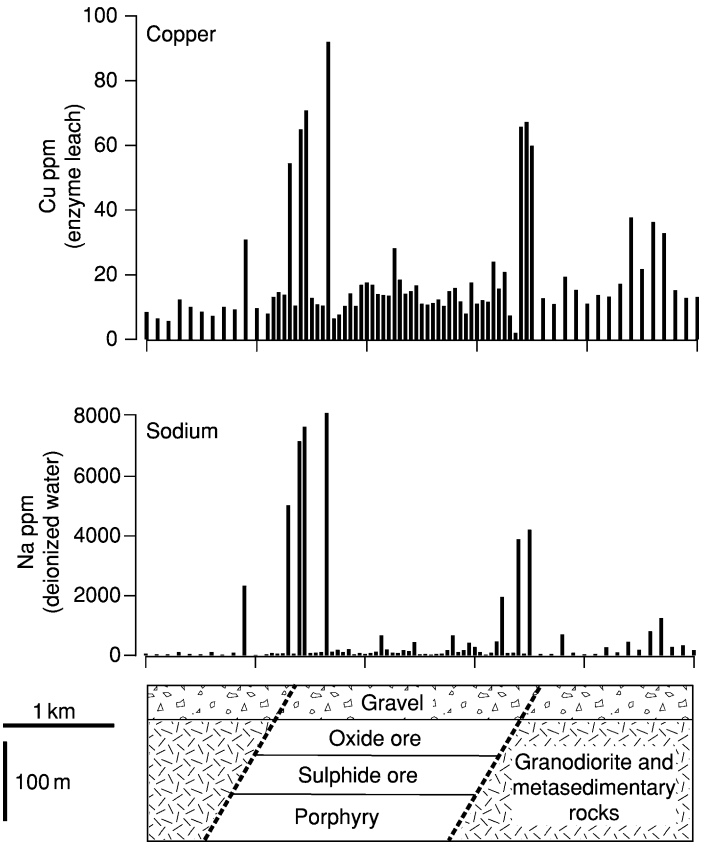


Figure 4 Analyses of soil samples for sodium and copper along a 5-km traverse across the Gaby Sur copper porphyry deposit, northern Chile. The deposit contains 400 million tonnes of 0.54% Cu and is overlain by piedmont gravels of Miocene age. Sodium was extracted by deionized water and copper was extracted by enzyme leach. Groundwater is found only in the basement below the gravel where drill holes intersect fracture zones, such as the boundary faults (shown by dashed lines). The anomalies for sodium and copper in the soils are interpreted to represent mineralized groundwater pumped to the surface during seismic activity.

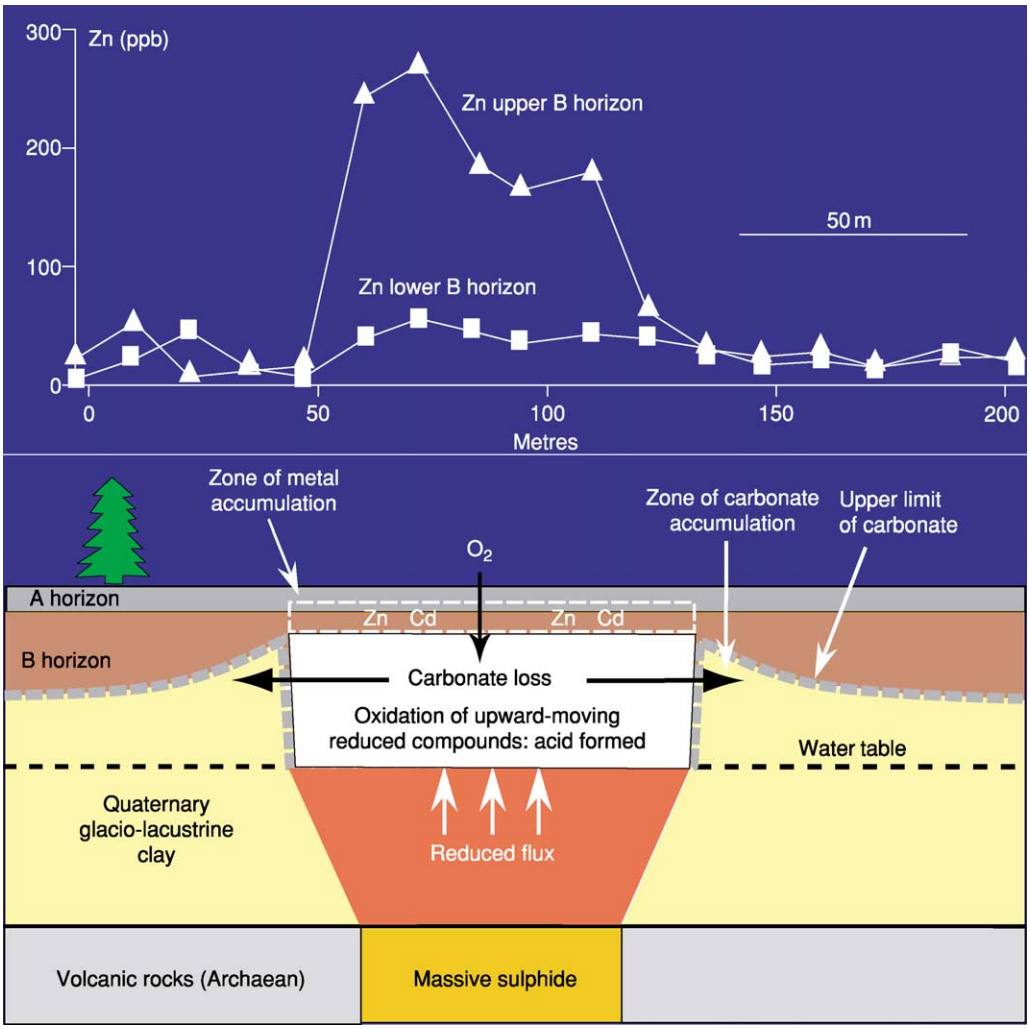


Figure 5 Diagram showing effects of an upward flux of reduced material from a subcropping massive sulphide body to the surface through impermeable clay. The oxidation of the reduced material by oxygen above the water table causes formation of acid, which removes the carbonate from the clay. Zinc and other base metals from the massive sulphide are precipitated in a thin zone of soils at the top of the B horizon. Zinc analyses of soil after extraction were by enzyme leach.

forming a zone of low pH and low carbonate in soils directly above the sulphides (Figure 5). Other elements from the sulphides, such as zinc and lead, also move to the surface and are precipitated in a thin zone, which may be the upper B horizon or Ae horizon, immediately below organic-bearing soils. Isotopic measurements of lead in this horizon show that it is similar in composition to the lead of sulphide mineralization and much different from the more radiogenic lead of the clays. Chemical diffusion is much too slow a process to explain the upward flux of these elements through 30 m or more of clay in a period of only 10 ky; other processes must be involved, and these are currently being investigated. This observation may be relevant to the use of clays to cover buried nuclear waste.

Analysis of Samples

Much of the progress in geochemical exploration has come as a result of rapid advances in analytical techniques. Today, analysts can measure a larger number of elements, to lower concentrations and at lower cost, than was possible even a decade ago. Most analyses are carried out using inductively coupled plasma mass spectrometry (ICPMS). Commercial laboratories routinely analyse samples for 50 elements at a cost of about $15 per sample. Some of the elements (rhenium, for example) are measured at concentrations below 1 part per billion (ppb). Rhenium occurs in low concentrations in molybdenite, which is present in most copper porphyry deposits. Rhenium serves as a far-travelling indicator of these deposits,

because it dissolves as an anion in groundwater. In soil from the porphyry belt of northern Chile, anomalously high concentrations of rhenium, indicative of porphyries, can be as low as 0.5 ppb, but this is still above the limit of detection when samples are analysed by a weak leach (see later) of 0.05 to 0.10 ppb Re.

As mineral deposits become harder to find, much effort has gone into amplifying the geochemical signal from the deposit, over the background noise. One of the first approaches to this was the collection of heavy minerals, as described in the preceding discussion on diamonds. Panning for gold is another example. A chemical method for amplification is selective removal or leaching of the mobile component of a soil or sediment, then measuring this component by ICPMS. 'Selective leach' reagents remove only a small fraction of the metal that might otherwise be dissolved by strong acids or by total dissolution, in the expectation that this fraction represents a more readily dissolved mobile phase derived from an ore deposit. The more abundant element fraction that comes from primary minerals forming the soil is called the endogenic phase, whereas that from external sources, including a mineral deposit, is the exogenic phase. The exogenic phase is initially introduced into the soil in water-soluble form and, as a result of soil-forming processes, is incorporated into secondary minerals, such as carbonates or iron and manganese oxides. Because the exogenic phase enters in water-soluble form, one approach is to use a weak leach that does not attack any of the minerals, but rather dissolves water-soluble salts and elements loosely adsorbed to mineral surfaces. Such is the 'enzyme leach' used to obtain the data for copper in Figure 4 and for zinc in Figure 5. For secondary minerals, ammonium acetate is commonly used to dissolve carbonate, and hydroxylamine hydrochloride at different concentrations and acidities is used separately to dissolve manganese oxides and iron oxides. Other reagents are used to dissolve the organic material that often accumulates with metals of exogenic origin.

From Regional Exploration Geochemical Surveys to Environmental Geochemical Mapping

Regional geochemical surveys using a variety of media (stream and lake sediments and waters or soils) have evolved in scale over time from covering hundreds of square kilometres to surveys that cover more than 100 000 km^2 in a field season. In the 1970s, 65% of the United States was surveyed to identify areas of potential for uranium and other minerals. Aircraft were key to this increase in productivity. Helicopters can ferry crews to sites where soils or sediments are sampled, and geochemists in low-flying aircraft can map the distribution of the radioactive series, i.e., uranium, thorium, and potassium. A few perceptive geochemists, notably JS Webb, saw early on that these surveys could reveal more information than had been mandated as their primary purpose (to identify mineral potential). In 1978, Webb and co-workers published a geochemical atlas of England and Wales based on stream sediment sampling. This showed trace element distributions that affected human and animal health. For example, areas underlain by molybdeniferous marine shales gave rise to pasture that caused molybdenum toxicity in cattle and a molybdenum-induced inability to absorb the essential trace element, copper. This problem was already known to veterinarians working in the most affected areas, but the survey revealed other areas where cattle had subclinical symptoms. Addition of copper to the cattle feed produced an increase in animal weight. Similarly, trace elements, either by their deficiency or their excess in soils and waters, can affect human health. Excess materials in the environment can have natural causes or may be the result of pollution. (*see* **Environmental Geochemistry**).

As the importance of environmental geochemical mapping became increasingly apparent, geological survey organizations began to include this purpose, in addition to resource evaluation, as a rationale for surveys. Today, environmental information has become the primary purpose of many surveys. This has resulted in modifications in the techniques applied, including the nature of the sampling media and the sampling density. Sampling of moss is now used to identify airborne pollution, for example, to map the emissions from nickel smelters in the Kola Peninsula. Early surveys for resource purposes were generally carried out at a high sampling density, with one sample per square kilometre being typical, but it became apparent that useful information could be obtained more rapidly and economically by sampling at lower densities. In 1972, the first survey of Canada's National Geochemical Reconnaissance program collected lake sediments and waters at a site density of 1 sample per 23 km^2, permitting 93 000 km^2 to be sampled in 6 weeks. There has been a progression of this trend to lower density sampling, particularly where surveys are being carried out on a multinational or global scale for environmental purposes. The Global Geochemical Baselines project collects samples on the basis of 160-km by 160-km cells. The mapping of Europe at this scale was completed in 2004 (Figure 6). Standardization of sampling and

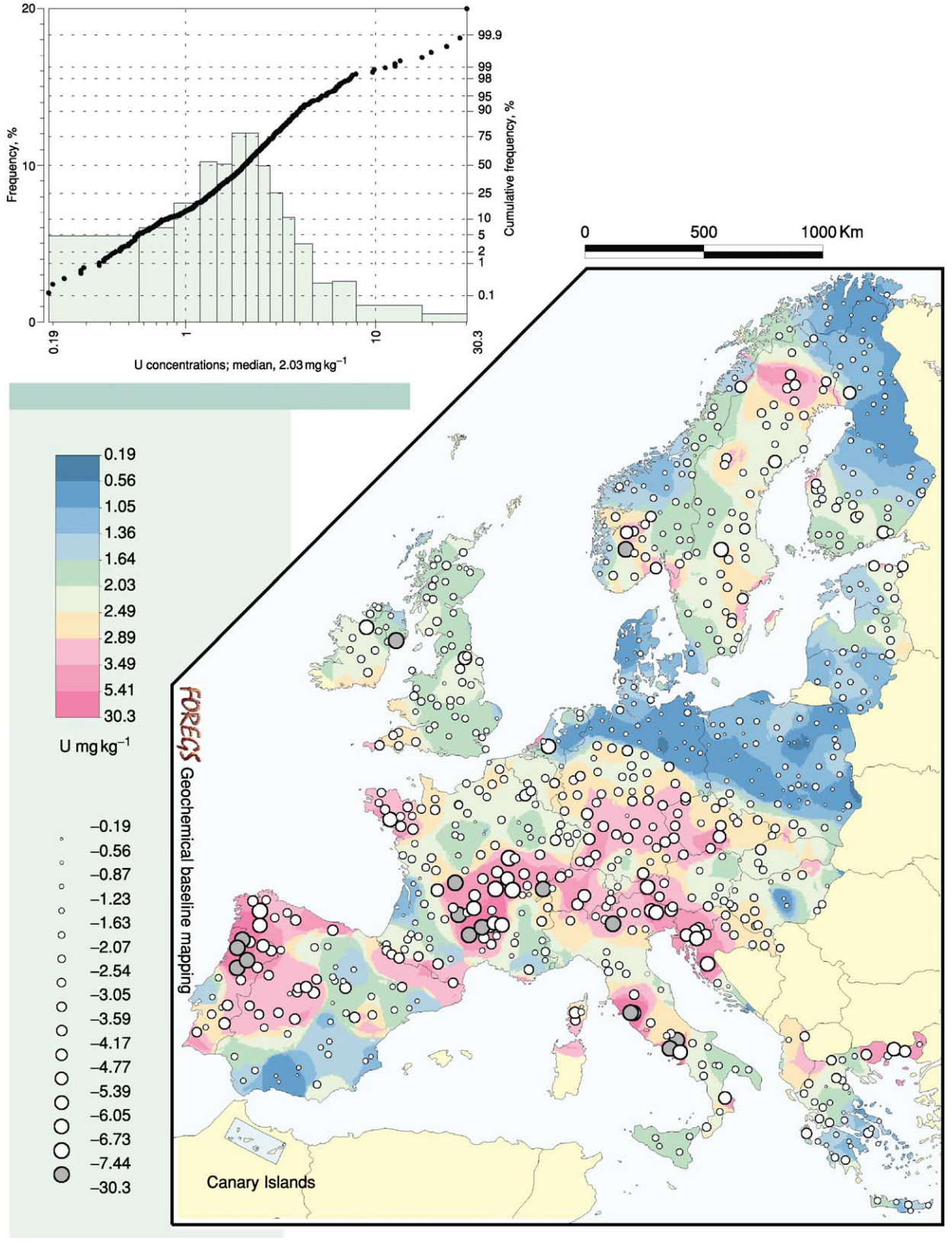


Figure 6 Geochemical map of Europe, showing the distribution of uranium in soils. Uranium concentrations of 764 samples were determined using inductively coupled plasma mass spectrometry (detection limit, 0.1 $\mu g\,kg^{-1}$). Courtesy of Reijo Salminen, Geological Survey of Finland. Survey sponsored by Forum of European Geological Surveys.

analytical techniques and quality control of the data were not always given full consideration in early surveys, but are now mandatory where comparisons are to be made on a global basis. Just as the mobile component of elements has been found to be most useful for identifying mineral deposits, for environmental surveys, emphasis must be given to measuring the fraction of the element that is bioavailable. Thus, selective leaches are used and, where waters are sampled, speciation calculations can be made to estimate the amounts of potentially toxic elements that are bioavailable.

See Also

Environmental Geochemistry. **Gold**. **Igneous Rocks:** Kimberlite. **Minerals:** Sulphides. **Mining Geology:** Exploration. **Sedimentary Processes:** Glaciers. **Soils:** Modern.

Further Reading

Brooks RR (1983) *Biological Methods of Prospecting for Minerals*. New York: Harper and Row.

Cameron EM, Hamilton SM, Leybourne MI, Hall GEM, and McClenaghan BE (2004) Finding deeply-buried deposits using geochemistry. *Geochemistry: Exploration, Environment, Analysis* 4: 1–26.

Carrigan CR, Heinle RA, Hudson GB, Nitao JJ, and Zucca JJ (1996) Trace gas emissions on geological faults as indicators of underground nuclear testing. *Nature* 382: 528–531.

Griffin WL (1995) Diamond exploration into the 21st century. *Journal of Geochemical Exploration* 53: 1–367.

Gurney J and Zweistra P (1995) The interpretation of major element compositions of mantle minerals in diamond exploration. *Journal of Geochemical Exploration* 53: 293–309.

Hale M and Plant JA (eds.) (1994) *Drainage Geochemistry*. Amsterdam: Elsevier.

Hall GEM (1998) Analytical perspectives on trace element species of interest in exploration. *Journal of Geochemical Exploration* 61: 1–20.

Maurice YT, Dyck W, and Strnad JG (1985) Secondary dispersion around the uranium–nickel deposit at Key Lake, northern Saskatchewan. In: Sibbald TII and Petruk W (eds.) *Geology of Uranium Deposits. Canadian Institute of Mining Metallurgy Special Volume*, pp. 38–47. Montreal: Canadian Institute of Mining, Metallurgy and Petroleum.

Plant J, Smith D, Smith B, and Williams L (2000) Environmental geochemistry at the global scale. *Journal of the Geological Society, London* 157: 837–849.

Rose AW, Hawkes HE, and Webb JS (1979) *Geochemistry in Mineral Exploration*. London: Academic Press.

Zeegers H and Butt CRM (eds.) (1992) *Regolith Exploration Geochemistry in Tropical and Subtropical Terrains*. Amsterdam: Elsevier.

GEOLOGICAL CONSERVATION

J E Gordon, Scottish Natural Heritage, Edinburgh, UK

Introduction

The long and varied geological history of planet Earth is reflected in the great diversity of its rocks, minerals, fossils, landforms, soils, and active geomorphological processes. In many parts of the world this diversity is under threat from a range of human activities. Traditionally, geological conservation has involved the protection, management, and interpretation of specific sites, or 'monuments', recognized to be of national or international importance on scientific grounds because of their geological and geomorphological features. More recently, a broader approach has centred on the idea of 'Earth heritage' (Earth heritage conservation) and latterly on the concept of 'geodiversity' (geoconservation). Geodiversity is the variety of geological environments, rocks, minerals, fossils, landforms, unconsolidated deposits, soils, and active geomorphological processes in a defined area. Geodiversity provides the foundation for life on Earth and for the diversity of natural habitats and landscapes, as well as for many aspects of cultural landscapes and built environments. It is therefore a valuable environmental, scientific, educational, cultural, and economic resource.

It is now increasingly recognized that proper conservation management of the non-living parts of the natural world is crucial for sustaining living species and their habitats. This leads to a broader role for geological conservation, incorporating not only the protection of geological and geomorphological features of scientific and educational value but also the management of natural environments and processes that support habitats and species. Such a role requires more integrated approaches to nature conservation and the management of sites and landscapes, which recognize the dependencies between biodiversity and geodiversity. It also embraces the involvement of people and society and their

interactions with geology, landforms, soils, natural processes, and landscapes. Geological conservation is therefore concerned with the protection and management of geodiversity both in designated sites and in the wider landscape. It includes promoting sustainable management of the natural world based on understanding Earth system processes, as well as influencing policy makers, planners, and land and water managers, and raising public awareness.

The Value of Geodiversity

The value of geodiversity ranges from the narrow scientific interest for geological research and education to the much broader economic and societal value of mineral resources and construction materials and the aesthetic and cultural benefits arising from activities such as geotourism. Geodiversity also has a key functional role in underpinning aspects of biodiversity.

Rocks, fossils, landforms, and soils are valuable resources for scientific research and education. Geology and its allied disciplines are fundamentally field-based, so that exposures of rocks and landforms are essential for continued research and education. As part of the process of scientific advancement, there is a continuing need for these sites to be available for further research in order to test and develop new theories and for demonstration and reference purposes. On economic and social grounds, such sites are essential for the training of new generations of geologists to help locate and develop the oil, gas, mineral, and aggregate resources that modern society ultimately depends upon. The applied value of geoscience is also apparent in the use of land for agriculture, forestry, mining, quarrying, building, and infrastructure, all of which are closely related to the underlying geology, landforms, and soils. Rocks, landforms, and sediments also record climate change and the dynamic history of the landscape. These environmental records allow present changes to be set in the context of past natural variations. This is important since past landscape changes provide a basis for understanding present changes and predicting future events and risks, such as those associated with flooding, earthquakes, volcanic activity, and coastal erosion.

From an Earth heritage perspective, the geological record contains a remarkable story of the Earth's history, including the evolution of life. This story embraces plate tectonics, continental drift, mountain building, volcanism, oceans opening and closing, dynamic surface processes, climate change from ice-houses to hothouses, and a record of life through evolving habitats, changing biotas, and species extinctions. It is a story that deserves to be more widely appreciated as a fundamental part of the natural world in which we live. On a worldwide scale, many geological sites are unique or contain exceptional features or landforms (e.g. the Grand Canyon, the Burgess Shale, the Channelled Scablands, Giant's Causeway, and Dinosaur Provincial Park). Many sites and areas also have great historic and cultural value because of their crucial role in the development of geoscience and their association with key historical figures. For example, many important geological principles have been developed in Scotland, through the work of James Hutton (*see* **Famous Geologists:** Hutton; Lyell), Archibald Geikie, James Geikie, Hugh Miller, Benjamin Peach, and John Horne, and applied worldwide. Many of the names for periods of geological time are derived from specific geographical localities (e.g. Cambrian, Ordovician, and Silurian) and form part of the local and national heritage. At a local community level, geodiversity provides the basis for landscape heritage and sense of place. Many sites have local-amenity, aesthetic, historical, industrial-heritage, and cultural values. Urban landscapes, too, can have a close affinity with geology; this is exemplified by the volcanic hills of Edinburgh, but is also apparent in the use of local building stone (*see* **Urban Geology**).

Rocks and landforms are the basis of landscapes and scenery, which are highly valued for aesthetic and practical reasons for tourism and a range of recreation and leisure opportunities. Conservation of geodiversity therefore helps to maintain an economic resource for geotourism and recreation activities.

Geodiversity also has much wider value through its functional links with other parts of the natural heritage. Rocks, landforms, soils, and geomorphological processes provide the basis for the diversity of valued habitats and species. Management of sites for the purpose of conserving biodiversity therefore requires an understanding of their geological and geomorphological settings and current process dynamics. For example, many internationally important habitats owe their origins to geological and geomorphological processes (e.g. species-rich chalk grasslands, coastal sand-dune systems, and the wintering sites of internationally important goose populations in the estuaries and salt marshes of the UK). Active geomorphological processes maintain dynamic habitats and ecosystems through sediment and water flows, nutrient cycling, and hydrology. For example, Atlantic salmon rely on the availability of river-channel features such as pools, riffles, and glides, which in turn depend on the underlying geology and ongoing geomorphological processes of floods, erosion, and deposition. Maintaining natural process systems is a key part of conserving biodiversity as

well as of maintaining the natural landscape. Degradation of landforms or soils or the interruption of natural processes through the building of river or coast defences or the construction of dams can have adverse impacts on biodiversity.

Soil is an often overlooked component of geodiversity but provides the medium for plant growth, agriculture, forestry production, and habitat development. It also performs key environmental functions (e.g. water filtering and storage, carbon storage, acting as a reservoir for biodiversity) and records climate and land-use changes.

Pressures on Geodiversity

Pressures on geodiversity are many and varied and arise principally from urban, industrial, and infrastructure developments and changes in land use. Geological sites typically occur as natural or man-made exposures. Some sites are therefore created and maintained by natural processes; others are created by human activities such as quarrying. While there is a view that rocks and landforms are robust, significant damage and loss of key sites have occurred in the past and are ongoing. The main impacts include physical damage, destruction or removal of the focus of interest, loss of visibility and access to exposures through burial by landfill or concealment by vegetation, damage to site integrity through fragmentation of the focus of interest, and disruption of active geomorphological processes. For example, exposures in disused quarries can be lost through landfill, prime glacial landforms can be destroyed by quarrying for sand and gravel, and key exposures can be sealed and natural processes disrupted by coastal protection measures, river-bank protection, and flood defences.

Mineral extraction can have positive and negative impacts. Quarries and gravel pits are a significant geological resource, particularly in areas where natural exposures are poor or scarce. Quarrying may reveal new sections of value, and many important sites are in former quarries, where the focus of the geological interest would not otherwise have been exposed. In some cases, quarrying may pose a direct threat to particular landforms; for example, limestone quarrying may destroy parts of cave systems and limestone pavements. In other cases, the loss of the surface landform and morphological integrity through quarrying, for example of an esker system, may need to be balanced against the potential value of new sections revealing the three-dimensional architecture of the constituent deposits.

Planning conditions usually require the restoration and landscaping of quarries, frequently involving landfill. The economic value of landfill space often competes against the value of conserving geological exposures. It is therefore crucial that potential geological interests are identified as far as possible at an early stage and incorporated into site restoration schemes. Road cuttings can also provide important geological exposures. Landscaping usually involves covering the cuttings with soil, seeding them with grass, and planting trees, leading to the loss of potentially valuable exposures.

A range of agricultural and forestry operations may affect geological sites and landforms. Landforms are often concealed beneath blanket commercial afforestation and damaged by extraction haul roads. Soils are under pressure from land-use practices and contamination, intensification of agriculture, afforestation, waste disposal, acid deposition, and urban expansion.

Traditional approaches to protecting coasts and rivers from erosion and flooding typically involve large-scale heavy engineering, which seals key exposures behind concrete seawalls, rock armour, or gabions. Natural processes of sediment supply and movement are disrupted, usually displacing the problem elsewhere.

There is a consensus that responsible fossil collecting can promote the science of palaeontology, providing that a code of good practice is followed. However, the irresponsible collecting of rare fossil and mineral specimens, often through the excavation of key sites for commercial gain, represents a significant loss to science as the context of the specimens is not recorded. Irresponsible collecting can also damage exposures and result in the loss of other specimens. Mechanical excavators, explosives, crowbars, and rock saws have all been used to remove fossil material, in a search for high-quality commercially saleable specimens.

Practical conservation therefore requires a combination of statutory protection and management for key sites and raised awareness of the value of geodiversity.

Conservation of Key Sites

Site Assessment

The identification and protection of key localities for research and education lies at the core of geological conservation. This approach is based on the selection of special or representative sites using scientific criteria and has been implemented in different ways in different countries through a variety of measures and instruments, including national parks, natural monuments, and other categories of protected site. It is particularly well illustrated by the British system of

national assessment, documentation, and protection of geological and geomorphological sites. Historically, this type of approach dates back to the mid-nineteenth century, early examples being the enclosure of the stumps of a former forest of Carboniferous lycopods at Fossil Grove in Glasgow in 1887 and the listing of erratic boulders in Scotland in the 1870s. The formal identification of key sites began in the 1940s with the compilation of a series of site lists, which were then added to in an *ad hoc* way. This process was superseded by the Geological Conservation Review (GCR), a major programme of systematic assessment of the conservation value of geological and geomorphological sites throughout Great Britain. Site assessment was undertaken between 1977 and 1990 and was the most comprehensive review of sites in any country. It was designed to reflect the full diversity of Earth heritage in Great Britain, spanning all the major time periods from the Precambrian to the Quaternary. Publication of the results in a series of 42 scientific volumes is now nearing completion. These describe the interests of individual sites and provide the scientific justification for their selection. Over 3000 individual localities were identified and form the basis of a network of Sites of Special Scientific Interest. These are accorded a measure of legal protection, which includes a requirement for consultation with the statutory conservation agencies over developments requiring planning consent and other notifiable activities.

The aim of the GCR was to identify sites of national and international geoscientific importance in Great Britain, based on a set of site-selection criteria and guidelines and extensive consultations within the geoscience community. Site selection is based on the concept of networks of sites representing the main features and spatial variations of geological events and processes during the main time periods. Three categories of site have been identified: sites of international importance, exceptional features, and representative features.

Many sites are of fundamental importance as international reference sites (stratotypes, type localities for biozones and chronozones, and type localities for rock types, minerals, or fossils), providing the building blocks for stratigraphy and the essential reference standards for the global correlation of rocks (e.g. Dob's Lin in the Scottish Borders is the boundary stratotype between the Ordovician and the Silurian). From a historical perspective, many sites are also internationally important classic localities in the development of geoscience, where features were first recognized or key concepts developed. In Scotland, the Northwest Highlands, Glen Coe, the island of Rhum, and Siccar Point have all provided crucial evidence for interpreting geological processes of global significance – respectively, the Moine Thrust, cauldron subsidence, magmatic processes and the origins of layering in igneous rocks, and a classic unconformity that provided the crucial evidence on which James Hutton developed the foundations of modern geology.

Some sites demonstrate unique or exceptional features (for example, the Rhynie chert in Scotland contains some of the oldest known fossils of plants and insects), while other sites demonstrate classic landforms or textbook examples of particular features, such as the Parallel Roads of Glen Roy and Chesil Beach. Many more sites contain nationally important representative examples of particular geological processes, environments, or events, which are essential for teaching and demonstration purposes and fundamental to understanding the geological history of Great Britain. Sedimentary rocks provide a valuable record of past environmental changes, and many contain valuable fossil remains that have helped to elucidate patterns of the evolution of life on Earth.

At a local level, geological conservation is pursued through the voluntary sector and the Regionally Important Geological/Geomorphological Sites (RIGS) movement. Sites of local importance are selected on the basis of their scientific and educational importance, historic interest, and aesthetic and cultural values, reflecting local rather than national values. Although these sites do not have statutory protection, many local authorities now have conservation policies for RIGS as well as other local wildlife sites. An important recent initiative has been the preparation of Local Geodiversity Action Plans in some areas. These should help to ensure greater protection for geodiversity as well as encouraging local awareness and involvement.

At an international level, many individual countries have compiled lists of geosites, particularly in Europe where there is a strong lead from ProGEO, the European Association for the Conservation of the Geological Heritage. Work is also in progress to develop international lists of sites under the auspices of the International Union of Geological Sciences, including a European initiative by ProGEO. In North America and Australasia, many geological and geomorphological features are protected by a variety of existing designations, including national and state parks, National Natural Landmarks, and provincial parks and nature reserves, although, with a few exceptions (e.g. Ontario, Tasmania, New Zealand), geological features have not been systematically assessed. A number of World Heritage Sites are so designated because of their geological features or

have significant geodiversity interests, but the list is not comprehensive or representative. Moreover, there are no international conventions or regional instruments for geodiversity comparable to those for biodiversity (e.g. the Convention on Biological Diversity or the EU Habitats Directive).

Site Management

Conservation is not only about the selection of sites but also about their safeguard. This requires the development of clear management objectives and periodic monitoring. In conservation management, there is an important distinction between 'integrity' and 'exposure' sites. Integrity sites include finite or relict features, which, if damaged or destroyed, cannot be reinstated or replaced because they are unique (e.g. fossil beds) or because the processes that created them are no longer active (e.g. Pleistocene glacial landforms). The prime management objective for integrity sites is to protect the resource. There are usually fewer options for compromise in reconciling conservation and development by employing practical management approaches such as the design of alternative solutions (e.g. excavation of a replacement exposure nearby). Conservation of active geomorphological sites depends on maintaining the freedom of the natural processes to operate across most or all of their natural range of variability.

Exposure sites are those in rock units or sediments that are spatially extensive below ground level, so that if one site or exposure is lost, another could potentially be excavated nearby. They include exposures in active and disused quarries, coastal and river cliffs, foreshore exposures, and natural rock outcrops inland. The principal management objective for such sites is to preserve exposure, but the precise location may not be crucial. Exposure sites are not usually damaged by quarrying or erosion, but are susceptible to landfill and other developments that obscure the sections.

In Great Britain, generic conservation management principles have been developed for different types of site, and site-condition monitoring is now a statutory requirement. Where natural exposures are poor, the appropriate restoration of disused quarries is crucial. However, the minerals industry now increasingly recognizes the importance of geological conservation afteruse. Through dialogue and partnerships between the statutory conservation agencies and the industry's trade organizations, good practice guidelines have been developed and practical conservation measures have been implemented for the retention of conservation sections and the provision of access to key geological exposures. Soft sediment exposures, however, continue to present long-term conservation problems, and, where the resource is finite, it may be necessary to keep the features buried and to excavate only for research purposes.

Practical conservation approaches and methods adopting low-impact less interventionist solutions have also been developed for river and coastal management. With more strategic assessment of environmental impacts, planning controls on floodplain and coastal developments, and the implementation of catchment plans and shoreline management plans, this should lead to outcomes that are more in harmony with natural processes.

Guidelines and codes of conduct for responsible collecting at fossil sites have been developed and applied. Commercial collectors are encouraged to work with specialists and museum curators to ensure that material is recorded and studied.

Sustainable Management of Natural Systems

In a wider context, the conservation of geodiversity is an integral part of sustainable management of natural systems. It is becoming clear that the management of sites for habitats and species cannot succeed without reference to the underlying geology, soils, and geomorphological processes. Understanding the links between geodiversity and biodiversity is crucial for conservation management in dynamic environments where natural processes (e.g. floods, erosion, and deposition) maintain habitat diversity and ecological functions. There is a strong case for more integrated approaches to conservation that would benefit both biodiversity and geodiversity, and this is now recognized by conservation agencies, notably in Great Britain, Tasmania, and Ontario, and also by bodies such as ProGEO.

The geological record clearly demonstrates how the Earth's natural systems have evolved in the past and how they might behave in the future. However, human activity is now an important 'geological force', reshaping the surface of the Earth through the movement of rock and soil, the building of cities, motorways, and dams, the fixing of the coast with concrete barriers, deforestation and soil erosion, and causing the extinctions of species. Not only are many resources being used at a greater rate than that at which they are being replenished, but also there are many uncertainties about the long-term environmental effects of the disposal of waste from human activities. Human activity is also having a potentially significant impact on global climate. A major challenge is to use an understanding of the Earth's processes to mitigate future human impacts, to contribute to the restoration of areas

already damaged by human activities, and to work with others to develop strategies for the sustainable use of the Earth's resources.

Geomorphological processes frequently impinge on human activity (e.g. through flooding, coastal erosion, and soil erosion), with resultant economic and social costs. Management responses often result in locally engineered solutions, such as riverbank and coastal protection measures, that are unsuccessful or simply transfer the problem elsewhere. Typically, management timeframes are based on human experience and are not sufficiently informed by the longer-term geological perspective. However, it is this perspective that is vital in assessing natural hazards and implementing sustainable management of natural resources.

Sustainable management of natural systems therefore depends on the effective application of earth science knowledge as part of the development of more integrated approaches; for example the maintenance of sediment transport at the coast or of natural flow regimes in rivers. Various guiding principles have been proposed.

- The inevitability of natural change should be recognized.
- Any management or intervention should work with, rather than against, the natural processes.
- Natural systems should be managed within the limits of their capacity to absorb change.
- The sensitivity of natural systems should be recognized, including the potential for irreversible changes occurring if limiting thresholds are crossed.
- Natural systems should be managed so as to maintain natural rates and magnitudes of change and their capacity to evolve through natural processes.
- Natural systems should be managed in a spatially integrated manner (e.g. at a catchment or coastal zone level).

Many of these principles are now being applied through a range of approaches or programmes in different countries, recognizing that landscapes are mosaics of geological, natural, and cultural features that need to be managed and interpreted in an integrated fashion. This is well exemplified by shoreline management plans and integrated river-catchment management and by the comprehensive ecosystem-based planning frameworks developed for the Oak Ridges Moraine and Niagara Escarpment in Ontario. The ecosystem approach, in particular, has been adopted as a primary framework for action under the Convention on Biological Diversity and provides a means for the closer integration of geodiversity and biodiversity on a wider scale. The importance of soil conservation is also gaining recognition. For example, in response to concerns about the degradation of soils in the European Union, the European Commission is developing a Thematic Strategy for Soil Protection that recognizes the value of the functions that soils perform and the need for soil protection to be integrated with other environmental policies.

Raising Awareness of Geodiversity

Raising wider awareness and involvement is a key part of geological conservation. At one level, as part of an integrated approach to management of the natural heritage, sustainable development, and landscape management, public bodies, industry, land and water managers, planners and policy makers, and their advisors should be aware of the value of geodiversity and its conservation, as well as having an understanding of natural processes, so that informed decisions can be made. At another level, there is a need to raise awareness of the value of geodiversity with the general public and in schools to help form the basis of a wider constituency of support for geological conservation. If communities value and take pride in their local geodiversity, they are more likely to support its stewardship and conservation, as well as contributing to public debate on the wider issues. Effective geological conservation will ultimately depend on better public awareness, understanding, and support.

Interpretation of geodiversity and geology-based tourism (geotourism) are not new, as demonstrated by the appeal of show caves, glaciers, and other natural wonders. Traditional geological interpretation, however, was often conveyed by interpretation boards using detailed and overly technical language, and providing information rather than interpretation. Recent developments have seen more effective communication, resulting in the production of more appropriate materials that are presented in stimulating ways using a range of media and based on best interpretive practices and sound educational principles. Such developments may involve more integrated landscape interpretation, linking geology, landscape, human activities, and industrial archaeology, for example. They include the designation of Geoparks and the promotion of geotourism (e.g. through the European Geoparks programme and the 'Landscapes from Stone' initiative in Ireland). As well as helping to raise awareness of Earth heritage, these activities have an economic dimension. Urban geology is also an important vehicle for raising public awareness through the exploration of the links between geology, use of building and paving stones, and architectural heritage.

Conclusion

Geodiversity is an integral part of our natural and cultural heritage that deserves to be better known and conserved for the benefit of future generations. However, in comparison with the conservation of biodiversity, the conservation of geodiversity has received much lower priority in national and international conservation programmes. This reflects, in part, a traditional focus on site-based protection for geological research and, in part, a lack of awareness at all levels of the wider significance and value of geodiversity. This has meant that geological conservation has lagged behind developments in mainstream conservation. However, there is now growing recognition of the functional links between geodiversity and biodiversity, and there are calls for more unified approaches, which will hopefully see better integration of geodiversity into conservation policy, protected area management, and sustainable management of natural systems. A key priority is to raise awareness among decision makers, policy makers and their advisors, planners, land and water managers, and the education sector, and at the same time to stimulate greater public interest and involvement.

See Also

Building Stone. **Environmental Geology**. **Famous Geologists:** Hutton; Lyell. **Quarrying**. **Soils:** Modern. **Urban Geology**.

Further Reading

Barettino D, Vallejo M, and Gallego E (1999) *Towards the Balanced Management and Conservation of the Geological Heritage in the New Millenium*. Madrid: Sociedad Geológica de España.

Bennett MR, Doyle P, Larwood JG, and Prosser CD (1996) *Geology on Your Doorstep. The Role of Urban Geology in Earth Heritage Conservation*. Bath: The Geological Society.

Commission of the European Communities (2002) *Towards a Thematic Strategy for Soil Protection*. COM(2002) 179 final. Brussels: Commission of the European Communities.

Doyle P and Bennett MR (1998) Earth heritage conservation: past, present and future agendas. In: Bennett MR and Doyle P (eds.) *Issues in Environmental Geology: a British Perspective*, pp. 41–67. Bath: The Geological Society.

Ellis NV, Bowen DQ, Campbell S, *et al.* (1996) *An Introduction to the Geological Conservation Review*. Peterborough: Joint Nature Conservation Committee.

Glasser NF (2001) Conservation and management of the Earth heritage resource in Great Britain. *Journal of Environmental Planning and Management* 44: 889–906.

Gordon JE and Leys KF (2001) *Earth Science and the Natural Heritage. Interactions and Integrated Management*. Edinburgh: The Stationery Office.

Gray JM (2003) *Geodiversity. Valuing and Conserving Abiotic Nature*. Chichester: Wiley.

Hooke J (1998) *Coastal Defence and Earth Science Conservation*. Bath: The Geological Society.

Johansson CE (2000) *Geodiversitet i Nordisk Naturvård*. Copenhagen: Nordisk Ministerråd.

O'Halloran D, Green C, Harley M, Stanley M, and Knill J (1994) *Geological and Landscape Conservation*. Bath: The Geological Society.

Parkes M (2004) *Natural and Cultural Landscapes – The Geological Foundation*. Dublin: Royal Irish Academy.

Sharples C (2002) Concepts and principles of geoconservation. Published on the Tasmanian Parks and Wildlife Service website at http://www.dpiwe.tas.gov.au/inter.nsf/WebPages/SJON-57W4FD?open

Stevens C, Gordon JE, Green CP, and Macklin MG (1994) *Conserving Our Landscape. Evolving Landforms and Ice Age Heritage*. Peterborough: English Nature.

Taylor AG, Gordon JE, and Usher MB (1996) *Soils, Sustainability and the Natural Heritage*. Edinburgh: HMSO.

Wilson RCL (1994) *Earth Heritage Conservation*. London and Milton Keynes: The Geological Society and the Open University.

GEOLOGICAL ENGINEERING

A K Turner, Colorado School of Mines, Colorado, USA

Introduction

This section explores the character of 'geological engineering' in the context of 'engineering geology'; the two distinct, but closely related professional fields. It also considers the relationships between geological engineering and several other specializations, including geotechnical engineering (*see* **Geotechnical Engineering**), ground engineering, environmental geology (*see* **Environmental Geology**), and hydrogeology (*see* **Engineering Geology:** Ground Water Monitoring at Solid Waste Landfills). The development of these

various specialized fields reflects the complexity of modern engineering design and construction, especially those designs involving the interface between naturally occurring earth materials and the engineered structure, or the use of naturally occurring materials within the constructed facility.

Geological engineering is primarily a reflection of legal and technological conditions within the USA. Technological developments in Canada, western Europe, and elsewhere generate very similar demands for individuals with appropriate technical skills, but without the same pressure to also meet professional engineering registration standards. A brief historical review of the relationships between engineers and geologists over the past 200 years provides some insight into the current situation surrounding the accepted professional stature and roles for geologists and engineers.

Individual practitioners are increasingly likely to become involved in litigation, professional liability has become an important concern for many professions in many countries, and geological engineers and engineering geologists are not immune from this condition. These concerns have led to increased professional registration options for both geologists and engineers, although the exact methods of achieving this vary from country to country.

Geological engineering has developed as a relatively small and unique specialization within the broader engineering profession. The skills of a geological engineer are becoming more desirable than ever as the technologies involved in construction continue to evolve. So the future of geological engineering would appear bright, except for the financial pressures faced by many universities, which are restricting the growth of 'high-cost' fields that attract relatively low student numbers. These trends suggest that it may be difficult for adequate numbers of geological engineers to be trained in order to maintain a viable cadre of professionals.

What is in a Name?

At first glance the terms 'geological engineer' and 'engineering geologist' appear synonymous. Because the two terms employ essentially the same two words, 'geology' and 'engineering'; although in reverse order, the opportunity for confusion is great. The word choices may be unfortunate, but the two terms represent distinct, although related, concepts concerning educational and professional endeavours.

The following attempts to differentiate a 'geological engineer' from an 'engineering geologist'. Before exploring the details, the following points should be understood:

- The term 'geological engineer' was developed in the USA in response to a combination of technical opportunities and the established legal processes for obtaining professional engineering registration in the USA.
- The initial demands for geological engineers came from the minerals and petroleum industries; the demand for significant numbers of specialists to work with civil engineers (in engineering geology) only developed in the latter half of the twentieth century.
- The term 'geological engineer' is thus most widely used in the USA. The term has been used only to a limited extent in Canada, which has a different professional registration structure, and only to a very limited extent and very recently in the UK, western Europe, and elsewhere.
- The term 'engineering geology' is widely used throughout the world in two contexts: to describe the application of geological principles relevant to engineering works, environmental concerns, and societal concerns, and to define specialist geologists ('engineering geologists') who are involved in such studies.
- In contrast to the 'geological engineer', who is trained as an engineer with additional geological knowledge, the 'engineering geologist' remains a scientist. This difference has ramifications for professional licensure and legal authority.
- Because engineers and scientists may be equally held liable for public safety and welfare issues; issues of certification, licensure, or registration increasingly affect the field of engineering geology, and the geological engineers, engineering geologists, and others involved in major engineering works. These issues are resolved in many forms in different parts of the world.

Defining 'Geological Engineering'

A geological engineer is trained as an engineer, but an engineer with a broad understanding of applied geological science. The concept of geological engineering originated in the USA in response to demands by North American industries for individuals who could apply both geological and engineering principles to the evaluation and design of projects involving earth materials, structures, and forces.

The legal circumstances in the USA related to professional practice pertaining to evaluation, design, approval, and operation of projects – in particular, the requirements enforced by official engineering registration boards in each state – demanded that these individuals have the academic credentials suitable to meet professional engineer (PE) registration criteria.

These demands developed in the early twentieth century as advances in both technology and engineering made larger and more complex engineering works feasible. New branches of engineering, such as petroleum engineering, developed in response to technological advances. Existing engineering disciplines, such as civil engineering, increased demands for new specialties to design ever larger and more complex dams, tunnels, and transportation systems. In particular, the minerals and petroleum industries required increasing numbers of exploration and production specialists and administrators, roles for which engineering training combined with geological knowledge were the basic requirements.

Consequently, a number of universities and mining schools in the western USA began to offer engineering programmes leading to a degree in 'geological engineering'. Graduates from these programmes were hired by petroleum and minerals exploration and production companies, and placed in positions where their combined geological and engineering training made them uniquely qualified. Subsequently, as professional engineering registration procedures became codified, these geological engineering programmes became accredited, allowing their graduates to later achieve the status of PE. Geological engineers that obtained PE status could legally approve designs for engineering works, an important consideration in some situations.

The term 'geological engineer' developed in the USA in response to both technological demands and to legal professional engineering registration procedures. Although the Canadian economy was even more dependent on the petroleum and mineral extraction industries, and thus experienced a similar demand for individuals with combined geological and engineering knowledge, the 'geological engineering' term was not adopted. Rather, in response to the demand for qualified graduates, many geology departments at Canadian universities either partially or entirely joined the faculties of applied science (in other words: engineering). This allowed many, if not all, of their graduates to achieve registration as professional engineers, and many individuals did so. The issues of professional engineering registration and legal liability issues are discussed more fully in subsequent sections.

The reader should note that most of the early geological engineers did not work on civil engineering projects. They were more likely to work on minerals exploration and exploitation projects with mining engineers, or on petroleum exploration and production projects with petroleum engineers. The employment situation was similar in both Canada and the USA. In the latter half of the twentieth century, major civil engineering projects following World War II placed new demands for specialists to work with civil engineers. In response to this new demand, many of the geological engineering academic programmes in the USA, as well as the Canadian geological programmes, began to provide 'options', usually three, with titles such as 'Petroleum Exploration', 'Mineral Exploration' and 'Engineering Geology'.

The majority of recent geological engineering graduates are now employed in practices focused on engineering geology (civil engineering applications), and ground water (hydrogeology) projects. Graduates may continue to specialize with more advanced degrees in such areas as geotechnical engineering, rock mechanics (*see* **Rock Mechanics**) (for tunnelling and underground construction), hydrogeology, contaminant transport (to evaluate ground pollution issues), or various geohazard mitigation studies (*see* **Engineering Geology:** Natural and Anthropogenic Geohazards) (landslides, earthquakes, or floods).

In all cases, the geological engineering programs in the USA retained their accreditation as engineering programmes, and their graduates, along with their Canadian counterparts, therefore, continued to have the right to achieve professional engineering registration.

Defining 'Engineering Geology'

In contrast to the geological engineer, the engineering geologist remains a scientist, albeit a rather applied one. Engineering geology uses geology to create more efficient and effective engineering works, to assess and allay environmental concerns, and to promote public health, safety, and welfare. The concept of geologists advising on engineering works dates from the earliest period of the science of geology, the historical evolution of relationships between civil engineering, and geology is discussed briefly in a subsequent section.

The term 'engineering geology' became widely accepted only as the previously noted demand for geological specialists to advise civil engineers developed in the latter half of the twentieth century. While the geological engineering programmes adjusted their focus to accommodate such demands, at the same time other universities in the USA and elsewhere began to offer an 'engineering geology' option within their science-oriented geological degree programmes. In general, the graduates from these science-oriented programmes cannot easily achieve professional engineering registration, although they often become members of multi-disciplinary teams undertaking a variety of construction projects and environmental evaluations.

The engineering geologist applies geological knowledge and investigative techniques to provide

quantitative geological information and recommendations to engineers for use during design and construction of engineering works, and in related professional engineering practice. Through cooperation, the engineering geologist and the civil engineer share the responsibility for ensuring the public health, safety, and welfare associated with geological factors that may affect or influence engineering works. In most cases, the public demands that the professional engineers be held responsible for the safety and integrity of their works. Thus, the engineering geologist may be considered as a specialist advisor to the design team, and may hold a position similar to an architect or other design specialist.

In this role, the engineering geologist often cannot provide legally binding approval of a design for an engineering project, many laws require a PE to make such judgments. In the USA, only a few individuals calling themselves 'engineering geologists' hold PE registration credentials; such persons are more likely to use the designation 'geological engineer'. However, there is a growing requirement for establishing a separate professional registration of geologists, especially for those individuals undertaking engineering geological investigations. In Canada, the situation is slightly different, with several Provinces having a single entity that supervises the professional registration of both engineers and geoscientists. The issues of professional registration and legal liability issues, for both geological engineers and engineering geologists, are discussed later.

Regardless of the existence of professional registration requirements, all engineering geologists carry a serious responsibility when applying specialist geological training and experience in communications with engineers. They must be aware of the requirements of the engineers during the design, construction, or operation and maintenance of facilities. They may be called upon to provide judgmental recommendations as well as quantitative data. The potential for misunderstandings must always be recognized; and a very high standard of written, graphical, and oral reporting is crucial to a successful engineering geologist (and geological engineer!).

In recent years the scope of engineering geology practice has expanded beyond its original closely defined connection with civil engineering. Many engineering geologists currently work closely with land-use planners, water resource specialists, environmental specialists, architects, public policy makers, and property-owners, both public and private, to prepare plans and specifications for a variety of projects that are influenced by geological factors, involve environmental modifications, or require mitigation of existing or potential effects on the environment.

Relationship of Geological Engineering to Associated Fields

The environmental movement has impacted on both geology and engineering, and on how these disciplines relate to the demands of society. In many western countries, resource extraction has become viewed with suspicion. At the same time, environmental concerns and demands for new and renovated infrastructure to support increasingly large urban populations (transportation, community expansion, water supplies, and waste disposal) have led to demands for new thematic products. National geological surveys continue to grapple with policies to define an appropriate response to these demands. Many now host important 'engineering geology' sections.

One result of the increased environmental awareness has been the development of a subject called 'environmental geology', which most engineering geologists consider to be largely within their field of expertise. Entirely new types of technical reports have been developed, aimed at the non-specialist and a broader public audience. Another aspect of the increased environmental awareness is the demand for assurances on safe and clean water supplies on the one hand, and their protection through the careful disposal of wastes on the other. Water is becoming the dominant factor for development throughout the world as populations increase and demands are placed on diminishing supplies. As a consequence, many geological engineers and engineering geologists are specializing in water-related topics, commonly considered the realm of 'Hydrogeologists'.

In response to the need for environmental protection on the one hand, and technological advances and economic forces on the other, engineering projects have become much more complex; bridges and tunnels have become longer, and high-speed transportation links have become common. Population growth has pushed developments into more complex geological locations where site conditions are less than optimal and geohazards more likely. These trends have led to an increase in the need for engineering specializations. 'Geotechnical engineering' (*see* **Geotechnical Engineering**) is a specialty that deals with the solution of civil, environmental, and mining engineering problems related to the interaction of engineering structures with the ground. Geotechnical engineers typically have expertise in soil mechanics and rock mechanics, and relatively little knowledge of geological science. They are predominantly civil engineers and are trained to design structures for foundations in soil or rock. For some projects their training limits their ability to account for the natural heterogeneity or complexity of naturally occurring

geological features. Under such circumstances, a consultation with engineering geologists or geological engineers is desirable. In many countries the term 'ground engineering' is used whenever soil mechanics or rock mechanics principles are employed in actions that modify the properties of naturally occurring materials. Usually these actions are directed toward making the materials stronger (i.e. capable of supporting larger structural loads) or reducing the permeability (e.g. to reduce the inflow of water into an excavation).

These disciplines (geological engineering, engineering geology, environmental geology, geotechnical engineering, and ground engineering) are frequently involved with construction sites, especially those for large and highly visible projects, and within or near large urban centres. Such applications lead to the use of two additional terms related to these typical locations: 'construction geology' and 'urban geology' (*see* **Urban Geology**).

In summary, geological engineering refers to a particular style of engineering that is predominantly engineering by training and experience, but utilizes special additional knowledge of geology. Geological engineers can perform a variety of tasks in the resource exploration and production fields, but in recent times most individual geological engineers are employed on civil engineering and environmental projects. They frequently work closely with engineering geologists, hydrogeologists, and geotechnical engineers, and there is in fact a continuum in training, experience, and background among individuals following these career paths. All such individuals may be employed in environmental geology, hydrogeology, or ground engineering projects and may further specialize in particular types of project that may be referred to as 'construction geology' or 'urban geology'.

Historical Interactions Between Civil Engineering and Geology

Geologists and engineers have interacted with varying degrees of support and antagonism for over 200 years. The industrial revolution demanded the transport of large quantities of heavy goods, first by canals and then by railways. When constructing civil works by large gangs of men using little more than picks, shovels, and wheelbarrows, the engineers had a considerable interest in exactly what they would be excavating, and how difficult the work might be.

William Smith (1769–1839), during his surveys for and construction of the Somerset Coal Canal near Bath in southwestern England, noted the regular succession of strata, and was able to correlate them with those in other locations by the use of fossils (*see* **Famous Geologists:** Smith). In 1799 Smith coloured his geological observations on a map of the Bath area: the oldest geological map in existence! In the same year Smith wrote a document, *Table of Strata near Bath*, and for this he became known as 'The Father of English Geology', although he continued to refer to himself as a civil engineer. In 1801, and subsequently in 1815, Smith produced further geological maps of England and Wales. He also conducted numerous civil engineering and geological investigations, and continued to assert the importance of geology to engineering.

Alexandre Collin, a French engineer responsible for the construction of several canals, conducted extensive field surveys to determine the characteristics of slope failures of cuttings and embankments. His 1846 treatise on the stability of clay slopes recognized that the characteristic circular failure surfaces were the result, not the cause, of the landslide movements. Even more importantly, he undertook laboratory experiments to determine the shear strength of these materials. He basically invented the subject of soil mechanics long before it became a popular civil engineering discipline. Unfortunately, his writings were poorly distributed and were largely ignored until about 100 years later.

The railways also required considerable earthworks and encouraged an ongoing close cooperation, and friendly relationship, between civil engineers and geologists during the latter half of the nineteenth century. The minutes of many meetings and the subjects of public lectures provide ample evidence of this collaboration – as, for example, the following quotation from an 1841 meeting of the Institution of Civil Engineers in London: "Mr. Sopwith called the attention of the meeting to the valuable Geological Sections presented by the railway cuttings ... the crops of the various seams of coal, with the interposing strata, were displayed in the clearest manner, developing the geological structure of the country which the railway traverses."

Beginning in the 1890s, the former close association of geology (and geologists) and civil engineering (and civil engineers) broke down. The introduction of powered machinery began to change the perceptions of many engineers, any job became feasible and 'successes', such as the completion of the Panama Canal by American engineers using much more powerful equipment (and also superior medical knowledge concerning tropical diseases) after the earlier failure of the French, merely served to increase this 'can do' attitude. A minority of engineers continued to strongly recommend that their colleagues seek geological advice. Chief among them in this period was Karl Terzaghi, 'The Father of Soil Mechanics' who strongly supported the linkage between of geology and engineering.

The attitudes of geologists during this period also contributed to the breakdown in the previously close working relationships. Increasingly, complex terminology was applied throughout the science, often based on Greek or Latin vocabularies. Engineers did not take kindly to this new language of the geologist. Contributions by geologists were increasingly considered to be irrelevant to the engineering design and construction processes. This separation of disciplines, and breakdown in communication and mutual respect, was most unfortunate, for the first half of the twentieth century witnessed the construction of major public works in many parts of the world. Yet, only the largest projects called for any significant geological consultation. Some spectacular failures were the result.

The introduction, debate about, and ultimate acceptance of the plate tectonics concept in the late 1960s completely revolutionized geology, providing a coherent underlying theory for evaluating descriptive geological observations. Engineering geological studies were able to make better predictions of subsurface conditions, especially in complex geological situations and on a regional basis. At about the same time the availability of powerful computers caused a complete change in engineering design procedures. No longer were dams and other structures designed using slide rules. Today modelling and optimization are required. These, in turn, place new demands for accurately predicting how geological materials will interact with the new engineered structure. At the same time, computers provide geologists with numerical analysis tools, including the ability to create and evaluate complex 3-D models of the subsurface.

As a consequence, in recent decades, the 'estrangement' of geologists and engineers has largely dissipated and the field of engineering geology, and the employment of geological engineers and engineering geologists on many projects, has begun to expand throughout the world.

Professional Liability Concerns

Liability is the legal responsibility for any loss or damage from ones actions, performance, or statements. Malpractice refers to improper, negligent, or unethical conduct or practice that results in damage or injury. Negligence claims are thus basically claims of malpractice. Court decisions are usually based on a legal responsibility for individuals to practice according to 'state of the art', 'best practice', or 'standard of practice' criteria.

Prior to the twentieth century, geologists were infrequently involved or concerned with legal matters. An exception was the involvement in 1839 of James Hall (*see* **Famous Geologists:** Hall) of the New York Geological Survey in assessing the conditions encountered during excavations to enlarge the Eire Canal locks at Lockport, New York. The contract specified a unit price for 'solid rock' and a lower price for 'slate rock and shale'; these classifications were subject to dispute. Hall was asked to provide expert testimony.

Today, geologists and geological engineers are increasingly likely to be involved in some aspect of litigation, as members of a team or as individuals, often as expert witnesses serving a client who is either a plaintiff or defendant. The probability of a geologist being sued for negligence or malpractice as an individual varies enormously depending on the nature of the work.

Geologists and geological engineers advising or employed by geological or geotechnical engineering consulting firms are most vulnerable to charges of malpractice or negligence. These individuals must abide by requirements for certification, registration, or licensing before working in any location. Those not meeting such legal requirements may be found guilty of breaking the law, may find their work defined as unqualified or unacceptable, and thereby be subjected to malpractice litigation.

There are differing legal opinions concerning responsibility whenever geological problems arise. In large firms, many junior geologists and engineers work under a principal who is registered, and in such cases the principal or the firm is responsible for the actions of the staff members. Individuals working on smaller projects sometimes seek to gain some protection from their clients by having them agree, by contract, to specifically indemnify them for consequences of their consulting services. Such actions are not always successful. Since geologists and geological engineers are increasingly involved in nearly all aspects of construction on or within the earth, or with earth materials, their degree of liability is growing.

Professional Registration and Certification Issues

The topic of professional registration and certification is complex, and is undergoing fairly rapid evolution in both North America and Europe. The pattern and procedures of registration of both engineers and geologists is quite different in the USA, Canada, and Europe. The following sections briefly consider them.

Geological Engineering Professional Registration in USA

Legal responsibility for the professional registration of engineers of all disciplines is delegated to 'Professional

Engineers Registration Boards' in each State. Although requirements and regulations do vary by state, most state boards require applications for 'registration' as a PE, to have: (1) obtained a university education from an engineering program accredited by the Accreditation Board for Engineering and Technology (ABET), (2) passed a 'Fundamentals of Engineering (FE)' exam, and (3) after several of years of experience, to have passed a PE exam. ABET has representatives from all engineering disciplines; it 'accredits' (that is reviews and approves) engineering programmes at universities and supervises the administration of the FE and PE exams. The Society of Mining, Metallurgy and Exploration (SME) represents the interests of geological engineering within the ABET organization.

State laws govern most environmental and hazard investigations, and many state laws require design documents for such projects to be signed by a PE. This requirement exists in spite of the fact that some projects involve geological aspects that an engineer is not necessarily competent to evaluate! The situation also means that a fully qualified and registered geological engineer can carry legal responsibilities that their engineering geologist brethren cannot. The laws reflect a reality; that the development and acceptance of an equivalent process to provide 'professional geologist' registration is occurring slowly and sporadically. However, some states, California being the earliest and perhaps the best example, do have considerable legal requirements for the registration of geologists.

Competing Approaches to Geologist Registration in the USA

Professional registration of geologists within the USA has been debated for about 20 years. Geologists employed in petroleum and mineral exploration have generally been opposed to calls for registration, while geologists involved in engineering, hydrogeology, and environmental projects, where public health and safety issues are readily apparent, have generally favoured registration efforts. State-by-State registration of geologists, following the engineering 'ABET' model appears to be the generally accepted method. California was the first state to legislate registration of geologists. Currently, 26 out of the 50 states require registration of geologists. State boards of registration, independent of the engineering boards, supervise the registration procedures in their state, and these boards cooperate through the National Association of State Boards of Geology (ASBOG).

ASBOG supervises the development and scheduling of two examinations: a 'Fundamentals of Geology (FG)' exam and a 'Practice of Geology (PG)' exam, which are administered by the state boards. The distribution of ASBOG examination questions reflects the importance of tasks performed by engineering geologists/hydrogeologists. This may be the natural result of the importance of engineering geology to ensuring public safety, this being the primary legal justification for professional registration.

A disturbing result of this examination process has been a very low pass rate for these exams, over the past decade only about 57% of candidates passed the FG exam and only about 68% of the candidates passed the PG exam. The percent passing has not materially changed from year to year over this period. These results suggest that many university graduates do not have an adequate grasp of the necessary geological skills. This has been used as an argument in support of registration of geologists in other states that currently do not require registration.

The American Institute of Professional Geologists (AIPG) also provides a 'certification' of geologists, giving those who are approved access to the title 'Certified Professional Geologist' (CPG). This certification is conducted by peer review of credentials without any examination. It has no legal standing in those states requiring registration, but does provide individuals with some 'national' credentials that may assist them when providing expert testimony and in similar situations.

Professional Registration Approaches in Europe

In Europe, the idea of 'professional registration' is quite different to North America. The professions generally are more self-regulating and professional credentials are tied to membership of professional societies within each country, and these in turn require completion of university degrees. In the UK, for example, civil engineers may become 'Members' of the Institution of Civil Engineers, thereby gaining the ability to use the designations 'MICE'. As a result, they qualify for registration with the Engineering Council as a Chartered Engineer (CEng).

With the creation of the European Union, Europe-wide credentials have evolved. European engineers can obtain approval to use the title 'European Engineer' and the designation 'Eur.Ing' from the European Federation of National Engineering Associations (FEANI), while in a similar fashion European geologists can obtain approval to use the designation 'Eur.Geol.' from the European Federation of Geologists (EFG). However, at the present time, whereas the designation 'Eur.Ing.' provides some specific legal standing, the 'Eur.Geol.' designation does not. Many European practitioners in the field of engineering

geology advertise their membership of the International Association of Engineering Geology and Environment (IAEG) as a credential showing competence, although it does not carry legal weight.

Professional Registration in Canada

Canadian registration procedures follow a path that lies between American and European registration practice. European procedures are largely related to membership of national professional societies. Canadian provinces have enacted registration laws in a similar manner to the American state legislatures. However, these laws generally designate appropriate provincial professional associations or societies as having the power of 'self-governance'. In the majority of provinces, a joint association supervises the registration of engineers and geoscientists. However, Ontario and Quebec have separate registration procedures; in the case of Ontario this developed when the creation of a joint registration procedure was stopped by the action of a group of engineers. In spite of some occasional evidence of friction between geologists and engineers, interactions are generally good and some individuals hold dual registration.

Conclusions Concerning Professional Registration

Several competing registration approaches have developed in the USA, Canada, and Europe. In North America, the professional registration of engineers has been legislated at the state/provincial level since the early twentieth century and has been accepted as necessary to protect the public interest. The case for an equivalent registration of geologists has not been so clearly made, and in fact there has been considerable opposition to such registration by many geologists.

In the USA, procedures to register engineers and geologists are administered quite independently by distinct official boards of registration. Whereas all states have engineering boards, only about one-half the states have geology boards. In Canada, the provincial legislatures delegate the registration process to professional associations, and in the majority of the provinces a single association supervises the registration of both engineers and geologists.

In the light of these developments, it is perhaps not surprising that the concept of 'geological engineering' should have arisen first in the USA, allowing engineers with specific geological knowledge and skills to become registered as engineers, while in Canada those geologists desiring registration and having the requisite skills and experience could obtain registration as geologists, engineers, or both.

The concept of professional registration is evolving in Europe. Once again, geologists are tending to lag behind engineers in embracing the need for professional registration.

International trends, especially the increased globalization of markets for consultation services as well as goods, have placed new pressures on the existing professional registration procedures. The requirements to have multiple registrations in several states or provinces in North America in order to undertake projects at several locations impose time and cost constraints on individual engineers and geologists, and their employers. Only limited reciprocal arrangements exist between Canada and the USA, in spite of the regulations embodied in the North American Free Trade Agreement (NAFTA). Similar trends within the European Union have led to new developments that promote European designations.

The concept of professional registration for geologists is still relatively young. Major constraints are the lack of public acceptance of the need for registration, the lack of 'official' legal standing, the objections of many geologists who see registration as restricting their mobility and freedom to conduct studies, objections by other professions, and competition among professional societies for authority to provide and supervise such registrations.

A Look to the Future

Geological engineering appears to be at a crossroads. Demand for geological engineering expertise to solve society's needs and desires for a better living environment points toward a bright future. Certainly, new and ever more challenging environmental issues will make the design and construction of new transportation and other facilities depend even more on an accurate prediction of geologic conditions. The increasingly sophisticated designs depend for their success on the involvement and acceptance of the geological engineer and engineering geologist.

Yet the entire capacity to educate and train fresh geological engineers is quite limited. The majority of the academic programmes in the USA have relatively small recruitment. The economic pressures facing many universities encourage the elimination of smaller 'specialist' or 'elitist' and high-cost programmes and departments. The establishment of geological engineering educational programmes beyond the USA has been, and continues to be limited because the same combination of technological and professional registration procedures that encouraged the establishment of geological engineering in the USA does not occur.

In western Europe and North America, the enrollment of students in engineering and science, especially geoscience, has been falling for several years. Many talented students are not selecting such 'tough' courses of study demanded by engineering and science fields. Topical areas perceived as narrow specialties, such as geological engineering, are apparently at a further disadvantage when attracting new recruits.

Geological engineering does not have the advocacy within the larger established professional societies to ensure its growth or even survival as a designated independent engineering specialization. Even the wider field of engineering geology practitioners, encompassing both geological engineers and engineering geologists, is facing a similar identity crisis. This is occurring in spite of expanding employment opportunities and the recognition of the need for such specialists by potential employers.

See Also

Engineering Geology: Codes of Practice; Natural and Anthropogenic Geohazards; Ground Water Monitoring at Solid Waste Landfills. **Environmental Geology**. **Famous Geologists:** Hall; Smith. **Geology, The Profession**. **Geotechnical Engineering**. **Rock Mechanics**. **Urban Geology**.

Further Reading

Kiersch GA (ed.) (1991) *The Heritage of Engineering Geology; The First Hundred Years*, Vol. 3, *Centennial Special*, p. 605. Boulder, Colorado, USA: Geological Society of America.

Legget RF (1962) *Geology and Engineering*, 2nd edn., p. 857. New York: McGraw-Hill.

Legget RF (1973) *Cities and Geology*, p. 578. New York: McGraw-Hill.

Paige S (ed.) (1950) *Application of Geology to Engineering Practice*. Berkey Vol., p. 327. Boulder, Colorado, USA: Geological Society of America.

Accreditation Board for Engineering and Technology, Inc.(ABET) – provides, operates and maintains an independent and objective accreditation system for applied science, computing, engineering, and technology education. http://www.abet.org/.

National Association of State Boards of Geology (ASBOG) – serves as a connective link among the individual state geologic registration licensing boards. http://www.asbog.org/.

National Council of Examiners for Engineering and Surveying (NCEES) – a national non-profit organization composed of engineering and land surveying licensing boards representing all U.S. states and territories. http://www.ncees.org/.

Canadian Council of Professional Engineers (CCPE) – provides links to all provincial/territorial engineering associations. http://www.ccpe.ca/.

Canadian Council of Professional Geoscientists (CCPG) – provides links to all provincial/territorial professional associations. http://www.ccpg.ca/.

Association of Engineering Geologists (AEG). http://www.aegweb.org/.

Geological Society of America (GSA). http://www.geosociety.org/.

International and European Societies.

International Association for Engineering Geology and the Environment (IAEG). http://www-cgi.ensmp.fr:88/.

Geological Society of London (GSL). http://www.geolsoc.org.uk/.

Institution of Civil Engineers (ICE). http://www.ice.org.uk/.

European Federation of Geologists (EFG). http://www.eurogeologists.de/.

European Federation of National Engineering Associations (FEANI). http://www.feani.org/.

GEOLOGICAL FIELD MAPPING

P Garrard, Imperial College London, London, UK

Introduction

Geological field mapping provides the fundamental scientific basis for most geological maps. It can range from large area reconnaissance to detailed mapping of areas only a few metres across, and sometimes has a specific aim such as mineral exploration. All-purpose, 'survey-type' mapping involves a geologist in the field examining rocks in their natural location, plotting data onto a base map, and recording details in a field notebook. Progressively, the field map displays the distribution of rocks and superficial deposits at the Earth's surface, their orientations, and the nature of the contacts between them. Fossils and rock samples may be collected for laboratory investigation. Analysis of structures indicates the tectonic development. From such information, it is possible to draw sections in any direction across the area, predicting geological relationships at depth, and to compile a report detailing the geological history. Other types of investigation, e.g., geophysical, geochemical, and remote sensing, can provide supplementary information.

National Geological Surveys publish district geological maps at scales of 1 : 50 000 or 1 : 25 000, based on field mapping typically performed at larger scale, for instance 1 : 10 000 in the case of the British Geological Survey (BGS). District maps hold a large amount of detail from routine fieldwork, plus data from boreholes, water wells, underground workings, and similar. Compilations give rise to regional and national maps, and the maps are also used by professionals requiring information on soils, economic deposits, land use, water supply, and hazard potential. Hence the field geologist needs to record all aspects of the geology.

Two points require emphasis. First, field investigations must be rigorous, making careful, comprehensive examination of the geology and recording the results accurately at the correct location. Inadequate or wrongly positioned data will result in incomplete or incorrect deductions. Second, a geological field map is partly interpretative. It contains 'factual' data, studied and measured in the field, but incomplete exposure means that the nature, position, or even existence of various boundaries depends on judgement by the geologist in the light of field observation, experience, and existing concepts. Further information may dictate the need for revision.

Basic equipment comprises a mapping board with plastic cover, base map, compass, clinometer, notebook, black and coloured pencils, sharpener, and eraser. Air photographs can be extremely useful. A hammer and hand lens enable fresh rock to be obtained and studied. A fine-line pen with waterproof black ink will be needed to ink-in the map and notebook, and it is helpful to record some map features in coloured ink.

Base Map

When topographical base maps are obtainable at different scales, the choice will be guided by the nature of the investigation. Scales of 1 : 100 000 to 1 : 50 000 are appropriate for regional exploration and reconnaissance. More detailed mapping is generally performed on scales of 1 : 25 000 to 1 : 5000. Very detailed work in connection with mines, quarries, and engineering sites may require scales of 1 : 1000 to 1 : 50, often available as company plans. The 1 : 10 000 Ordnance Survey sheets used by the BGS show contours at 5 or 10 m intervals, the positions of streams, buildings, fences, and roads, and a numbered grid. It is relatively easy to locate a position on the map, record it as a grid reference, and identify morphology by 'reading' the contours. Accurate contoured maps can be made from air photographs using specialized photogrammetric techniques.

Where suitable topographical maps have not been produced, have poor accuracy, or are unavailable because of political or military restriction, the geologist must find an alternative. Small-scale maps can be enlarged photographically, or scanned or digitized and redrawn by computer, but only have the detail of the original.

Satellite images, obtainable for all the Earth's surface, can be processed to give a coarse resolution base map suitable for reconnaissance work. However, the creation of contoured maps is restricted to satellites which 'see' in three dimensions, through overlapping images.

Air Photographs

The most commonly used air photographs are those made on black and white panchromatic film from level flight with the camera axis vertical. Approximate scales of contact prints are generally on the order of 1 : 8000 to 1 : 40 000. Typically, there is 60% photograph overlap along the line of flight and 30% overlap of adjacent lines, allowing the area to be viewed as a three-dimensional image through a stereoscope. As a result of spacing, the vertical scale is markedly exaggerated, enhancing topographical features that might not otherwise be obvious.

Air photographs are not maps. Only the centre (principal point) is viewed from directly above, comparable to a map. Away from the centre, the top of a vertical object appears displaced radially outwards relative to its base. The scale varies according to terrain height relative to flying height and to camera tilt.

Despite such limitations, air photographs can be immensely useful for locating position and tracing boundaries. Data recorded on overlays in the field are subsequently transferred to the base map. In three-dimensional view, photographs display relationships between terrain, drainage, and geology over a broader area than can be seen from a ground position. Variations in tone and texture, plus the dip of strata and patterns of fractures, can be interpreted to yield a photogeological map showing solid and superficial units and the size, shape, and orientation of the principal structures. Field investigation is essential to identify rock types and measure structures, but photo-interpretation is strongly recommended prior to fieldwork and each evening to give a rapid, detailed overview and allow optimum planning of traverses. Methods of interpretation are described in relevant books. Photogeological information is recorded on the base map in a colour that distinguishes it from field data.

Equipment

Map Board/Case

The simplest combination is an A4 clipboard or plywood, plastic, or aluminium sheet of about 33 cm × 23 cm, plus elastic bands to hold the field maps and a large polythene bag for weather protection. More complicated map cases may have a lid and pockets and slots to hold pencils, notebook, and other items. Field sheets can be held by spring clips, but as these affect a nearby compass it is better to use elastic bands. Cut the base map to suitably sized field sheets, say 18 cm × 25 cm, number all grid lines, and mount or photocopy onto thin card leaving a blank protective margin.

Pencils

Sharp graphite pencils of H or 2H hardness are recommended for map work because softer grades tend to smudge. Structural readings and map notes should be made in pencil in the field, not directly in ink. Pencil notes can be repositioned if their map space is needed, but ink errors are difficult to remove. At the end of the day's work, however, all pencilled data must be inked. Coloured pencils for recording lithologies should be good quality, thin-leaded, and waterproof. About 12 contrasting colours serve most purposes. Carry spare pencils, plus a sharpener and eraser.

Mapping Pens

Waterproof black and coloured inks (red, blue, and green) are used to make a permanent record of the day's work on the map and in the notebook. Barrel-type technical pens giving a line width of 0.15–0.2 mm are a common preference for maps and 0.2–0.3 mm pens for notebooks. Fine-line, waterproof, fibre- or ceramic-tipped pens of nominal 0.1 mm line size are an economical option for the colour pigments.

Compass and Clinometer

A compass is used to take bearings on distant objects to establish one's map position and to determine the orientation of structures. A clinometer measures the inclination of a planar or linear feature relative to a horizontal reference plane.

Several instruments combine both functions. A popular unit, the Silva Ranger Type 25TDCL, has a liquid-filled compass cell, graduated 0–360°, which can be rotated in its rectangular baseboard. The hinged lid has a sighting line and a mirror to view an object and the compass simultaneously. The transparent cell base has edge markings around half the circumference, with 0° in the centre and 90° at each end. When the ends are aligned with the baseboard edge and the instrument is held in a vertical plane, a centrally mounted, free-swinging arrow registers the angle of inclination. The cell base is also inscribed with parallels for alignment with the map grid and an arrow shape to register the position of the compass needle. An adjusting screw allows the arrow to be turned relative to the parallel lines, thus correcting for the angle between magnetic north and grid north. The use of such an instrument is illustrated in Figure 1. Readings taken from the rock are plotted immediately on the map. The orientation of distant or inaccessible surfaces can be found by moving to a position along strike (sharpest edge view), taking a compass reading, and then tilting the clinometer in mid-air to match the dip.

Hand Lens

A lens is essential and it is worth paying for optics which give a flat field and sharp undistorted image to the edge. A magnification of ×10 is the most useful and the lens should be worn on a cord around the neck so that it is constantly at hand to examine rock textures, mineral grains, and the like.

Hammer and Chisel

Geological hammers are used to break off a fresh surface of rock for examination or sampling, and are available in a range of sizes, shapes, and materials. For general use, a weight of 0.7–0.9 kg is suitable. Heavier ones (1.1–1.8 kg) are better suited for hard igneous and metamorphic rocks. One end of the head has a square face; the other may be chisel shaped, useful for layered rocks, or pick shaped, for prising cracks or digging into soil. Handles can be of wood, fibreglass, or steel.

Because of the hardened steel face, one must never use a hammer as a chisel and hit it with another, as metal shards can fly off and cause injury. Instead, use a soft steel chisel. Safety goggles should be worn when hammering or chiselling rocks.

Field Notebook

The general preference is for a pocket-sized notebook of around 12 cm × 20 cm. The paper must be of high quality, still usable after several cycles of soaking and drying, and securely bound between hard covers. Books designed for geologists are stocked by relevant dealers. Ordinary stationers' notebooks are not suitable. Notebooks can have an end or side hinge, and the pages may be plain, lined, gridded, or mixed. A 5 mm square grid is useful for recording measured stratigraphical sections and scaled plans.

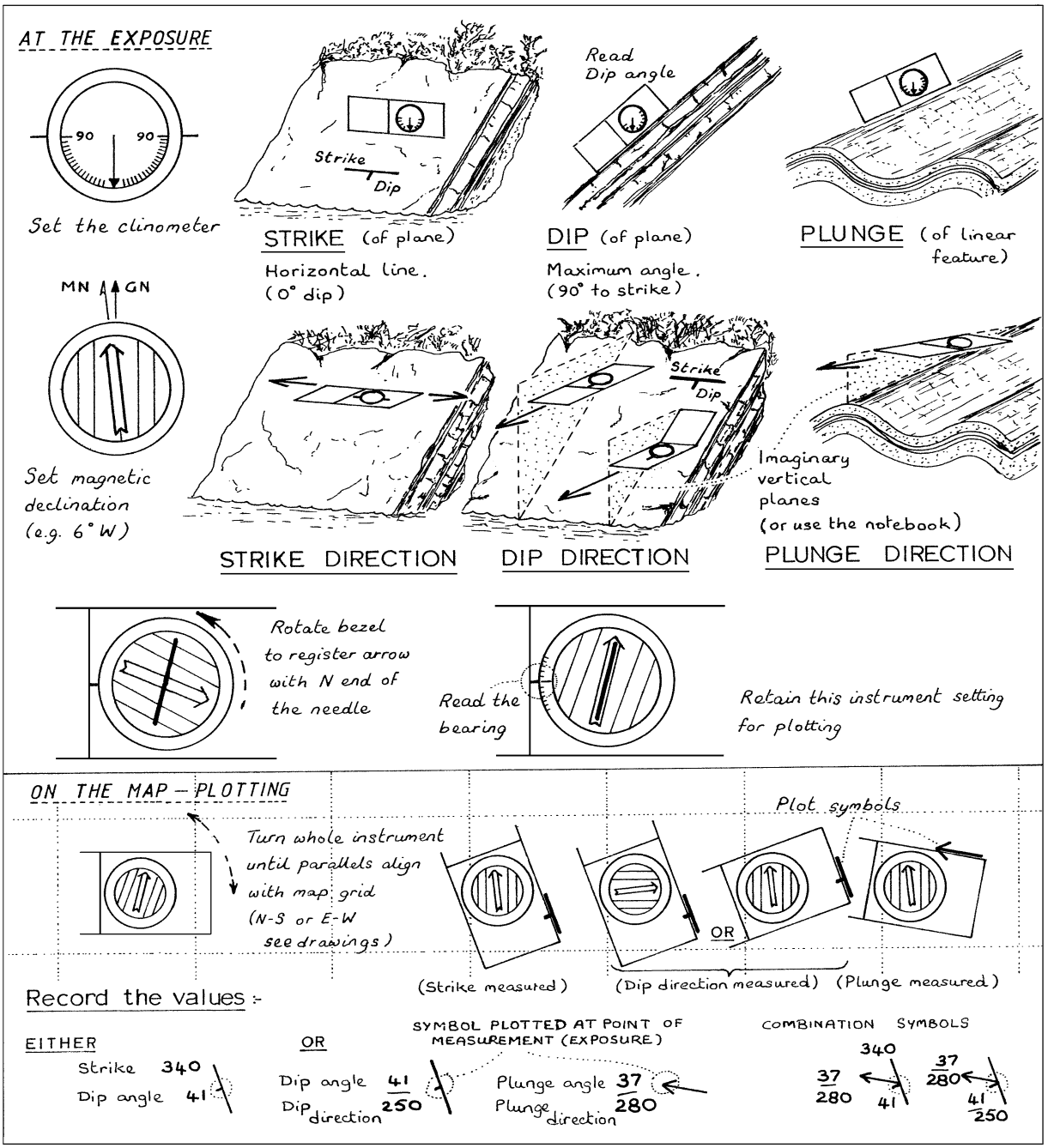


Figure 1 Use of compass–clinometer.

Other Equipment

For general field mapping, the following are often or occasionally used.

- Small dropper bottle containing 10% hydrochloric acid. A drop causes calcite to effervesce, but shows no reaction with dolomite.
- A 3 m steel measuring tape to determine bed thicknesses, and perhaps a 30 m linen tape for distances.
- A scale rule and protractor for measuring distances and angles on the map.
- A Global Positioning System (GPS) instrument. Although rather expensive, this is increasingly being seen as a standard field item for its ability to locate position to within about 10 m.
- An altimeter can assist positioning if the field map has reliable contours, but is less useful than GPS and subject to weather-related pressure changes.

- A pocket stereoscope is important if mapping involves the field use of air photographs.
- A stereographic projection net assists in understanding structural relationships.
- Binoculars allow the study of inaccessible rock faces and can help in locating exposures and routes.
- A camera, preferably with the ability to take close-up as well as distant shots.
- A penknife for conducting scratch tests, amongst other uses.
- Sampling bags.

Field Clothing

Clothing appropriate to the conditions allows the geologist to work comfortably throughout the day and not to be stressed by heat, cold, or wetness to the point at which concentration lapses and safety is at risk. Typically, a day rucksack is carried, large enough to take spare clothing, food, water, mapping equipment, safety items, reference literature, and collected specimens. In cold wet climates, the aim is to be warm and weatherproof by wearing items such as an undervest, shirt, fleece or sweater, and waterproof hooded anorak, plus a warm hat, scarf, and gloves. Trousers are best made of quick-drying, close-weave cotton, not jeans, with further protection from waterproof overtrousers. Well-cleated strong leather boots giving firm ankle support are recommended for most fieldwork, with wellington boots for wading activities.

Duvet clothing, warm padded footwear, and windproof outer garments are used in very cold dry climates. In hot dry conditions, protection against direct sun radiation is best given by a loose, long-sleeved cotton shirt, long trousers, and a wide-brimmed sun hat, but short trousers and sleeves are preferred by many. Corresponding footwear is lightweight whilst still providing support and grip.

Health and Safety

A doctor can advise on what inoculations, vaccinations, and other medical precautions are needed for work in different parts of the world. A first-aid kit should always be part of the field geologist's equipment. Items which come under the general heading of safety include goggles, a hard hat, torch, whistle, phone/radio, and survival bag. Brightly coloured field clothing will help a search party, should the need arise.

Mapping Preliminaries

Preparatory work before entering the field enables the survey to be conducted more efficiently. Administrative aspects include obtaining permits, establishing how to contact medical and emergency services, and arranging that others know your daily whereabouts. Relevant literature, maps, and borehole records will indicate what is known about the rocks and the nature of uncertainties. Air photograph interpretation and the production of a photogeological map will identify the type of terrain, the range of solid and superficial units, and the principal structures.

One or more days will be needed for reconnaissance before mapping starts to provide first-hand knowledge of the rock types, mappable formations, nature of exposure, and style of structures. Notes are made on access routes, areas of good and poor exposure, and parts that are inaccessible or hazardous. An important act is to make contact with local people to explain the work and confirm permission.

Following the field visit, decisions can be made on mapping methods. Often the most productive traverses are along streams which flow across the general strike and have cut down to bedrock. Parts of an area may be best mapped by tracing the outcrop of a marker unit. Broad area reconnaissance may dictate widely spaced traverses on parallel compass lines, whereas a mineral prospect might need enlarged scale mapping controlled by plane-tabling or compass and tape. Measured stratigraphical sections and key structural areas can be drawn at large scale in the notebook. The overall aim is to measure and record as many exposures as possible to cover the area within the allotted time, constantly remembering that the map and notebook must be easily read, understood, and followed by another geologist. Hence exposures must be precisely located, symbols drawn neatly and accurately in conventional forms, numbers legible, and notes comprehensible. Map and notebook must both distinguish between observation and inference. Various symbols, colours, and abbreviations will be used to depict the geology. Prepare a 'key' on the map to explain these, with formations in stratigraphical order.

Symbols

There is no universal set of mapping symbols, but published lists have broad similarity and an example set is given in Figure 2. Planar structures are generally depicted by a bar in the strike direction and a tick or other shape in the dip direction. Linear structures are shown as arrows, with various head, shaft, and tail ornament. Orientation measurements for lineations are given as an inclination angle (plunge) and its bearing (plunge direction) in the form 20/232. Two methods are in use for the orientation of planes. One

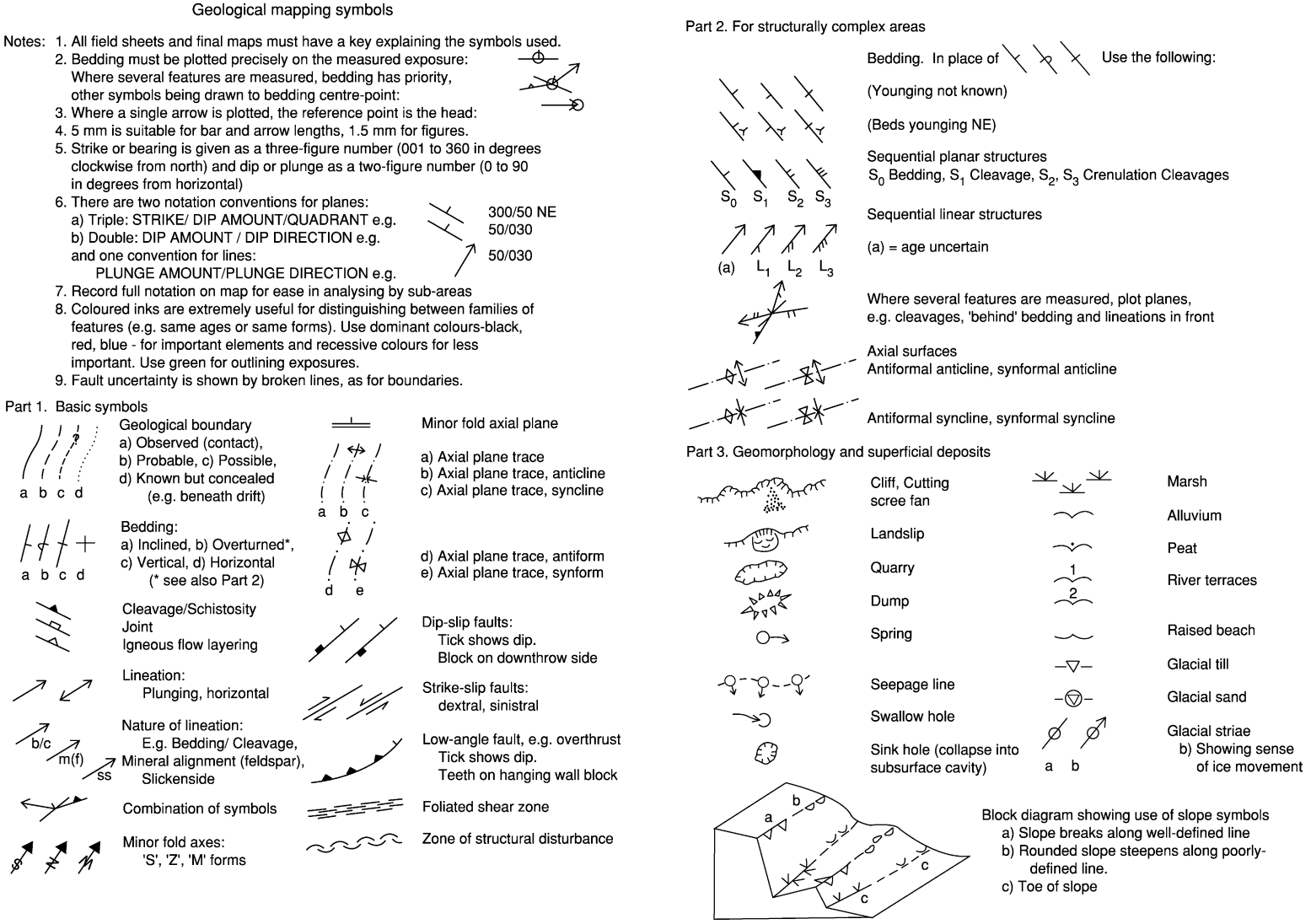
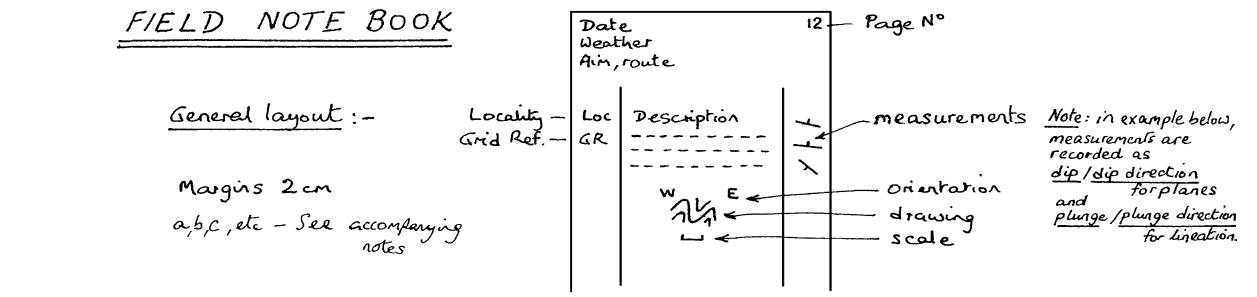


Figure 2 Geological mapping symbols.



8 July 1999 (a)

Dry, 5/8 cloud, slight breeze. Drizzle in afternoon.

Aim: Traverse downstream from 978475 for 2km, then E, from granite across cover rocks.

(b)

Loc 41
9782
4747

Exposure: low smooth slabs, 3m across, in banks & stream bed (c)

Granite (d): pink, med-crs.gr., equigranular (5mm). Pink felds. 55%, grey glassy qtz, 40% black biotite, 1-3mm, 5% (e).

Homogeneous. No obvious joints

Loc 42
9796
4748

Expos: low smooth surface 3×2m granite in sharp intrusive contact (f) with 1m width hornfels. Contin. expos. of granite from 41 (g).

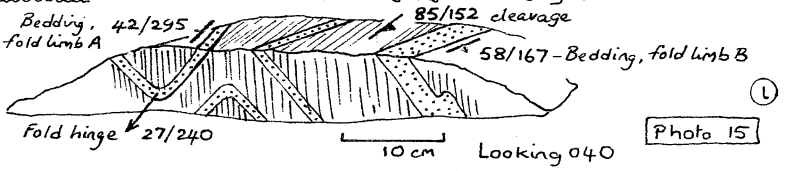
Granite: as at 41. No obvious change in grainsize (no chilled margin). Hornfels: dk grey fine gr. metasiltstn with qtz, prob. felds, biotite and 5-15% 1mm grey-black spots (? cordierite), the latter in 3-7cm bands (= bedding?). SPEC 23 (h) – hornfels.

Contact strikes 340, dip uncertain (?80° NE). Spotty banding: 50/174 (i)

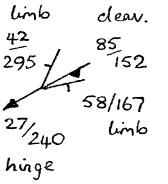
Loc 43
9820
4742

Expos: 3×3m on N bank, oppos. small trib. Patchy (20%) expos. of h'fels from 42. Alternating cleaved siltstns and thin ssts, folded. [NB. No spotting here, hence outside thermal aureole (j)]. Siltstones: 60%, in 4-10cm beds, med.grey, well cleaved.

Sandstones: 40% in 1-5cm beds, fine gr, greenish grey (chloritic?) (k).



(l) Photo 15



Loc. 44
9853
4751

Expos: 2m crag 5m long, 40m N of stream. Similar crags visible for c.100m to NW (m).

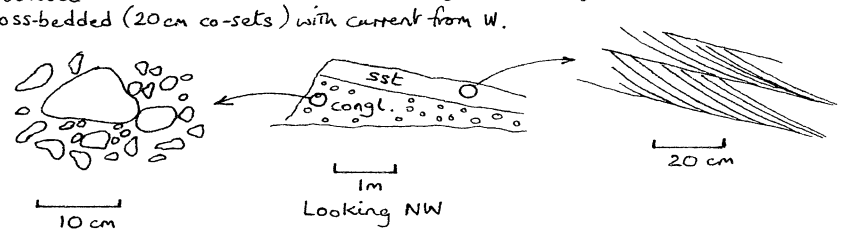
No expos. (turf and heather) betw. 43 and 44. Boundary perhaps along crag line? (n)

1.5m pebble/cobble conglom. overlain by 0.5m sst. Conglomerate? (o) 40:60 clast : matrix ratio.

Clasts: 40% pink med.gr.granite, sub-rounded, 2-10cm; 30% white vein quartz, sub-angular, 2-3cm; 15% grey h'fels, sub-ang, 1-2cm; 10% fine gr. greenish sst, rounded, 1cm; 5% siltstone, sub-ang., 1cm. Matrix: poorly sorted fine-crs.gr. sand (qtz + 20% felds)

[NB. Clasts derived locally. Presum. unconformity betw. 43 and 44].

Sandstone: abrupt contact with congl. Grey brown med.gr. quartz + 5% felds. Cross-bedded (20cm co-sets) with current from W.



15/050 Bedding

45/062 Foreset

(P)

Figure 3 Field notebook. (a) The daily heading gives the date, weather conditions, general aim, and route. (b) Number the localities consecutively. Do not restart each day. An eight-figure grid reference (eastings then northings) locates the position to 10 m. (c) Describe the nature of the exposure, particularly its size. A 1 m exposure of sandstone might be part of a shale formation, whereas a

Continued

records strike / dip angle / quadrant, as in 330/30 SW. The other gives dip angle/dip direction, as in 30/240. The latter is advised. It has fewer components, is less prone to error, and is in the same form as lineation data.

Use of Field Notebook

The notebook is the essential constant companion to the field map. While the map depicts the important readings, boundaries, and formations in spatial relationship, the notebook holds further readings, detailed descriptions, drawings, and much extra information. The order of use differs amongst geologists, but a common practice is to make a thorough examination of the exposure, record it in the notebook, and then plot the salient features on the map. An example of notebook layout and entries appropriate to a mapping survey is shown in Figure 3. Corresponding explanatory notes are also given, extended in a few cases to cover other geological situations. Should the field map be lost, it would be possible to reconstruct much of it from the notebook, although not boundaries drawn directly from field observation. Drawings should always be made in pencil, to be inked subsequently when the correct line has been achieved. Notes can be in pencil or ink.

Use of Field Map

Mapping involves placing a record on the map of: (1) the presence of rocks at the surface (exposures); (2) the type of rock(s) exposed; (3) the orientation of bedding, contacts, and tectonic structures; (4) the nature of superficial cover; and (5) landform features, such as river terraces. Information from various sources is combined to allow: (6) boundaries to be drawn on the map, thus giving the 'outcrops' of mappable units.

At each location, proceed as follows.

1. Determine the position on the map by GPS, by reference to nearby features and contours, or by resection using two or three compass bearing lines 60°–90° apart.
2. Outline the limits of the exposure.
3. Identify the rock type(s) and formation, and colour the exposure on the map.
4. Record the locality, grid reference, rock descriptions, measurements, drawings, etc., in the notebook.
5. If a formation contact is present, plot it on the map. Scan surrounding country for clues to extend the line.
6. Measure the orientation of bedding and plot the symbol on the map, precisely where the reading was taken. Record measurements in 1.5 mm figures.
7. Plot other features, such as cleavage, drawing the symbols to the centre of the bedding bar. In complex areas, plot the most important structures, leaving others in the notebook. No reading should be plotted in a position which was not the point of measurement.
8. Add locality number and relevant notes, particularly for key stratigraphical, fossil, and structural localities. Notes can be written on unused areas, such as the sea, and linked to the exposure by a dotted line.
9. Move to another locality, observing continually. If unexposed ground is thought to be of the same formation, shade it more lightly with the same colour.

If the next exposure is of an entirely different formation, there must be a boundary somewhere in the intervening ground. Whether it is sedimentary, igneous, or tectonic is commonly inferred from its geological setting. Its position is judged after assessing

30 m cliff of sandstone clearly belongs to a sandstone formation. (d) Underline key words. (Use the formation colour employed on the map.) (e) Give size, shape, colours, and proportions of the different rock constituents, not just a rock name or mineral list. (f) Contacts demand particular attention. Is the contact sedimentary, igneous, or tectonic, sharp, gradational, sheared, discordant, etc.? What are the age relationships? With igneous contacts look for features such as chilling, baking, xenoliths, and veins. (g) Note the nature of ground between localities, e.g., continuous, patchy or no exposure, type of vegetation. (h) Record specimen numbers. Collect typical rocks: abnormal ones are additional. Make notes if a specimen has been oriented (marked in the field such that it can be oriented identically in the laboratory). (i) Plotting symbols and figures in the margin as they will appear on the map provides a check against possible map errors, allows easy data retrieval, and provides a visual record of changes in orientation. (j) Include deductions or implications arising from the geological observations. (k) Proportions and thicknesses are more informative than descriptions such as 'thinly bedded sandstones and siltstones'. If fossils are present, record their identity, size, state, distribution, and abundance. (l) In deformed areas, record all planar and linear elements (bedding, cleavages, axial planes, fold axes, lineations). Always draw folds, because readings alone can never convey shape. Draw accurately, label the rock types, and avoid ornamental lines that might be mistaken for geological fabric. Incorporate a scale and an orientation (N arrow for plan views, viewing direction for sections). It helps to show readings on the drawing. (m) Note geological features in view, as well as those at the exposure. (n) Record any evidence, or lack of it, for a deduced boundary. (o) Conglomerates indicate which rocks were exposed to erosion at a particular time. The proportions and shapes of the constituents provide evidence on transport and depositional environment. (p) The measurement of bedding and truncated foresets shows younging of the sequence and gives information to calculate the current flow vector. Other sedimentary structures have similar potential.

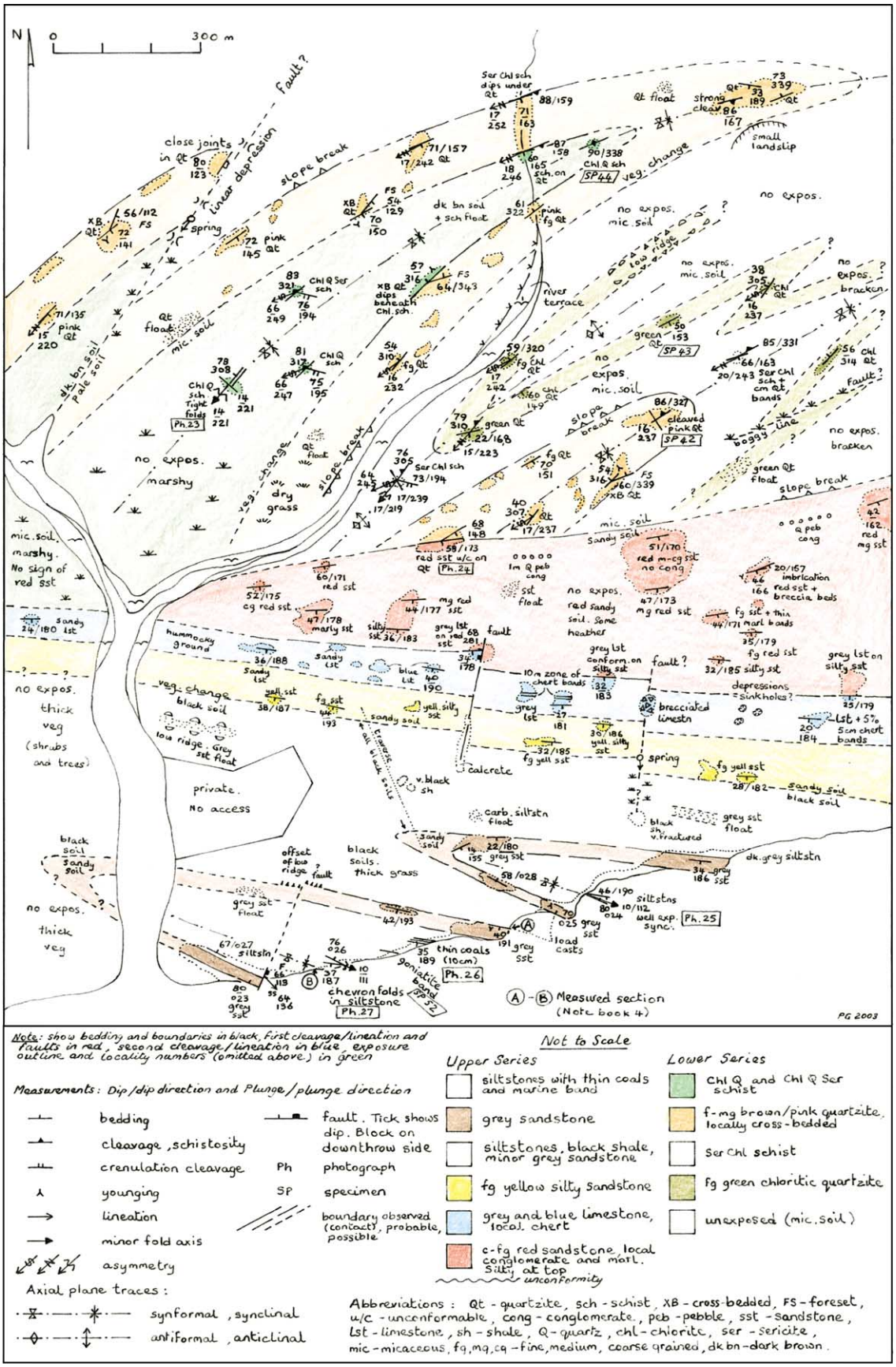


Figure 4 Example of field map entries. Map grid, contours, and locality numbers have been omitted for clarity. The map symbols would be in coloured ink. Exposures would be outlined in a continuous, not dotted, green line.

field evidence, such as a change in soil composition, soil colour, slope change, topographical feature, spring line, and change in vegetation, and recording these on the map. Study of air photographs may help. A boundary is pencilled on the map in a position which fits the evidence best, drawn as a full line if well constrained, or long or short dashes for decreasing confidence.

It is clear that geological field mapping means making a map in the field whilst the evidence is in sight, not placing locality numbers on the map to plot the data later and draw arbitrary lines between rock changes. An example of field map entries is given in Figure 4.

At the end of the field day, the map and notebook are inked in to make a permanent record, using coloured ink on the map as appropriate (Figure 4). Ink the notebook drawings accurately to preserve the lines made whilst the object was in view. Analyse structural data on a stereonet and construct trial cross-sections. Never leave a field area without drawing at least one geological cross-section. Specimens collected during the day are given permanent labels and laid out for further reference. It is helpful at this stage to summarize the day's findings, place it in context, and make notes for the final report.

Superficial Deposits

Some areas are deeply weathered and covered by residual soils or regolith through which only the most resistant rocks emerge at the surface. Others are blanketed by transported materials, such as wind-blown sands or glacial deposits. For residual soils, information can be obtained by a study of auger samples, recording parameters such as soil colour, composition, and consistency. Look also for fragments ('float') brought to the surface by fallen trees, ploughing, or burrowing animals. Air photographs can prove to be useful in indicating boundaries, particularly where vegetation is well controlled by soil type and has been little disturbed by human activity.

For transported materials, the geologist aims to outline the deposit and classify the type. The 'drift' maps of the BGS, for example, show various kinds of gravel, sand, and clay deposits formed from standing, moving, or melting ice, plus river terraces, fluvial and estuarine alluvium, peat, and others. Some can be clearly seen and mapped from air photographs or a high vantage point. Relevant observations are placed on the field map by a selection of colours, overprint symbols, and notes. Areas of slope failure should be marked.

Ancient soils are not commonly preserved in the stratigraphical column but may have left traces, such as weathered and reddened surfaces; here, a knowledge of soil processes helps in assessing the palaeoclimate.

See Also

Engineering Geology: Geological Maps; Natural and Anthropogenic Geohazards; Site and Ground Investigation; Ground Water Monitoring at Solid Waste Landfills. **Environmental Geology**. **Geochemical Exploration**. **Geological Maps and Their Interpretation**. **Geological Surveys**. **Geomorphology**. **Remote Sensing:** GIS. **Seismic Surveys**. **Soil Mechanics**. **Soils:** Modern.

Further Reading

Allum JAE (1966) *Photogeology and Regional Mapping*. Oxford: Pergamon.

Barnes JW (1991) *Basic Geological Mapping*, 2nd edn. *Geological Society of London Handbook*, Milton Keynes: Open University Press.

Boulter CA (1989) *Four Dimensional Analysis of Geological Maps*. Chichester: Wiley.

Compton RR (1985) *Geology in the Field*. New York: Wiley.

Lattman LH and Ray RG (1965) *Aerial Photographs in Field Geology*. New York: Holt, Rinehart and Winston.

Lillesand TM and Kiefer RW (1979) *Remote Sensing and Image Interpretation*. New York: Wiley.

Maltman A (1990) *Geological Maps: An Introduction*. Milton Keynes: Open University Press.

McClay KR (1987) *The Mapping of Geological Structures*. *Geological Society of London Handbook*. Milton Keynes: Open University Press.

Moseley F (1981) *Methods in Field Geology*. Oxford/San Francisco: Freeman.

Ramsay JG and Huber MI (1987) *The Techniques of Modern Structural Geology*, vol. 2, *Folds and Fractures*. London: Academic Press.

Ray RG (1960) *Aerial Photographs in Geologic Interpretation and Mapping*. *United States Geological Survey Professional Paper 373*. Washington: US Government Printing Office.

Tucker ME (1982) *The Field Description of Sedimentary Rocks*. *Geological Society of London Handbook*. Milton Keynes: Open University Press.

GEOLOGICAL MAPS AND THEIR INTERPRETATION

A Maltman, University of Wales, Aberystwyth, UK

Introduction

Geological maps at their simplest are portrayals of the distribution of geological materials at the Earth's surface (*see* **Geological Field Mapping**). Any covering, say of agricultural soil or concrete, is ignored. Such a picture of areal distributions is itself highly useful, but the real power of geological maps arises from the further interpretations that can be made from them – into the third-dimension below the ground and back into geological time. Because of this breadth of possibilities, geological maps are a fundamental tool of both academic and applied geologists, and the governments of most developed countries underwrite a national body responsible for the geological surveying of the nation. Indeed, geological maps have been referred to as 'the visual language of geologists'. Today, yet new possibilities are blossoming as GIS (*see* **Remote Sensing:** GIS), spatial databases, and digital imagery are applied to geological map methods. How best to harness all this tremendous potential is the subject of much current dialogue. Fundamental map principles, however, are unchanging, and it is these that are summarized here.

The Nature of Geological Maps

The distributions of the different kinds of rocks are plotted onto some kind of topographic base image, traditionally a topographic map. Scales vary from very small, such as 1:1 000 000 or less (which may summarize the geology of entire regions), through medium scales such as 1:100 000 to 1: 25 000 (showing particular areas in reasonable detail), to large-scale maps at scales of 1:500 or even greater (typically used for specific sites of some commercial activity). Small-scale maps commonly have to neglect topographic relief, but at larger scales topographic contours can be shown, and the interaction between the physiography of the land surface and the underlying geology can be important in three-dimensional interpretations. The distribution plots are increasingly being linked into Digital Elevation Models, to aid visualisation of the interplay between geology and topographic relief.

The areal extents of the various materials are shown by different ornaments in the case of black and white maps, or by different colours in the case of most published maps, sometimes with additional letter or number symbols to aid distinction. The geological ornaments/colours and symbols are explained in an accompanying key or legend. While in some cases the Earth materials in a given area fall naturally into groups that are convenient for depiction at a particular map scale, in many cases the surveyor has to judge how best to make appropriate sub-divisions. Small-scale maps commonly portray divisions based on the geological period in which the material formed; some more specialized geological maps may show divisions according to the fossils contained in deposits, or some other particular characteristic.

The majority of geological maps, however, are litho-stratigraphic: the divisions are based on the type of Earth material and their stratigraphic position. The various divisions of material types are, conventionally, arranged in the map key in order of their genesis, upwards from oldest to youngest. The actual geological ages may or may not be added. On medium- and larger-scale maps, each division is commonly referred to as a map 'unit' or 'formation'. It is important to understand that although a particular formation may be coloured boldly on the map, and given some imposing name, it has been defined subjectively by the surveyor. It may be quite indistinct on the ground; another worker may have divided the units differently. It is also the convention to make a major distinction between solid bedrock and any overlying, unlithified, geological materials. Some geological maps ignore the latter, except where they are especially significant, and are thus referred to as bedrock or – in the UK – solid maps. On the other hand, maps that emphasize them are variously referred to as surficial, superficial, Quaternary or – again in some UK usage – drift maps.

Many published geological maps, which by definition depict things in the horizontal plane of the Earth's surface, are accompanied by geological cross-sections. These portray the arrangement of the material in the vertical. It is this complementarity of maps and sections that underscores the three-dimensional aspects inherent in geological maps, a topic expanded in the following section.

Mention was made above of the subjectivity of the units into which the Earth materials are divided for the purpose of a map. Also intrinsic in a geological map is the interpretive nature of the distributions shown. Unlike almost all other kinds of maps, much of the information shown is interpretive. Except in

the most arid parts of the land surface, the covering of soil, etc., means that the surveyor has had to judge the areal extent of the underlying Earth materials, from scattered exposures, in cliffs, river banks, road cuttings and the like, and other, indirect, evidence. Geological maps, therefore, are continually being refined as additional evidence becomes available.

The Interpretation of Geological Maps

Many of the uses of geological maps involve dealing with geological information in three dimensions. New IT techniques are becoming available to facilitate this but much work still relies on interpreting the third-dimension from the flat map – the sheet of paper or computer screen. Sedimentary rocks, being arranged in approximately tabular strata, lend themselves to this kind of treatment. The strata comprise a three-dimensional configuration, originally a horizontal succession but in the case of marine deposits that have been uplifted to form land – the most widespread situation – now in some combination of horizontal, tilted, folded, and faulted layers. The present-day land surface is an eroded slice, roughly horizontal, through this array. Properly, therefore, a geological map is a plot of the intersections or outcrops of the strata with the present land surface, projected onto the horizontal plane of the map. Note incidentally, in this context, the geological terms exposure and outcrop differ in meaning. An outcrop of a formation, i.e., where it intersects with the land surface, and hence what is depicted on a geological map, may or may not actually be exposed, i.e., lacking modern cover and be directly visible. From the patterns made by the outcrops on the map, a trained geologist can infer the three-dimensional configuration of the formations, that is, how they were arranged above the land surface before their erosion and, more importantly, how they are disposed below the ground, in the 'subsurface'. The geologist can do this in a quantitative or a non-quantitative way, the latter by simply assessing the geological patterns on a map and mentally visualizing – perhaps while sketching rough cross-sections – the spatial arrangements. Some examples follow.

Visual Assessment

A sequence of map formations passing progressively from older to younger implies: (i) that the succession has been tilted (if they were still horizontal only the uppermost division would outcrop), and (ii) the direction of dip (Figure 1). Even with such a preliminary interpretation it becomes possible to gauge what lies below the land surface in various places, apart from those sites where the lowest, oldest known formation is outcropping. This simple principle, that

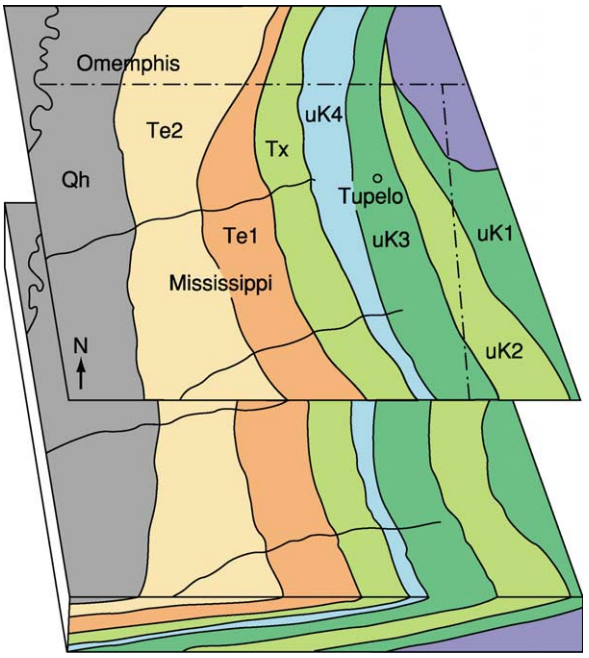


Figure 1 Block diagram and map to show how formations dip towards the direction in which they become younger (ignoring, as may be reasonable on a small-scale map, topographic relief). In this example, the geological map shows that a journey westwards across the State of Mississippi would encounter progressively younger formations. The simplest and much the most likely interpretation is that the formations are dipping westwards, as evident from the front panel of the block diagram. uK1, uK2, uK3, and uK4 are progressively younger Cretaceous formations, Tx, Te1, and Te2 are progressively younger Tertiary formations, and Qh is the youngest formation, of Quaternary age. Air photographs and satellite images may be helpful; geological maps of the sea-floor commonly involve seismic surveying.

outcropping formations pass from older to younger in the direction towards which they are dipping, needs care with larger-scale (usually contoured) maps of rugged country, where the topographic relief may interfere with and complicate the outcrop patterns. However, further simple principles can be applied in such cases. A dissected landscape developed in horizontal formations can give what at first glance is a bewildering geological map, but the outcropping formations, discerned by their boundary traces, will be exactly parallel to the topographic contours (because they, too, of necessity are horizontal; Figure 2). The outcrop patterns progressively deviate from this parallelism as the dip angle increases until, at the other extreme, vertical formations outcrop without any influence of topography. Their outcrops cross topographic contour lines to give straight traces. The more general case of dipping formations is best interpreted by assessing the direction and shapes of the V-shape outcrops made where they intersect with topographic valleys (Figures 3 and 4). Note that these are geometric principles arising from the intersections

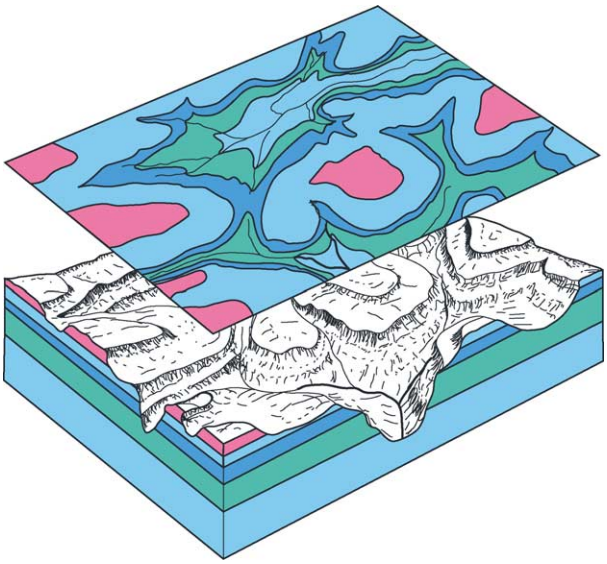


Figure 2 Block diagram and map to show how outcrop traces of horizontal formations parallel topographic contours (as they must also be horizontal). In the example shown here, the geological map is deceptively intricate. The outcrops are irregular, despite the simple, horizontal arrangements of the strata, because the landscape is highly dissected.

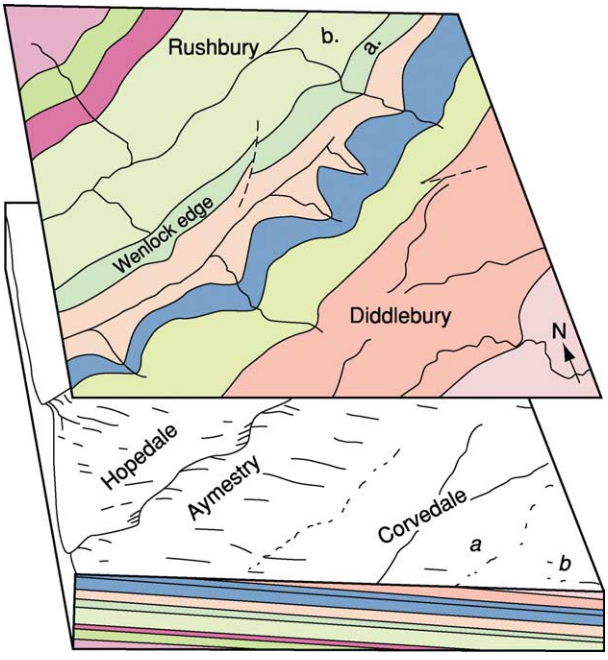


Figure 3 Block diagram and map to show the V-shape made by the outcrop of a dipping formation crossing a valley. The V in map-view points in the direction of dip, here towards the southeast. Note also the influence of land-slope on outcrop width. For example, the outcrop of formation *a*, at the steep escarpment of Wenlock Edge, is much narrower than that of formation *b*, at the flat land of Hopedale, even though they have similar thicknesses.

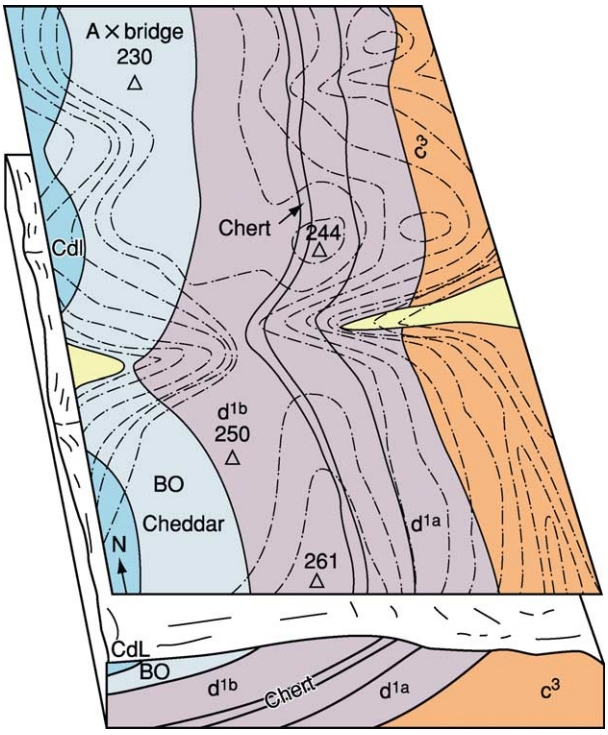


Figure 4 Block diagram and map to show how outcrops cross valleys with a more open shape where the formation dip is greater. As in **Figure 3**, the V in map view points in the direction of dip, here towards the W. The outcrop of more steeply dipping formations such as c^3 makes a more open V than move gently dipping units such as BO. The superficial deposits (yellow) are approximate horizontal and make the tightest V shape, parallel to topographic contours.

of surfaces – any geological surfaces, whether or not they are sedimentary strata. Thus geological faults (*see* **Tectonics:** Faults) commonly appear on maps as more or less straight traces, not because they are faults but because fault surfaces are commonly very steeply dipping or vertical. Gently dipping thrust faults, on the other hand, will give traces that are close to parallelism with the topographic contours (**Figure 5**).

The relative widths of outcrops on a map of a given scale reflect the thicknesses of the formations. Hence the outcrop width will vary if the material was deposited in different thicknesses at different places. Other things being equal, thicker units give wider outcrops. On larger-scale maps, however, the dip angle of the formation and the topographic slope will further influence the outcrop width. In general, steeper dips give narrower outcrops, as do steeper topographic gradients (**Figure 3**).

Symmetrical, roughly mirror-image repetitions of outcrops indicate folding (*see* **Tectonics:** Folding). Older formations symmetrically flanking younger ones imply that the strata have been downwarped

into the structure known as a syncline; younger formations flanking older indicate an upwarp or anticline (Figure 5). In addition, the shape of the symmetrical pattern contains information on the three-dimensional arrangement of the fold. Outcrop sequences that repeat in a translation pattern rather than a mirror-image reflection (Figure 5) imply the presence of a fault. The effects of faults on outcrop patterns are diverse, depending on the nature of the fault and its orientation with respect to the displaced strata. On published maps, faults are normally marked by a particular symbol – commonly a heavier line – indicated in the map key. The actual nature of the fault displacement is commonly impossible to interpret from a map, because the same outcrop patterns can be generated by different kinds of fault motion. Additional, often field-based, evidence is normally needed. Unconformities are recognised on maps by the absence of strata representing a significant length of geological time, and also by discordant outcrop patterns in the case of an angular unconformity (*see* **Unconformities**). The latter is important in three-dimensional map interpretations because the geometric arrangements will differ above and below the surface of unconformity.

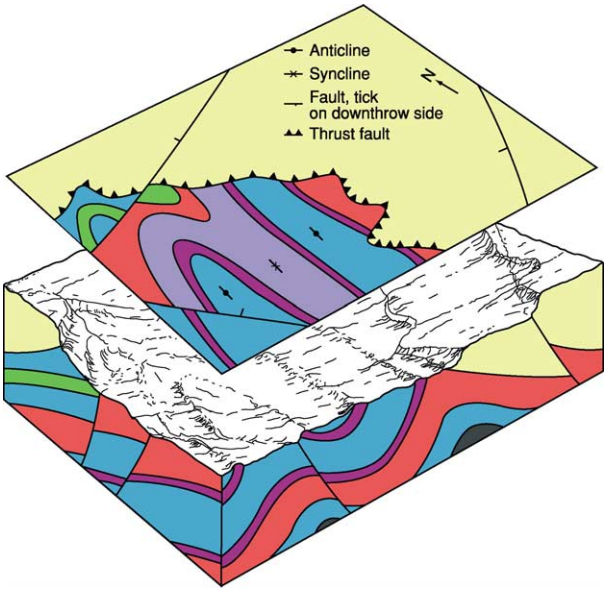


Figure 5 Block diagram and map to show outcrop patterns typical of some structures. The thrust fault is gently dipping and hence its trace is closely influenced by topography; steeply dipping faults crop out with relatively straight traces. The translatory repetition of outcrops is a typical effect of faults. The mirror-image repetition of outcrops is characteristic of folds.

Quantitative Treatments

In practical applications it is commonly not enough just to visualize or sketch the subsurface arrangements; the actual subsurface depths of particular formations and their thicknesses are often required. Quantitative answers to such questions are usually arrived at through some combination of three methods. First, a scaled geological cross-section can be constructed through the site of interest and the required data measured directly from the diagram. Larger-scale maps commonly provide symbols indicating the dip angles measured at exposure by the surveyor, and extrapolations of surfaces to depth are based on these values. However, with cross-sections there can be construction inaccuracies or even mistakes – irrespective of whether the work is done manually or by computer – though this pictorial approach has the advantage of making errors more readily apparent. A second approach utilises trigonometric calculations (Figure 6). Both these approaches rely on assumptions, such as the dip angle as measured at exposure being maintained at depth, and they

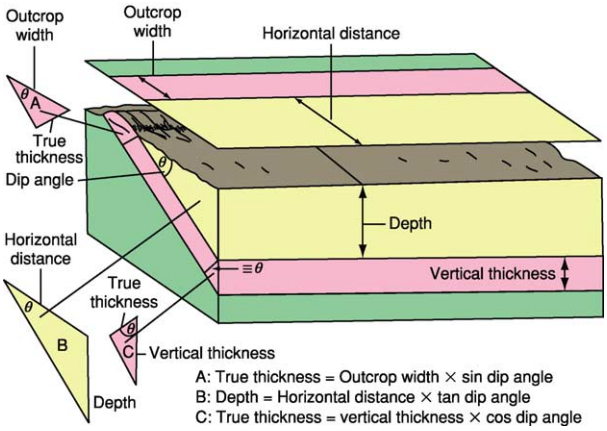


Figure 6 Schematic block diagram and map to illustrate how the thickness and subsurface depth of dipping formations can be calculate trigonometrically from map information.

only provide the quantitative information along the specified line, in the case of a cross-section, or, with the trigonometric approach, at the specified point. Computer-generated visualisations can be deceptive in their visual attractiveness: as with traditional methods they involve assumptions and subjective interpretations (and simplified algorithms).

Much more powerful is a third approach, using structure contours (Figure 7). Exactly analogous to topographic contours of the land surface, structure contours portray quantitatively the form of some specified surface below the ground. The surface might be the upper boundary of a formation of particular interest, perhaps an aquifer or a mineral-bearing horizon; it could be an unconformity or a fault surface. The contour values can represent the depth below ground of a surface or, more usefully where the land is not flat, the elevation of the geological surface, with respect to sea-level or some local datum. The numerical difference in altitudes between topographic contours for the land surface and the structure contours gives the depth of the contoured geological surface, at any desired location in the map area. Values for locations intermediate between contour lines can be interpolated. The numerical difference between contours for the upper and lower boundaries of a formation gives the vertical thickness of the material. For dipping formations this value will differ from the true thickness, at right-angles to the bounding surfaces, but the trigonometric corrections are straightforward. Contours that directly portray thicknesses are known as isopachytes.

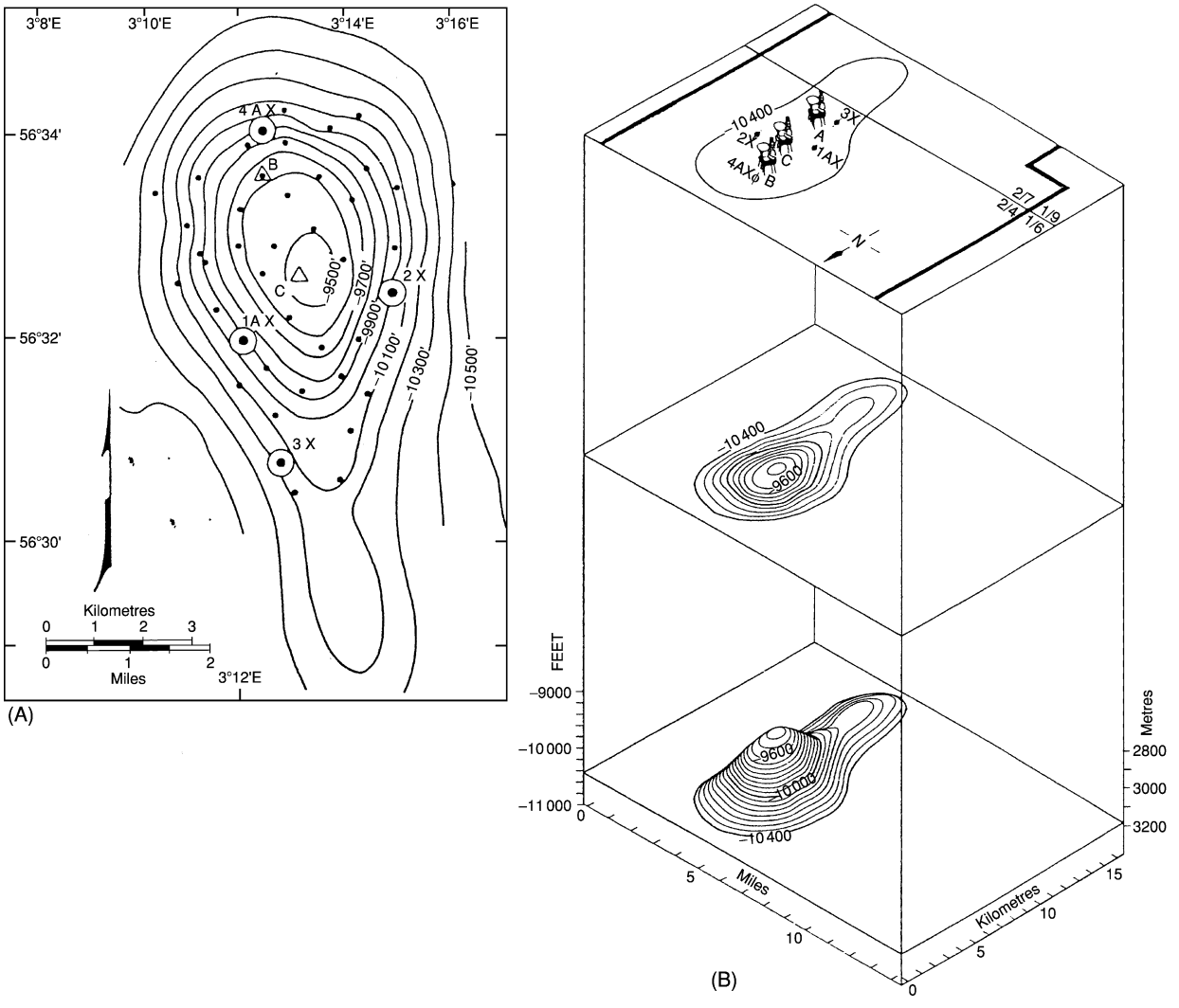


Figure 7 An example of a structure contour map: the Ekofisk oilfield, North Sea. (A) Structure contour map of the top of the oil-bearing formation. (B) Oblique view of the form of the contoured surface (bottom level of drawing) and structure contours (drawn at an arbitrary level) and the sea-bed (top level). Reproduced with modification from van der Bark and Thomas, 1980, American Association of Petroleum Geologists; used by permission.

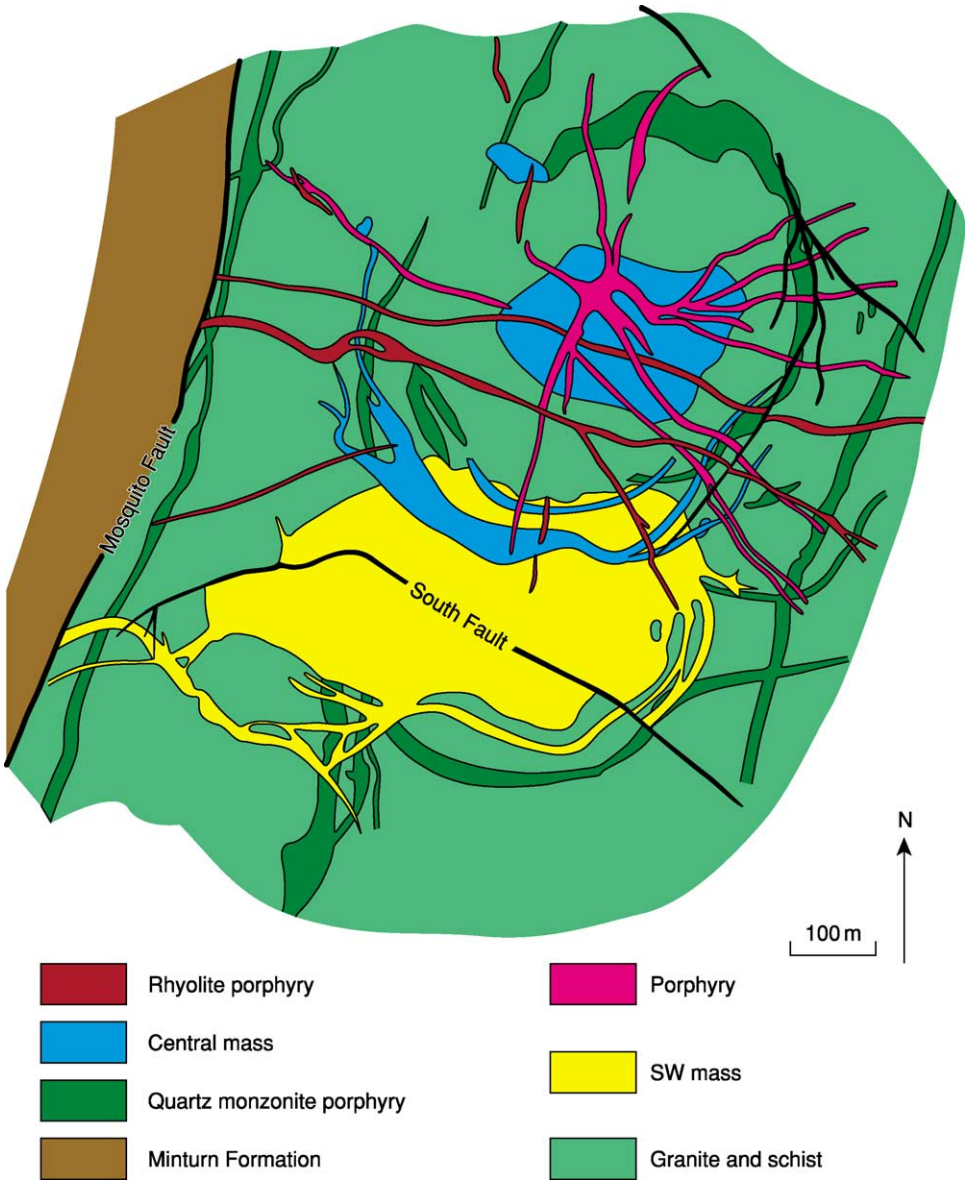


Figure 8 Map of the Climax molybdenum mine, Colorado, to illustrate the use of cross-cutting relationships in determining geological histories from maps. The Mosquito Fault displaces all the rock units and therefore appears to represent the youngest geological event. The dykes of rhyolite porphyry represent the youngest rocks E of the Mosquito Fault as they cross-cut all the units there; their relationship with the Minturn Formation, only seen to the W of the fault, is unclear from the map. Further cross-cutting relationships indicate that the rocks are progressively older in the following sequence: porphyry, the ''central mass'' of the Climax igneous complex, the SW mass, and quartz monzonite porphyry. The latter two units are displaced by the South Fault, which must therefore be younger than the time of their intrusion. All the igneous rocks are intruded into granite and schist, which is therefore the oldest unit of these units. Because no igneous rocks are shown in the contiguous Minturn Formation, interpretation of the map suggests it is the youngest unit. (In fact the Minturn Formation does contain related igneous rocks, and these have been dated as Tertiary. The Minturn Formation is Pennsylvanian (Upper Carboniferous) and the granite/schist host rocks are Precambrian in age.) Adapted from a map by Wallace, Dahl and others, Graton-Sales Volume, AIMME, 1968, used by permission.

Of course, the structure contours themselves have to be interpreted from control points such as boreholes or any sites where the formation of interest is exposed (at a known altitude), and their reliability varies with the density of control points available. If such control data can be derived from a geological map (e.g., by noting where outcrop traces cross topographic contours) it is possible for the map reader to construct structure contours for the surface of interest. A minimum of three points is needed. The construction is common on maps produced for specialised purposes – depth to bedrock, for example,

in engineering applications (*see* **Engineering Geology:** Geological Maps), or gravel thickness in the assessment of aggregates – but some general geological maps include structure contours for selected underground surfaces, to give a more quantitative picture of the subsurface arrangements. Developed oilfields or mineral deposits are likely to have a high density of well information and hence can yield structure contours of great accuracy.

The principles outlined above, utilising geometric surfaces, apply best to tabular sedimentary materials rather than other forms of rocks. Even so, extrapolations of depths and thicknesses of sediments have to be handled cautiously where the values are known to vary laterally, and this is particularly true of superficial sedimentary deposits. Because many such materials formed on land, where the depositional processes are much more localized than submarine sedimentation and where they are soon vulnerable to erosion, they characteristically lack lateral persistence. Igneous materials of tabular form, such as lava flows, sills, and dykes, can be interpreted using the basic geometric ideas, as can tabular mineral deposits. Less regular bodies will be difficult to interpret geometrically. Metamorphism does not itself affect application of the three-dimensional principles, but because it is commonly accompanied by substantial deformation, the configurations of metamorphic rocks can be complicated and more difficult to interpret from maps.

Geological Histories

In addition to the sub-surface extrapolations outlined above, another important interpretive aspect of geological maps is the reconstructing of geological histories. From the relative ages of formations normally indicated in the key, together with principles such as that a feature cross-cutting another must be the younger of the two, it is possible to erect a sequence of the geological events in the history of an area (Figure 8). By then linking in the lithologies and the conditions under which they must have formed, it becomes possible to reconstruct past geological circumstances. In the case of sedimentary rocks, for example, past depositional environments and their evolution through time can be deduced. It may even be possible to create a geological map for the land surface as it was at some past time – what is properly called a palaeogeological map. Such historical aspects are not only of academic interest but find applications in the petroleum industry, for example, where the linking of geological with palaeogeological maps, together with cross-sections, structure contour and isopachyte maps, is routinely used in hydrocarbon exploration and extraction. Such manipulations are these days largely carried out by computers, but a geologist familiar with the principles of geological maps still has to evaluate the output, and to be alert to shortcomings of the software and any consequent unnatural results.

Applications of Geological Maps

In ways mentioned above, geological maps find applications in different spheres of academic and commercial geology, and a host of related disciplines. Geological maps are relevant, to mention just a few examples, in civil engineering, hydrogeology, and archaeology. They help understand the landscape around us, both physical aspects such as landforms, slope instabilities, and building materials, and more human aspects, such as the soils derived from the bedrock and hence the crops that are viable, and settlement characteristics – in some cases even the place names. A recent development is the rise of geological maps for environmental planning, in which the geological aspects of an area relevant to, say, subsidence problems, landfill sites, contaminated ground, and radon concentrations are highlighted. Some national surveys (*see* **Geological Surveys**) now produce folios of such maps, each of which emphasises a particular environmental parameter, in an effort to foster sensible forward planning. Finding such wide usage, geological maps are so much more than mere plots of areal distributions.

See Also

Engineering Geology: Geological Maps. **Geological Field Mapping**. **Geological Surveys**. **Remote Sensing:** GIS. **Tectonics:** Faults; Folding. **Unconformities**.

Further Reading

Maltman A (1998) *Geological Maps: an Introduction*, 2nd edn. Chichester: John Wiley and Sons.

Soller DR (ed.) (2001) *Digital Mapping Techniques '01 – Workshop Proceedings*. US Geological Survey Open-File Report 01-223. http://pubs.usgs.gov/of/2001/of01-223/

United States Geological Survey (2002) *Three-Dimensional Geologic Maps and Visualization*. http://3d.wr.usgs.gov/docs/wgmt/3d/pp02.html

GEOLOGICAL SOCIETIES

G L Herries Davies, University of Dublin, Dublin, Ireland

Introduction

The development of modern geology has brought into being many types of geological institutions, ranging in character from geological surveys and mineral exploration companies to geological museums and departments of geology within seats of higher learning. But among all these types of institutions, a very special and significant niche is occupied by the world's geological societies. The seventeenth and eighteenth centuries saw the establishment of many a society dedicated to the advancement of what had been termed 'the New Learning'. Of these, the most famed was – and is – the Royal Society of London for the Promotion of Natural Knowledge, which received its charter from King Charles II on 15 July 1662. Phenomena such as rocks, minerals, fossils, landforms, and Earth history were all subjects that engaged the eager attention of the devotees of the New Learning. In the Royal Society's *Philosophical Transactions* number 52 (17 October 1669) is an account of minerals found following an eruption at Mount Etna, and in number 76 of the same periodical (22 October 1671), Martin Lister (1638?–1712) writes from York on the subject of local fossils. Similarly, the *Histoire de l'Académie Royale des Sciences* for 1771 (published in 1774) contains a paper on the volcanoes and basalts of the Auvergne by Nicolas Desmarest (1725–1815). In 1788, the Royal Society of Edinburgh (founded 1783) published in its *Transactions* the famed 'Theory of the Earth; or an Investigation of the Laws Observable in the Composition, Dissolution, and Restoration of Land upon the Globe', by James Hutton (1726–97). In 1799, the American Philosophical Society Held at Philadelphia for Promoting Useful Knowledge (founded 1769) published within its own *Transactions* a paper on fossil bones from Virginia by its president, Thomas Jefferson (1743–1826), who shortly was to become the third president of the United States.

Similar eighteenth-century societies made various other types of contributions to the development of the Earth sciences. Within the British Isles, the Dublin Society for Improving Husbandry, Manufactures, and other Useful Arts and Sciences (founded 1731; it assumed the title 'Royal' in 1820), between 1801 and 1832, published statistical surveys of 23 of Ireland's 32 counties. Several of these surveys, especially the 1802 survey of County Kilkenny by William Tighe (1766–1816), contain comprehensive essays in regional geology. In 1808, the same society offered a premium of £200 (it was never claimed) for a geological map of County Dublin, and in 1809, the society commissioned a geological survey of the Leinster coalfields. In London, the Society for the Encouragement of Arts, Manufactures, and Commerce (founded 1754; styled the Royal Society of Arts after 1908) offered a gold medal or a premium of 50 guineas for the first mineralogical map of England and Wales, of Scotland, and of Ireland. The offer was readvertised annually until February 1815, when William Smith (1769–1839) claimed the premium for his pioneering geological map of England and Wales. Again in London, at the Royal Institution (founded 1799), the professor of chemistry, (Sir) Humphry Davy (1778–1829), in 1805 and in subsequent years, delivered a course of lectures in geology; in support of those prelections, the Institution established a comprehensive geological museum.

The foundation of learned societies of a generalist nature continued in Britain down to the close of the eighteenth century. The Lunar Society of Birmingham was founded in 1775; the Manchester Literary and Philosophical Society, in 1781; the Royal Irish Academy, in 1785; and the Literary and Philosophical Society of Newcastle-upon-Tyne, in 1793. All of these societies brought geology with their purview, but as science progressed, there came into being specialists who were interested in the foundation of societies dedicated exclusively to individual sciences. In London, for instance, there were founded the Medical Society of London (1773), the Linnean Society of London (1788), the short-lived British Mineralogical Society (1799), the Horticultural Society (1804; it was granted the title 'Royal' in 1861), the Institution of Civil Engineers (1818), the Astronomical Society (1820; it received the title 'Royal' with its Charter of 1831), and the Zoological Society of London (1826).

The British Mineralogical Society was dissolved in December 1806 after a life of less than 8 years. One of the society's members had been the eminent London physician William Babington (1756–1833), and following the society's demise, a group of its former members, together with other gentlemen, began to hold informal scientific meetings at Babington's London home, which, from 1807, was at 17 Aldermanbury. Most of the meetings were held over breakfast, before the start of Babington's busy

daily medical rounds, but when the distinguished Humphry Davy joined the group – he termed his companions 'the Geophilists' – he urged that future meetings should be held following a dinner arranged within the convivial atmosphere of some favourite hostelry. His proposal found favour. A dinner was arranged to be held at the Freemasons' Tavern, in London's Great Queen Street, on 13 November 1807. Eleven gentlemen attended. One of them was Davy. He expected to see the Geophilists transformed into a properly constituted dining club (what he termed 'a little talking Geological Dinner Club'), but what he actually witnessed that evening was an event of a very different order of magnitude. He saw founded the Geological Society of London. It was the world's first society to be devoted exclusively to the science of geology. The resolution adopted that November evening read as follows:

> That there be forthwith instituted a Geological Society for the purpose of making Geologists acquainted with each other, of stimulating their zeal, of inducing them to adopt one nomenclature, of facilitating the communication of new facts, and of ascertaining what is known in their science, & what remains to be discovered.

Sir Joseph Banks (1743–1820), the president of the Royal Society, asked to be enrolled within the new body, but he soon had second thoughts. He had believed himself to be joining a geological dining club; he now discovered that the former Geophilists were aspiring to become a full-blown scientific society. For over a 100 years, geology had been a constituent of the Royal Society's empire of science. Now this upstart society seemed to be laying claim to geology as its own special preserve. Further, for a loyal Fellow of the Royal Society to become a member of the Geological Society was then seen as akin to being guilty of scientific bigamy. Banks, Davy, and several others who in all innocence had joined the new society promptly submitted their resignations.

Geological Society of London

Despite so inauspicious a beginning, the Geological Society of London prospered. Geology flourished: it was the most 'popular' science of the greater part of the nineteenth century. And for much of that century, the Geological Society of London was a cynosure for the world's geological eyes. The society had reached a membership of 450 by 1825, the year in which the society was granted the Royal Charter of Incorporation, under which its members were to be transformed into 'Fellows'. Honorary members were elected from all over the British Isles during the society's early years; from 1814, overseas geologists were elected as foreign members, and in 1863, there was added the additional category of foreign correspondent. The society first acquired its own premises in 1808; in 1828, the government provided the society with rent-free accommodation, first at Somerset House (1828–74) and then at Burlington House (1874–present). By its centenary in 1907, the society possessed a total of 1356 Fellows and overseas members.

From the outset, a prime function of the society was the holding of regular meetings at which papers were to be read and discussed. The first two papers, by Richard Knight (1768–1844) and Jacques-Louis, Comte de Bournon (1751–1825), were read following a dinner held at the Freemasons' Tavern on 4 December 1807, and since then tens of thousands of papers have been presented before the society. On 5 March 1824, the society received its first communication from a woman, Mrs Maria Graham (later Lady Maria Callcott; 1785–1842), and from 1887, the society received many papers from female authors. It was nevertheless March 1919 before the society finally resolved to open its fellowship to women geologists. The first woman to hold the presidency of the society, between 1982 and 1984, was Janet Vida Watson (1923–85).

The young society set about the collection of geological information out of which, it was hoped, there might emerge a true theory of the earth. To assist geological observers, the society in 1808 compiled and circulated a questionnaire entitled *Geological Inquiries*. This 20-page pamphlet was widely noted and was even reprinted in the United States within the *American Mineralogical Journal*. As a further manifestation of its desire to assemble and use geological information, the society in 1808 began the compilation of a geological map of England and Wales. This task was carried out under the direction of George Bellas Greenough (1778–1855), the society's first president (1807–13). The map was published, at a scale of 1 inch to 6 miles (1:380 160), on 1 May 1820 (not 1 November 1819, as recorded on the map), and subsequent editions were published by the society in 1840 and 1865.

The society in 1808 founded a museum to contain rocks, minerals, and fossils from around the world, but unsolicited gifts eventually converted the museum into an incubus. Inadequately housed and curated, the collection was disbanded in 1911, part of it being given to the Museum of Practical Geology in London's Jermyn Street (now the Museum of the Geological Survey at Keyworth, Nottinghamshire) and the remainder going to the British Museum (Natural History) (now the Natural History Museum in London's Cromwell Road). Happier is the story

of the society's library. Coeval with the society, by 1826, the library contained 1072 volumes and many pamphlets and maps, and substantial annual growth has continued to the present day. Today the library houses one of the world's most comprehensive collections of geological literature. The library is much consulted by the Fellows and by a wide range of individuals having need of geological information. The library is the society's proudest possession.

The great majority of papers read to the society have been published in one or other of the society's periodicals. These include the *Transactions of the Geological Society of London* (1811–56), the *Proceedings of the Geological Society of London* (1826–45), the *Quarterly Journal of the Geological Society of London* (1845–1970), and, since 1971, the *Journal of the Geological Society*, which appears bimonthly. Among the thousands of papers published by the society, highlights include the work by William Henry Fitton (1780–1861) on the geology of the opposed coasts of France and England (1826); (Sir) Henry De La Beche (1796–1855) on the geology of Jamaica (1827); Louis Agassiz (1807–73) on the glacial theory (1840); Henry Clifton Sorby (1826–1908) on the microscopical structure of Yorkshire grits (1850); Joseph Beete Jukes (1811–69) on the geomorphology of the south of Ireland (1862); Charles Lapworth (1842–1920) on the Moffat Series (1877); Benjamin Neeve Peach (1842–1926) and colleagues on the northwest highlands of Scotland (1888); Arthur Vaughan (1868–1915) on the stratigraphy of the Bristol region (1904); (Sir) Edward Battersby Bailey (1881–1965) and colleagues on the cauldron subsidence of Glen Coe (1909); Owen Thomas Jones (1878–1967) on the evolution of a geosyncline (1938); and John Frederick Dewey (1937–present) on the development of the South Mayo Trough (1962).

Since 1831, the society has added another important spoke to the wheel of its activities, by assuming a leading role within the reward system of the international community of the earth sciences. This was first possible when, on 10 December 1828, the society's council learned that William Hyde Wollaston (1766–1828), a mineralogist and a member and Fellow of the society since 1812, had left to the society the sum of £1000, the income from the investment to be used "in aiding or rewarding the researches of any individual or individuals, of any country". Further, Wollaston enjoined the society not to hoard the income, but to strive to make an award every year. The council resolved to bring into being a medal to be struck in gold (the medal has also been struck in palladium, a metal discovered by Wollaston), and the first of the long line of Wollaston Medals was awarded to William Smith on 18 February 1831. Since 1835, one or more Wollaston Medals have been awarded every year, and the medal is today recognized as the premier award within the world of geology. The roll of the medal's recipients glitters with distinction, as the following sample of recipients reveals: Louis Agassiz, 1836; Charles Darwin, 1859; Sir Charles Lyell, 1866; James Dwight Dana, 1872; Eduard Suess, 1896; Grove Karl Gilbert, 1900; Albert Heim, 1904; Baron Gerard Jacob de Geer, 1920; Reginald Aldworth Daly, 1942; Arthur Holmes, 1956; Alfred Sherwood Romer, 1973; John Tuzo Wilson, 1978. When Sir Roderick Impey Murchison (1792–1871) died, he left to the society a bequest sufficient to endow the annual award of a Murchison Medal, and at the death of Sir Charles Lyell (1797–1875), a similar bequest allowed the endowment of a Lyell Medal. Today the Wollaston, Murchison, and Lyell medals, together with sundry other younger awards placed within the gift of the society, are all made at the President's Evening of the society held early in May each year.

Other Geological Societies

The young Geological Society of London aspired to a standing that was national, if not international, but it speedily became the prototype for other more local geological societies that soon began to arise in many parts of the British Isles. Several of these societies of lesser ambition were rooted in regions where a nearby mining industry imparted to geology an especial significance. The following represent the principal British and Irish geological societies founded during the nineteenth century, in imitation of the Geological Society of London: The Royal Geological Society of Cornwall (1814), the Geological Society of Dublin (1831; restyled the Royal Geological Society of Ireland in 1864), the Edinburgh Geological Society (1834), the Geological and Polytechnic Society of the West Riding of Yorkshire (1837; after 1877, the Yorkshire Geological and Polytechnic Society; after 1905, the Yorkshire Geological Society), the Manchester Geological Society (1838; after 1903, the Manchester Geological and Mining Society), the Dudley and Midland Geological Society (1842; refounded as the Dudley and Midland Geological and Scientific Society and Field Club in 1862), the Geologists' Association (1858), the Geological Society of Glasgow (1858), the Liverpool Geological Society (1859), the Norwich Geological Society (1864), and the Hull Geological Society (1888). One other British foundation merits mention. At York in 1831, there was established the British Association for the Advancement of Science, based, somewhat, on the model of the German Gesellschaft Deutscher

Naturforscher und Ärzte (1822). Each year since 1831, with some interruptions during the two world wars, the 'B.A.', or the 'British Ass', has held its annual peripatetic gatherings in cities throughout the British Isles and in Commonwealth locations such as Montreal (1884), Toronto (1897, 1924), South Africa (1905, 1929), Winnipeg (1909), and Australia (1914). The Geological Section – today Section C – of the Association has always been one of the most important, successful, and popular of the Association's numerous sections.

The geological societies named so far have all been bodies dedicated to the earth sciences in general, but since the middle decades of the nineteenth century, increasing specialization within the earth sciences has encouraged the development of a new generation of societies dedicated to just one field within the earth sciences. In Britain, the earliest of these more narrowly focused societies was the small London Clay Club, established in 1838 for the collection, description, and illustration of the local Eocene mollusca. Considerably more important is the Palaeontographical Society, founded following a meeting of the Geological Society of London held on 3 February 1847, and dedicated to the publication of monographs devoted to British fossils. Other specialized British bodies founded within the field of the earth sciences include the Mineralogical Society (1876), the Institution of Mining and Metallurgy (1892), the Institution of Petroleum Technologists (1913), the Palaeontological Association (1957), the British Geomorphological Research Group (1961), the British Micropalaeontological Society (1970), and the Geological Curator's Group (1974). Within the Geological Society of London there are now specialist groups, of which the following organizations serve as a sample: the British Geophysical Association, the British Sedimentological Research Group, the Environmental and Industrial Geophysics Group, the Geochemistry Group, the Geological Remote Sensing Group, the Geoscience Information Group, the Hydrogeological Group, the Marine Studies Group, the Metamorphic Studies Group, the Petroleum Group, the Tectonic Studies Group, and the Volcanic and Magmatic Studies Group. The *Quarterly Journal of Engineering Geology* (now the *Quarterly Journal of Engineering Geology and Hydrogeology*) has been published by the Geological Society since September 1967.

During the past 150 years, the pattern of geological society foundation, evident in Britain since 1807, has been mirrored the world over as the global population of earth scientists has undergone dramatic expansion. The increasing interest in higher education, the growth of the hydrocarbon industry, the bourgeoning demand for industrial minerals, the incessant call for building materials, and the world's never-to-be-assuaged thirst for water have all proved to be the detonators of a geological population explosion. It has been an explosion such as could never have been imagined by those 11 founders of the Geological Society of London who dined together at the Freemasons' Tavern on 13 November 1807. In six continents, thousands of geologists have combined within the convenient and comfortable ambiance of geological societies. There geologists have sought intellectual companionship, there they have relished the inspiration of a geological milieu, there, for their research discoveries, they have found both fora for discussion and channels for communication, and there, by their labours, they have earned for their science that public prestige and influence that comes from nicely modulated collective activity. Further, in recent years, large numbers of enthusiastic amateurs have been attracted into geological societies. Never since the furore surrounding Darwin's *Origin of Species* (1859) have the earth sciences attracted as much attention as is theirs today. Plate collision, mass extinction, *Jurassic Park*, Creationist claims, lunar rocks, Martian images – these have all served as eye-catching billboards for geology. As a result, internationally, droves of amateurs have joined the ranks of the world's geological societies, there to make absorbing contact with a fundamental element of the human environment.

The Republic of Ireland has a population of only 3.9 million people, but it affords a microcosm of the type of recent institutional development that has characterized geology throughout the world. The Royal Geological Society of Ireland, dating back to 1831, died in 1894, partly as a result of the completion of the primary geological survey of Ireland in 1890, and partly as a result of the collapse of the small Irish mining industry as new and more profitable mines were opened overseas. With the second half of the twentieth century came revival. The Irish Mining and Quarrying Society was established in 1958 for those involved in the renewed and expanded extractive industries. The Irish Geological Association was launched in 1959 to offer a regular programme of events of interest to both the professional and the amateur. The Irish Association for Economic Geology was founded in 1973 to serve all those with a professional involvement in the economic aspects of the earth sciences. In 1978, there was founded a new journal, the *Journal of Earth Sciences Royal Dublin Society* (today the *Irish Journal of Earth Sciences* of the Royal Irish Academy). The Cork Geological Association was born in 1992 out of the enthusiasm of some Cork citizens

who had just attended an evening diploma course at the local University College. Finally, in 1996, there was established a Mining History Society of Ireland (since 2000, the Mining Heritage Trust of Ireland) "for all those persons interested in Ireland's historical mining industry".

The Irish societies are little more than dust on the international scene of geology. But on that same scene, other societies are massive landmark boulders, visible from all quarters of the geological world. These are the major societies that have played no little part in giving to the earth sciences their present character. Among those societies, pride of place must be accorded to the Geological Society of London. As the world's senior such society, it sometimes emphasizes its standing by strengthening the definite article and by dropping the geographical designation, to become simply 'the' Geological Society. It has been hugely influential within geology. At its centenary celebrations in 1907, the geological world paid a tribute to the mother of all geological societies, and many another tribute will doubtless be proffered at the society's approaching bicentenary in 2007.

Among the world's other notable societies are the Société Géologique de France (1830), the Società Geologica Italiana (1881), the Geological Society of South Africa (1895), the Palaeontological Society (1909), the American Association of Petroleum Geologists (1917), the International Quaternary Association (1928), the Geological Society of Australia (1952), the Sociedad Venezolana de Geólogos (1955), the Geological Society of India (1958), and the Nepal Geological Society (1980). Finally, among these leading societies, there is the Geological Society of America (1888). With a membership of around 20 000 geologists, it is the world's largest such society. The *Bulletin of the Geological Society of America* (1890) is one of the world's most significant geological journals, and the annual conference of the society is a major highlight in the calendar of international geology. The headquarters of the society was built in 1972 in Boulder, Colorado, at 3300 Penrose Place, named after Richard Alexander Fullerton Penrose (1863–1931), who was a major benefactor of the society. In Boulder, a fascinating location, both the grounds and headquarters building are replete with features of geological interest. Even the handles on the main doors are shaped in Scandinavian labradorite!

Although not a geological society in the strictest sense, one other geological institution, the International Geological Congress, does here merit mention. The story of the Congress begins in the United States. In 1876, the American Association for the Advancement of Science was meeting in Buffalo, New York, when there crystallized the notion that regular international gatherings of geologists might prove of benefit to their science. More specifically, the Buffalo resolution referred to such gatherings as being "for the purpose of getting together comparative collections, maps, and sections, and for the settling of many obscure points relating to geological classification and nomenclature". A successful congress was held in Paris in 1878, during the Paris Exposition Universelle, with 312 geologists present from 22 different countries. There followed seven other nineteenth-century congresses, at Bologna (1881), Berlin (1885; a congress planned for 1884 was abandoned because of an outbreak of cholera in southern Europe), London (1888), Washington (1891), Zurich (1894), St Petersburg (1897), and Paris (1900). Since 1900, 24 additional congresses have been held, although sometimes international events have interfered with the measured regularity of the congress procession. Between 1913 (Toronto) and 1922 (Brussels), World War I prevented the holding of any congresses, and the 1940 congress, scheduled for London, was cancelled on the outbreak of World War II in 1939. Most dramatic of all, the 1968 congress in Prague had to be abandoned on its second day as a result of the Soviet invasion of Czechoslovakia on the evening of 20 August 1968. That congress, like its predecessor in New Delhi (1964) and all subsequent congresses, was organized under the aegis of the International Union of Geological Sciences founded during the course of the Norden congress of 1960.

Further Reading

Eckel EB (1982) *The Geological Society of America: Life History of a Learned Society*. Boulder, CO: Geological Society of America.

Schneer CJ (1995) The geologists at Prague. History of the International Union of Geological Sciences. *Earth Sciences History* 14(2): 172–201.

Société Géologique de France (1930) *Centenaire de la Société Géologique de France: Livre Jubilaire 1830–1930*. Paris: Société Géologique de France.

Woodward HB (1907) *The History of the Geological Society of London*. London: Geological Society.

GEOLOGICAL SURVEYS

P M Allen, Bingham, Nottingham, UK

Introduction

'Geological survey' is the generic term used for the government organizations that carry out geological mapping; in addition a survey may also provide other basic geoscientific information and services to government, industry, commerce, and the general public. In the English-speaking world, survey organizations are often named simply 'Geological Survey', with country name as some part of the formal organization title. Elsewhere, the literal translation of the organization name into English may also be 'Geological Survey'; for example, the Sveriges Geologiska Undersökning (SGU) is the Swedish Geological Survey. Worldwide, however, there is an enormous variety in organization names (Table 1) and most geological entities have undergone several name changes during their history.

The first record of a geologist being employed by government was in late sixteenth-century Russia. In the late eighteenth and early nineteenth centuries, the French government and some of the states in the USA employed geologists specifically to make geological maps. This was stimulated by a need for information to support agriculture and mineral exploration, but when the first modern geological survey, the Geological Survey of Great Britain (GSGB), was founded in 1835, their remit had broadened. Now, most countries and many states, provinces, and regions have an organization that is recognized by international bodies as the representative governmental geological survey group. In some poor countries, particularly in Africa, the geological surveys are effectively moribund or are sustained by foreign aid. Nearly all geological surveys are funded by and are part of government. There is a common core element to their work. For most, their main purpose has been and remains to support the minerals industry, but among the more mature surveys this role is reducing as issues relating to the environment become more important in society.

The Work of Geological Surveys

The principal objective of geological surveys is to collect and interpret geoscientific data and put the information into the public domain. The work falls within five essential categories:

- Geological mapping on land and the continental shelf. In many countries, geochemical, geophysical, and hydrogeological (groundwater) mapping projects are also carried out by the survey.
- Appraisal and assessment of mineral, energy, water, and land resources.
- Baseline studies on geohazards (volcanic, seismic, and ground stability) and the environment.
- Maintenance of a geoscientific archive or database.
- Publication of maps, books, reports, and data packages about the survey work.

As a secondary objective, some geological surveys carry out work commissioned by government departments outside the parent organization, or by the private sector, which will pay for it at the market price.

In the USA, Canada, and several colonial African countries, among others, many geological surveys had to make topographic maps as a base for their geological maps. In most cases, topographic mapping is now done in another part of government, but the United States Geological Survey (USGS), still retains responsibility for it as well as for biology, hydrology, and hydrogeology. All geological surveys do geological mapping, but not all carry out work in all the other categories. Responsibility for groundwater resources is often vested in a separate body within government. Where hydrocarbons are important to the economy, they are sometimes excluded from the remit of the geological survey. A typical example is the Division of Geological and Geophysical Surveys (DGGS) in Alaska's Department of Natural Resources. It has no responsibility for oil and gas or even basin analysis, which are covered by another division in the state government.

Many geological surveys carry out activities beyond those regarded as essential. Some operate a minerals bureau, both issuing exploration licences and collecting production statistics. They carry out research in mining, mineral dressing, and metallurgy and may provide support for small-scale mining activities. Others, particularly state surveys in the USA and Italy, have an educational function. The Geological Survey of Zambia has a gemmology unit. Geological surveys in Indonesia, New Zealand, Japan, the USA, Italy, central America, and the Andean countries all have specialist volcanological units and carry out research in earthquake hazard. The British Geological Survey is one of a small number of surveys that does research in geomagnetism. Several surveys carry out overseas aid programmes and some, such as the Geological Survey of

Table 1 National and federal geological surveys founded in the nineteenth century

Country	*Present name of geological survey*	*Year founded*	*Total staff*[a]
Austria	Geologische Bundesanstalt	1849	80
Belgium	Belgian Geological Survey	1896	17
Bulgaria	Directorate of Geology and Protection of Substrate	1880	25
Canada (federal)	Geological Survey of Canada	1842	667
Denmark	Geological Survey of Denmark and Greenland	1888	354
Egypt	Egyptian Geological Survey and Mining Authority	1896	2600
Finland	Geological Survey of Finland	1885	671
France	Bureau de Recherches Géologique et Minière	1868	848
Germany (federal)	Bundesanstalt für Geowissenschaften und Rohstoffe	1873[b]	660
Hungary	Geological Institute of Hungary	1869	143
India	Geological Survey of India	1851	2900
Indonesia	Geology Research and Development Centre, Directorate General of Geology and Mineral Resources	1850[c]	N/A
Ireland	Geological Survey of Ireland	1845	51
Italy	Agency for Environment Protection and for Technical Surveys	1867	86
Japan	Geological Survey of Japan/AIST	1882	c300
New Zealand	Institute of Geological and Nuclear Sciences Limited	1867	258
Norway	Geological Survey of Norway	1858	198
Philippines	Lands Geological Survey Division, Mines & Geosciences Bureau	1886	394[d]
Portugal	Instituto Geologico e Mineiro	1857	290
Russia[e]	Geological Institute, Russian Academy of Sciences	1883	N/A
Spain	Instituto Technológico GeoMinero de España	1849	335
Sweden	Swedish Geological Survey	1858	268
Switzerland	Swiss National Hydrological and Geological Survey	1872	16
United Kingdom	British Geological Survey	1835	815
United States of America	The United States Geological Survey	1879	1600[f]

[a]In 2002, unless noted otherwise.
[b]The Prussian Commission of Surveying, the predecessor organization of the Prussian (then German) Geological Survey, was founded in 1841.
[c]The first government-funded geological research was initiated by the Dutch colonial government in 'Dienst van het Mijnwezen' based in Bogor, in 1850. The organization moved to Jakarta in 1869 and to Bandung in 1924, undergoing name changes each time.
[d]Total staff in 1994.
[e]The current organization in Russia, founded in 1883, can trace its predecessor organizations back to 1584.
[f]In July 2003.

India and the USGS, have work programmes in Antarctica. The Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) in Germany runs a research ship and carries out geoscientific research in the world's oceans.

Types of Geological Survey

Geological surveys are conducted at three levels, national, state, and federal.

National Geological Surveys

National geological surveys are solely responsible within their country for the national geological mapping programme. In most countries, the national geological survey will be the only geoscientific institution at any level of government that carries out geological mapping. Countries with strong regional, state, or provincial governments may have geological surveys that are administered at that lower level of government, but, except in certain federal nations, overall control of the national geological mapping programme is still retained at the national level. This applies in Brazil, where some provinces have their own geological surveys, but the national survey has offices in all the provinces and controls the mapping programme centrally. In Italy, all the regions have their own geological survey as a branch of their administration and fund them independently of the central government, but the national geological survey, the Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici (APAT), coordinates and provides funding for the national mapping programme, even though the work is done by the regional surveys or universities.

In some countries where there is a single, national geological survey, not all of the full range of survey activities are carried out by the main organization. In France, where the modern Bureau de Recherches Géologique et Minière (BRGM) did not come into being until 1959, other government bodies carry out geophysics and marine geoscience, and engineering geology is almost entirely done within universities.

In countries that were once part of the Soviet Union, or influenced by it, government-funded geoscientific research was always dispersed among several institutes, often in different academies of science, and it is not uncommon for geophysics and geochemistry to be done in separate institutes to geological mapping.

British Geological Survey The Geological Survey of Great Britain (GSGB), founded in 1835 as a branch of the Ordnance Survey, acquired independent status within Government and the right of access to private and public land for the purposes of making geological maps through Act of Parliament in 1845.

The Survey's aim, now nearly achieved, was to produce a uniform series of geological maps at the scale of one inch to the mile (now 1:50 000) for the whole of Great Britain. Coalfields and other areas of economic importance were resurveyed several times. From the early 1860s Survey geologists used maps at the scale of six inches to the mile (now 1:10 000) for field recording.

A core of field geologists was supported by petrographers, mineralogists, palaeontologists and chemists. The Museum of Practical Geology was part of the GSGB until 1984. The Survey first addressed matters relating to groundwater in 1872 and used geophysics to aid mapping from 1926. Staff in the two decades after 1945 was around 140.

In 1965 the GSGB was merged with the Overseas Geological Surveys to create the Institute of Geological Sciences (renamed the British Geological Survey in 1984), within the Natural Environment Research Council. Disciplines brought in from OGS included photogeology, isotope geology, applied mineralogy, mineral intelligence, economics and statistics, modern chemical laboratories, applied geophysics and responsibilities for mapping overseas. Geomagnetism and seismology units were also added at this time.

By 1980 major new programmes included mapping the continental shelf, bulk and metalliferous mineral resource evaluation, national geochemical and geophysical surveys, geothermal energy, offshore hydrocarbons assessment, deep geology, environmental pollution, radioactive waste disposal, geohazards and urban geology.

In 1971 the Rothschild Report, recommending that direct funding should be partly replaced by a customer/contractor relationship with government departments, was implemented. Direct funding fell sharply and the Survey's mapping programme, which no single department would fund, was reduced to a minimum.

The Butler Enquiry (March 1987) restored balance to the work programme. From 1990 direct funds were made available for a core programme of geoscientific surveys overseen by a Programme Board made up of representatives of the user community. Subsequent government enquiries led to progressive commercialisation of the BGS. By 2004 its income came from direct Government grants, research commissioned by Government departments, the European Union, research grants and commercial activity.

The Survey complement grew from 501 in 1965 to 1200 in 1984, declining to around 800 in 2004. Gross turnover exceeds £30 million annually.

State Geological Surveys

At state level, geological mapping and many of the other activities of a geological survey are carried out exclusively within the boundaries of a state, province, or region. Usually, state geological surveys are part of the administration of state, provincial, or regional levels of government and are funded by them. State-level bodies are found in countries that have a national geological survey, such as Italy and Brazil, as well as in countries with federal systems of government, such as Australia, Canada, Germany, and the USA, in which nearly all states and provinces have their own geological surveys.

Federal Geological Surveys

Federal geological surveys are funded by the federal government to provide a variety of umbrella functions, including the generation of nationwide coverage of small-scale geological maps, but they are not responsible for the systematic, nationwide, medium-scale geological mapping programme. The mapping is done by the geological surveys at the lower level of government. There are four federal geological surveys: the United States Geological Survey, Geoscience Australia, the Geological Survey of Canada, and the Bundesanstalt für Geowissenschaften und Rohstoffe in Germany. Typically, federal bodies provide specialist laboratory facilities and take responsibility for hydrocarbons research, maritime surveys in home and foreign waters, overseas activities, research that has a nationwide relevance, and the compilation of national, small-scale, overview maps.

United States of America Established within the Department of the Interior, the United States Geological Survey (USGS) was founded in 1879. Its remit included, "classification of public lands, and examination of the geological structure, mineral resources and products of the national domain." To carry it out the USGS then, as now, had to do both topographical and geological surveys.

The USGS started with a staff of 38 and a budget of $106 000. In its centenary year the available budget

was $765 million and there were 12 000 staff. Funding throughout its history has come from direct Government appropriation.

The Survey's programme was always practically based, but in the 1890s Director Charles D Walcott broadened its remit to include basic scientific research.

Originally concerned with mapping publicly owned mineral lands, the Survey's programme grew by responding to national and international economic and political pressures. It began its hydrographic programme in 1888 after a severe drought. The first staff sent overseas went to Nicaragua in 1897. Groundwater became a Survey interest in 1894; oil in 1901; coal and iron in 1905; strategic minerals in 1914. After a period of decline from 1920 to 1939, the USGS began a major period of expansion, partly stimulated by the Cold War. Its focus for geological mapping shifted to the search for radioactive minerals and to support engineering projects in 1950. By then there were geological maps for less than 10% of the country and the responsibility for systematic geological mapping was taken over by the State surveys. In 1956 the USGS began evaluating the impact of underground nuclear explosions. Extraterrestrial studies began in 1959 with a photogeological map of the moon. Antarctic exploration began in 1958; marine studies in 1962 and earthquake research after the Alaska earthquake in 1964, adding to existing research on volcanological and other natural hazards. Geothermal research began in 1974 after the oil crisis. From 1970 the USGS became increasingly involved in all aspects of environmental geology.

The USGS has suffered and benefited from political interference. There have been many reorganizations, culminating in major changes in the late 1990s following a threat to its existence. Now there are four major divisions: Biology; Geography; Geology and Water. The Survey's headquarters, in Washington D.C. up to 1973, are in Reston Virginia. There are regional offices at Menlo Park, California and Denver, Colorado; Earth Science Information Centers at Rollo and Sioux Falls, and representatives in all states.

Australia Government-appointed geologists were working in western Australia in 1847, but the first state survey in Australia was the Mines Department in Victoria, which by 1861 was producing geological maps. The federal survey, called the Bureau of Mineral Resources (BMR), based in Canberra, was not established until 1946. It formed by merging the Aerial Geological and Geophysical Survey of northern Australia (1935) and the Mineral Resources Survey (1942) with the aim to increase Australia's metal output. It took responsibility for geological mapping in northern Territory, Papua-New Guinea, and Canberra, where there were no geological surveys, and worked in conjunction with the geological surveys of Queensland and western Australia on their mapping programmes. The BMR subsequently changed its name to the Australian Geological Survey Organisation and later to Geoscience Australia. Both northern Australia and Papua-New Guinea now have their own geological surveys.

Canada The Geological Survey of Canada (GSC) was founded in 1842 to carry out a reconnaissance geological survey of Canada. It made the first topographic maps and carried out biological surveys. In 1934, restructuring made the GSC concentrate essentially on geological surveying. This task has now passed to the provincial surveys, though the GSC carries out geological mapping in provinces that do not have their own geological survey. The provincial surveys tend to work closely with the mining industry. Many were originally closely associated with the provincial universities, sharing staff with them. Among the first to form, in 1864, was the Geological Survey of Newfoundland. The youngest is the Yukon Geological Survey, founded in 2003.

Germany The German Geological Survey arose out of the Prussian state geological survey, which was founded by merging the Prussian Commission of Surveying with the Mining Academy and the Museum of Economic Geology in 1873. The Commission of Surveying had been carrying out geological mapping since 1841 and had established a standard scale of 1:25 000 for its work in 1866. The Mining Academy was removed from the survey in 1916. After the First World War, the German Geological Survey became more concerned with economic geology, especially oil exploration. From 1939, all the German state geological surveys came under central control, and throughout the war, the German Geological Survey attempted to coordinate geological surveying and oil exploration in all of the occupied countries. The structure collapsed in 1945, but was reconstituted by the British military government as the Geological Survey of the Federal Republic. Headquarters were established in Hannover, shared with the Geological Survey of Lower Saxony. Now it is called the Bundesanstalt für Geowissenschaften und Rohstoffe (the Federal Institute of Geoscience and Raw Materials). Before reunification, western Germany consisted of nine states, each with its own geological survey. The work at state level is distinctly different from that at the BGR, being concerned with strictly local geological matters, including geological mapping. In the German

Democratic Republic, there was a single geological survey, the Zentrales Geologisches Institut (ZGI), until reunification. Now there are five state geological surveys in the former East Germany.

The Start of Geological Surveys

The history of government funding for geology starts in 1584, when the Russian government set up a 'stone department' (*Kamenay prikaz*) to regulate and encourage private enterprise in prospecting for mineral deposits as it colonized Siberia. In 1700, Peter the Great transformed that body into the Department of Mining and later (1719), the Ministry of Mining. This ministry continued to organize geological research and mapping and in 1883 established a 'Geological Committee', modelled on the Geological Survey of Great Britain, to stimulate a national geological mapping programme. Under various names, this organization has continued to the present day. In France, government charged the Corps des Mines to gather geological information about the country in 1794 and authorized the preparation of a geological map of the country in 1822. It was not until 1868 that the Service de la Carte Géologique de la France, the precursor to the present BRGM, was founded to oversee the preparation of a national geological map.

In the United States, federal funding for geological work was made available in 1804 in connection with military expeditions to unexplored territories. In 1810, the Academy of Science in Connecticut approved the appointment of a geologist to map the state. Work began in 1820. Geological surveys were established in the Carolinas in 1823 and 1825 and in 14 other states between 1830 and 1839. In many cases, the stimulus for the foundation of the geological bodies was to provide support for agriculture, the mainstay of the US economy at that time. In 1834, the United States Congress attempted to establish a federal geological survey by permitting the Topographical Bureau of the US Army to spend money on a geological map of the United States. This was abandoned after 2 years and it was not until 1879 that the United States Geological Survey was founded.

In Great Britain, a geologist was appointed to the Trigonometrical Survey (Ordnance Survey) in England in 1814, and in 1826 in Ireland. As early as 1815, William Smith, in the memoir accompanying his geological map of England, listed beneficiaries ranging from agriculture to water supply to mining and quarrying. When the Geological Survey of Great Britain was founded in 1835, then, there was no single justification for it. The GSGB differed from all other similar institutions at that time. Small-scale geological maps of Great Britain, created independently of any government funding, were already in existence. Thus, the GSGB founder, Henry de la Beche, proposed to map the whole country systematically at the comparatively large scale of 1 inch to 1 mile. He also started a Museum of Practical Geology, a School of Mines (now the Royal School of Mines), an educational programme for the common people, and a Mining Records Office. None of these now remain within the field of responsibility of the British Geological Survey, but de la Beche's vision provided a model that was copied worldwide.

Later Developments

De la Beche's broad view had appeal throughout Europe and directly stimulated many countries to institute their own geological surveys on the GSGB model (Table 1). The German states of Prussia, Westfalia, and Rhine all started geological mapping programmes in 1841; Ireland followed in 1845, and Austria and Spain, in 1849. By the end of the nineteenth century, more than half the European countries had national geological surveys. Today, all except the small states (Andorra, San Marino, and Liechtenstein) have them, the most recent being Latvia, in 1995.

Throughout the nineteenth century, the demands of the British manufacturing economy and the developing industrial economies of Europe and North America led to many geological surveys being established with the purpose of providing geological maps for use in the exploration for natural resources for export. In the British Empire, geological surveys were founded in Canada in 1842, India in 1851, the state of Victoria in 1861, and the Cape of Good Hope, South Africa in 1895. Elsewhere in the British Empire, government-funded geological reconnaissance was carried out throughout the second half of the nineteenth century by geologists seconded from the GSGB. Geological surveys were established in several parts of the Empire in the first quarter of the twentieth century (*see* **Colonial Surveys**), but most came in a second wave in the 1940s and 1950s during the post-World War II drive to find resources for reconstruction. There is a similar picture among the French colonies. In Indochina, three organizations were established to provide geological information for what became Vietnam, Laos, and Cambodia: Service des Mines de la Cochinchine (1868), Service des Mines de l'Indochine (1884), and the Service Géologique de l'Indochine (1898). In Africa, however, the French colonial geological surveys were either small or nonexistent until after the Second World War, when considerable expansion took place for essentially the same reasons as in the

British Empire. The Dutch, Spanish, and Portuguese administrations all started geological surveying in their colonial possessions. In Indonesia, it was in 1850, 53 years before a national geological survey was established in The Netherlands. Outside Europe and its colonial possessions, most countries were late to found their own geological surveys. Japan (1882) was among the earliest; China's was set up in 1913. In central America, South America, and the Caribbean, where there are now over 20 geological surveys, those in Argentina (SEGEMAR) and Columbia (SERGEOMIN) were founded in 1904 and 1916, respectively, but nearly all the rest were established after 1940.

By the end of the twentieth century, nearly all countries had a geological survey and, outside Europe, many are still primarily concerned with supporting the mining industry. Though geological mapping remains an essential activity for all surveys, the focus of attention in the more mature surveys has now shifted away from providing maps and other information to support mineral exploration and exploitation to environmental issues such as groundwater supply and quality, rehabilitation of 'brown' land, and on providing information on undermined areas, ground stability, and geohazard.

Geological Mapping

There is a natural progression in the way in which geological mapping (*see* **Geological Field Mapping**) is carried out in geological surveys:

- Stage 1 is reconnaissance, to find out what is there and to construct a small-scale map for the whole territory. Some of the early state surveys in the United States were closed down after this stage was completed. The scale of the reconnaissance map is small. The first map of the whole of the USSR, completed in 1956, and in China, completed in 1999, was 1:1 million. In France, the first geological map, published in 1841, was at 1:500 000, as was the first in Portugal, in 1876. In Czechoslovakia, the first map, completed in 1960, was scaled at 1:200 000.
- Stage 2 is the first, detailed, systematic survey. Its purpose is to generate medium-scale geological maps that fit into a regular grid pattern covering the whole of the territory. The scale chosen for this stage was 1:200 000 in China and 1:100 000 in Argentina. Most countries, even some large ones such as India, Pakistan, and South Africa chose a scale of 1:50 000 for complete, detailed national cover. This is the preferred scale for most European countries. Some state surveys, including the Prussian Geological Survey and the Ontario Geological Survey, have chosen 1:25 000 or a combination of 1:25 000 and 1:50 000. In some American states, medium-scale maps are drawn at 1:24 000.
- Stage 3 is a systematic, but local, large-scale survey. Because this mapping is demand driven, it will not normally lead to full, national map cover, but will be confined to areas of particular economic importance. In Great Britain, the chosen scale is 1:10 000; in France, 1:20 000; in Russia, either 1:50 000 or 1:25 000; and in India, 1:25 000. In many countries, urban areas are mapped at 1:5000.

The Size of Geological Surveys

Among the smallest geological surveys are the Service Géologique du Luxembourg and the Planning Directorate in Malta, both with a staff of eight. The largest was, and remains, the China Geological Survey, which in 1994 employed 1.1 million; in 1999, it was reorganized and staff was reduced to 6500. Much of the work the previous organization carried out was devolved to commercial enterprises in the provinces. After the breakup of the Soviet Union, many large eastern European geological surveys underwent drastic staff reductions, mostly accomplished by reducing the numbers of non-scientific staff. The most savage reduction was in the Bulgarian geological survey, which in 2001 was 25% of its size in 1996. Within Europe, the large geological surveys (Table 2) employ 600 to 850 persons; medium-sized surveys employ around 300 and small ones have 100 or fewer staff. Outside Europe, the ratio of scientific to non-scientific staff continues to vary considerably, so that comparisons of size using total staff numbers are not meaningful. Turkey employs 3456; the much larger India has a staff of only 2900. Because the prime function of any geological survey it to make geological maps, a useful measure of whether the survey is appropriately sized for its country is to divide the area of the country (in square kilometres) by the number of permanent staff. Another measure is to relate the size of the survey to population (Tables 2 and 3).

Relationships with Government

Most geological surveys are embedded within the structure of government as a division within a department or ministry. The home department will be most commonly one with responsibility for mines, minerals, or natural resources; energy; the environment; science and technology; or industry. Some geological surveys are publicly funded research institutes standing apart from government, but with a government department acting as a supervisory body. Among these, the Geological Survey of South Africa, called the Council for Geoscience since 1993, is a

Table 2 Ratios of area and population to total staff for 30 national geological surveys[a]

Country	*Ratio of area (km^2) to staff*	*Total staff*	*Ratio of population (in thousands) to staff*	*Rank of country in terms of population ratios*
Netherlands	130	315	50.1	12
Greece	159	827	12.9	2
Turkey	225	3456	19.3	5
United Kingdom	300	815	72.2	17
Luxembourg	323	8	53.9	14
Malta	323	8	48.6	11
Egypt	385	2600	26.3	8
Estonia	417	108	12.9	3
Finland	502	671	7.7	1
Lithuania	567	115	31.9	9
France	641	848	69.7	15
Hungary	650	143	70.2	16
Pakistan	718	1120	139.7	23
New Zealand	1041	258	15.0	4
Austria	1048	80	102.6	20
India	1092	2900	349.5	27
Japan	1232	300	442.4	28
Ireland	1350	51	73.1	18
Poland	1431	725	53.5	13
China	1475	6500	196.5	25
Spain	1507	335	118.3	21
Norway	1636	198	22.6	7
Sweden	1672	269	33.1	10
Belgium	1795	17	597.2	29
Iran	2354	700	96.7	19
South Africa	3822	310	130.2	22
Bulgaria	4436	25	329.0	26
Brazil	7170	1187	143.3	24
Ukraine	9432	64	788.4	30
Namibia	9931	83	20.8	6

[a]Staff numbers taken from various sources dated 2001 to 2003. Population figures and land areas are from the GeoHive Global Data website at www.geohive.com.

Table 3 Ratios of area and population to total staff for some state and provincial geological surveys[a]

State or province	*Ratio of area (km^2) to staff*	*Total staff*	*Ratio of population (in thousands) to staff*	*Rank in terms of population ratios*
Lower Saxony (Germany)	207	230	34.3	4
Delaware (USA)	353	15	53.8	6
Indiana (USA)	1089	86	71.6	7
Alabama (USA)	2526	53	84.7	8
Tasmania (Australia)	2974	23	20.6	3
Ohio (USA)	3148	34	335.9	11
Victoria (Australia)	5685	40	121.4	9
Texas (USA)	5906	117	186.2	10
South Carolina (USA)	7326	11	373.4	12
Alberta (Canada)	11 030	60	51.9	5
Western Australia	19 022	133	14.4	1
Alaska (USA)	40 281	38	16.9	2

[a]Staff levels are taken from geological survey websites in 2003. Population and land area figures are from the GeoHive Global Data website at www.geohive.com.

free-standing science council, and the BGS is part of the Natural Environment Research Council, a non-ministerial government department. The most extreme development of this kind is the Institute of Geological and Nuclear Sciences Ltd in New Zealand, which is a limited liability company. Some of the state surveys in the United States have no organizational links with government, but are attached to universities.

Funding

Most geological surveys are funded entirely or mainly by a direct grant from government, and some are prohibited by law from being involved in commercial activity. Others, particularly in Europe, are semicommercial in their mode of operation. Until 1994, The BRGM in France owned and operated a mining company fully commercially, though no funding was transferred from mining to the survey activity. Since 1973, the British Geological Survey has derived most of its income from commissioned research for government departments other than the supervisory one, and in the past decade, by an increasing amount of commercial activity. This pattern is repeated in many other countries, and by 2001 nine European surveys received less than 70% of their income as a direct grant from government. The lowest percentage grant from government: the lowest being Iceland, with only 30%. The balance in these surveys is made up of income earned from the European Union research grants, some from the private sector, but mostly on commission from various government departments. The Council for Geoscience in South Africa is divided into business units and its commercial freedom extends to allowing it to own and run hotels.

Associations and Resources

Geological surveys throughout the world have banded together to form loose, common-interest groups. Among these are the International Consortium of Geological Surveys (ICOGS); the Commonwealth Geological Surveys Forum (for countries in the British Commonwealth); the Forum of the European Geological Surveys Directors (FOREGS), which arose out of the Western European Geological Surveys (WEGS, founded in 1973); the Association of the Geological Surveys of the European Union (EuroGeoSurveys), a lobbying group to act on behalf of the geological surveys within the European Union; the Asociación de Servicios de Geología y Minería Iberoamericanos (ASGMI), which consists of the geological surveys of Spain, Portugal, and Latin America; and the Coordinating Committee for Offshore and Coastal Geoscience Programmes in South-east Asia (CCOP), which contains all the geological surveys of the region, though membership is not exclusive to them.

Information about geological surveys is dispersed. Some geological surveys have their own written histories, which can be acquired directly from them or through a library. In most cases, the best information is obtained through survey/association websites. Three groups maintain gateway sites that can be used to access geological survey information: the British Geological Survey, at www.bgs.ac.uk/geoportal; the Open Directory Project, a citizen-editor/contributor database project, at www.dmoz.org/Science/Earth_Sciences/Geology/Organizations; and McCully Web, at www.mccullyweb.com. Other sources include newsletters of the International Consortium of Geological Surveys and publications of the Forum of the European Geological Surveys Directors.

See Also

Colonial Surveys. **Geological Field Mapping**. **Geological Societies**.

Further Reading

The report on the symposium meeting on the Organization of Geology Overseas, in the *Proceedings of the Geological Society of London, No. 1633, Sept 1966,* gives a good account of the origin of geological surveys in Australia, Canada, China, Czechoslovakia, France, the United States, and the former USSR.

GEOLOGY, THE PROFESSION

G L Jones, Conodate Geology, Dublin, Ireland

Introduction

Geology deals with the Earth and its processes. The variety of ways in which geologists work has been evolving for millennia. It is now the most complex and diverse profession in society, so that geologists could be regarded as modern-day polymaths. The profession continues to evolve, with many areas of expertise, ranging from palaeontology to geophysics and from volcanology to environmental geology. Along with other modern professions, it has adopted the professional qualification system, which is based on the four pillars of academic training, professional experience, a code of ethics, and continuing professional development (Figure 1).

Ancient History

Geology as a profession can be traced back to the birth of civilization. At the moment that our hunter–gatherer ancestors picked up a stone as an implement, the profession of geology was born.

Soon, the discerning, evolving, human being realized that some rocks have better properties than others depending on the needs at hand. Some were harder and did not shatter when used as a hammer; others could be worked to produce a sharp edge and wielded as axes, arrows, or spears. As an interest in rocks developed because of their use as building materials, metal sources, agricultural ingredients, pharmaceuticals, jewellery, and wealth, the *Homo sapiens* with a geological eye became valued by society.

The specialized quarryman, ore smelter, or miner was representative of geologists for thousands of years. Their keen observation and skill was exemplified by their ability to hue megaliths and other objects for use in many aspects of Mesopotamian, Egyptian, Greek, Roman, Aztec, and Inca life. Their constructions produced such marvels as elegant obelisks, the impressive Parthenon, dramatic amphitheatres, and pyramids – to mention only a few. These men knew their geology, and they understood the nature of bedded rocks and how joints broke up the beds into usable blocks. So the profession was divided for a long while into miners, quarrymen, masons, brick makers, etc.

However, even in ancient times people thought about the causes of phenomena such as earthquakes and volcanoes. Philosophers such as Aristotle, Lucretius, Herodotus, and especially Avicenna made astute observations based on meagre information, and these data were preserved through the dark ages only by Arab intellectuals. Then, towards the end of the fifteenth century, Leonardo da Vinci observed that fossils had once been living organisms and that the land had once been covered by the sea. So it was with the development of the Enlightenment that ancient ideas were re-examined and new thoughts advanced by such Italians as Vallisneri and Moro. They followed Bishop Nicholas Steno, the founder of modern geology, who developed the ideas of stratigraphy in Florence in 1669. Then, in the late seventeenth century, Descartes in France followed by Leibnitz in Germany formulated the concept of the development of the planet from vapour through molten rock to the solid surface.

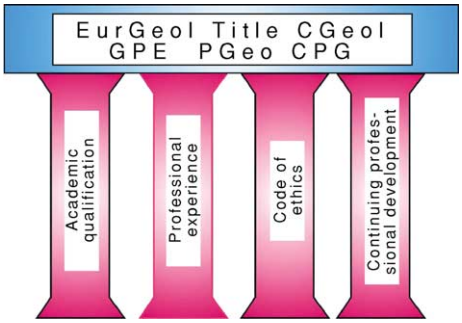


Figure 1 The four pillars of professional geological qualifications. EurGeol: European Geologist (Europe), CGeol: Chartered geologist (UK), GPE: Geologo Profesional Especialista (Spain), PGeo: Professional Geologist (Ireland) and Professional Geoscientist (Canada), CPG: Certified Professional Geologist (USA).

More Recent History

In the eighteenth century geology became the preserve of the gentleman polymath. These men of substance, quite a number of them questing clergymen or surgeons, found that they could read the rocks like a book and could see that certain layers always appeared above others and that they contained distinctive fossils. Gradually they put together the sequence of deposition of the rock layers.

In France, the Académie Royale des Sciences was the seat of the discussions that crystallized French learning about the Earth, such as Desmarest's 1771 memoir, in which he presented his theory of the

volcanic origin of basalts. The founding fathers of modern geology included men such as James Hutton (*see* **Famous Geologists:** Hutton), who published his '*Theory of the Earth*...' in 1795, and Charles Lyell (*see* **Famous Geologists:** Lyell), whose *Principles of Geology* was published in 1833. In England gentlemen organized themselves into the Geological Society of London in 1807; this was initially a discursive dining club, but it eventually became the home of geology (*see* **Geological Societies**).

Geological maps were being drawn at that time in many countries and included Tighe's Kilkenny map of 1802 in Ireland and Mantell's 1822 map of the South Downs in England. However, one man who was involved with the practicalities of the industrial revolution made enormous strides. A blacksmith's son working on the development of canals throughout the UK drew detailed maps of the changes in the rocks across the countryside. So William 'Strata' Smith (*see* **Famous Geologists:** Smith) drew the first geological map of most of Britain in 1815.

Thus during these times geological learning evolved from a restricted interest in mining or mineralogy to broader geology. For a long time, in many parts of Europe, geology was seen as a subservient branch of engineering, and amazingly in some countries a geologist still cannot sign his own professional report, but must ask an engineer to do it for him!

The Breadth of Geology

It can be seen that the practice of geology developed and enlarged over the millennia and embraced many disciplines. It has burgeoned into perhaps the broadest discipline to be found in society.

Geologists include palaeontologists and palynologists who specialize in zoological and botanical sciences; their understanding of the processes of evolution is central to their work. Sedimentologists image past processes and compare them to modern ones. Mineralogists and crystallographers deal with minerals and crystals. Metamorphic geologists look at the changes that take place in rocks under extreme heat and pressure. Hydrogeologists deal with the crucial area of the movement of groundwater and our ability to provide clean water to society; a subgroup specialize in the disposal of society's waste in a safe, clean manner. Geophysicists, through mathematics and physics, use diverse techniques, such as the micromeasurement of gravity, the response of the ground to electrical currents, the reflection of radar waves, and, of course, the measurement of seismic waves, whether produced naturally by earthquakes or by artificial sources. Marine geologists look at geology below the surface of the sea, imaging the seafloor and working with their geophysicist colleagues to see deep into the rock strata. Petroleum geologists understand the formation of oil and gas and how these vital commodities are caught in traps in the rocks; they help to deliver these resources to society. Volcanologists or igneous geologists deal with active volcanoes and igneous processes and with their now fossilized equivalents in the geological record. Geochronologists use their knowledge of the radioactive decay of some elements as a crucial tool in the absolute dating of rocks and sediments, for example by using the decay of uranium into lead to date ancient intrusive igneous rocks that are between 1 Ga and 400 Ma old, the decay of potassium-40 into argon do date rocks formed during the last 500 Ma, or the decay of carbon-14 into nitrogen to date more recent organic sediments that are a few tens of thousands of years old. Geochemists specialize in the mineral constituents of rock, following on from the smelting of ores for thousands of years, and this was an important aspect of the development of the science of chemistry. Structural geologists apply mathematical and geometrical knowledge to the way that rocks behave under varying conditions of pressure and temperature. Engineering geologists must know about the technical properties of rocks and sediments and be able to communicate this information to their engineering colleagues; they also deal with natural hazards such as danger from landslips, etc. Environmental geologists have become central to the management and development of society; their input into land-use planning improves the quality of life for the inhabitants and ameliorates the risks posed by geohazards. Planetary geologists learn from our planet and its moon and work with their astronomical colleagues in investigating distant bodies. Remote-sensing geologists use satellite imagery to examine the surface of our planet.

Our science overlaps into a plethora of other disciplines.

Academic Education

As geology gradually evolved, existing university natural history courses began to include geological options. In the five years after Waterloo in 1815, the universities began to formalize matters: Buckland was taken on as reader in Oxford, and Sedgwick (*see* **Famous Geologists:** Sedgwick) received the chair in Cambridge. These chairs were in natural history, mineralogy, geology, or a combination, but the course was becoming a geological one. The degree awarded was still a Bachelor of Arts, of course.

This trend continued, and colleges around the world began teaching geology. Degree courses

evolved until 3 years of education were required to receive a Batchelor of Science degree, or 4 years with Honours. Most universities now offer postgraduate Master of Science degrees, involving 1–2 years work, through thesis and/or examination, whilst Doctor of Philosophy (PhD) degrees may take between 3 and 5 years of research to complete. In Europe, this may take up to 10 years. In general it is considered that most geologists need to complete at least 5 years of education before becoming suitable for a long-term geological career.

The Learned Societies

In order to cater for all the interests of these geologists, various bodies came into existence. They fall into the categories of learned and technical societies, but a new function or even a new type of society has evolved with the recent birth of the professional societies (*see* **Geological Societies**).

We see that the major national societies such as the Geological Society of London, the Union Français des Geologues, and the Ilustre Colegio Oficial de Geólogos cater for many areas of geological life and interest, especially when they are able to set up special interest groups. The specific concerns of large groups of specialized geologists, especially when these are of a global nature, have lead to the establishment of large organizations such as the International Association of Engineering Geologists, the European Union of Geological Sciences, the Petroleum Exploration Society of Great Britain, the Micropalaeontological Society, the International Association of Hydrogeologists, and the Society of Exploration Geophysicists. All cater for a particular area of geological interest, and most produce a journal carrying technical peer-reviewed articles.

The Professional Bodies

The evolution of the professional geologist, working in academia, government, or industry, has lead to the need for organizations to look after them. Some learned societies, such as the Geological Society of London, have evolved easily to cater for this new function. In Ireland it was necessary to create a new independent body – the Institute of Geologists of Ireland – to look after Irish professional geologists, and in the USA the American Institute of Professional Geologists took on this role. The European Federation of Geologists (Figures 2 and 3) took on the continental role of representing the national associations, some of which were professional bodies and others learned societies. Similarly, in Canada the Canadian Council of Professional Geoscientists represents the professional interests of the provincial bodies (Table 1).

All of these bodies are striving towards the same goal of ensuring professional standards and representation. They are establishing mutual recognition and other agreements to benefit their members worldwide.

The Profession

The practice of geology occurs in three principle areas: academic, governmental, and industrial.

Clearly, the teaching of geology is crucial, and it is amazing that so little is taught in primary or secondary schools; geology is simply seen as a small aspect of geography! It is in tertiary education that geology comes into its own, and universities develop courses that give the neophyte geologist a grounding in so many disciplines. So, most geologists working in academia are based in university geology departments, though some may be found in the allied disciplines of geography, archaeology, etc. Some can also be found in Schools of Mines or Technical Institutes, where they instruct mining geologists, geosurveyers, and geological technicians.

In Government, the traditional area of practice has been in geological surveys. These have long been seen as the providers of modern geological maps, and this is indeed one of their main functions. It has been interesting to see the variation in the response of geological surveys around the world to modern changes in such areas as digital data management. Those surveys that have remained traditional have come under great pressure, and some have had to close. Others that understand the demands of the digital era have been quick to adjust to being data managers and providers, not just map makers, and so have become a necessary organization rather than an 'appendix'. Consequently, geologists working in these areas have also developed their geological skills, often into new areas such as heritage or marine. Other bodies have also created opportunities for geological employment, and in the last few years departments such as Environmental Protection Agencies have sprung up to look after the world we live in. Geology naturally also has a role to play in the supply of renewable energy, allowing countries to be less reliant on fossil fuels.

Industry is a huge area of employment for geologists. The supply of fossil fuels or mineral resources depends on the geological setting and on the location of the deposit to be exploited. So a geologist must be mobile and able to travel to any part of his or her country and even to any part of the world if necessary. Consequently, an ability with foreign languages can be highly advantageous at times. However, in the area

Figure 2 Mission and objectives of the European Federation of Geologists.

of industrial minerals, geologists deal with the supply of basic raw materials such as limestone, sand and gravel, and aggregates; many of these materials are supplied very close to their extraction source since transport is the major cost.

The birth of new areas of geology, taking over to some extent from the older ones, has lead to more home-based occupation, including hydrogeology, waste disposal, pollution remediation, environmental geology, geotechnical geology, renewable energy geology, and geophysics.

Professional Qualifications

Professional qualifications have been developed during the last half of the twentieth century. This profession has set up requirements that geologists must demonstrate not only that they have a good academic education, but also that they have achieved a high standard of professional expertise over a number of years and that they can demonstrate this in front of their peers by examination of both their work and their knowledge. Another part of this process has been the signing by geologists of agreements to obey codes of ethics demonstrating that they will work to high ethical standards. These codes are backed up by disciplinary committees. A further aspect that has become formalized is the agreement to demonstrate that continuing professional development is being carried out by the geologist.

Around the world this movement has proceeded and the standards that have been set up are very similar in most countries, so that across Europe there is a single standard to be achieved, and this is recognized to be equivalent to the standard achieved in other continents, with reciprocal rights and recognitions existing between different international groups, allowing geologists to travel and practice around the world.

Also of importance has been the acceptance by government and business bodies of this professional standard. For them, this quality-assurance mark is a

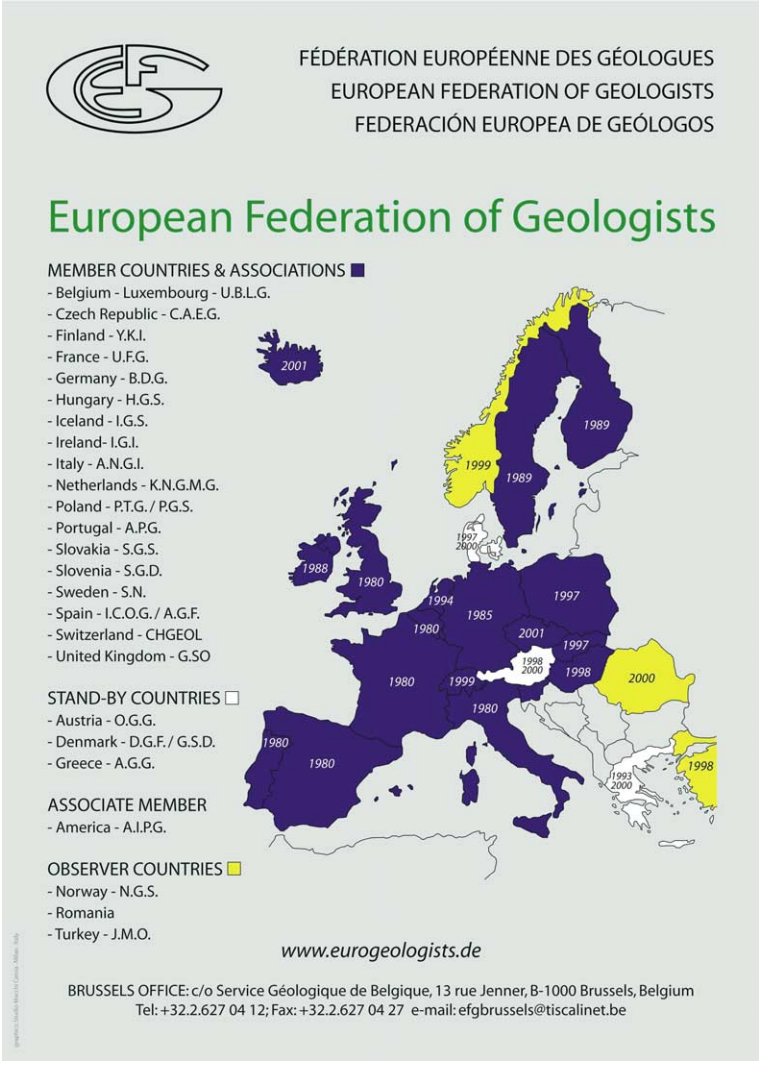


Figure 3 Members of the European Federation of Geologists. Single dates are years that countries became observers or members. Double dates are periods for which countries were members.

Table 1 Professional websites

European Federation of Geologists	www.eurogeologists.de
Geological Society of London (UK)	www.geolsoc.org.uk
Illustre Colegio Oficial de Geólogos (Spain)	www.icog.es
Institute of Geologists of Ireland	www.igi.ie
American Institute of Professional Geologists	www.aipg.org
Canadian Council of Professional Geoscientists	www.ccpg.ca
Berufsverband Deutscher Geologen, Geophysiker und Mineralogen EV (Germany)	www.geoberuf.de

way of ensuring high standards in the geological input to projects.

Part of the motivation for this movement has been the necessity to demonstrate high standards to local, national, and continental governments. Especially in this litigious age, the use of appropriately highly qualified people in all aspects of life is becoming a necessity.

Another powerful motivation has been the recent requirement by financial bodies such as the Canadian, Australian, British, and Irish stock exchanges that all sections of bankable reports on mineral

deposits are signed by competent people. The holders of these professional geological qualifications, available in Europe and North America, are accepted by these bodies as competent people in their fields of geological expertise. Government departments are also beginning to recognize that these qualifications help to ensure the quality of reports submitted to them by professional geologists.

In Europe, the European Federation of Geologists is the representative body for geologists, and it is composed of the national geological associations of each of the nineteen member countries (Figure 3). It administers the professional title of European Geologist, which is equivalent to the professional titles of Professional Geologist in Ireland, Chartered Geologist in the UK, and Titulo de Geologo Profesional Especialista in Spain. For these countries and for those that have no national professional qualification it provides a European route to a continentally recognized qualification.

Regulation

In some countries the State has taken upon itself to regulate the profession. In Italy and in Spain, in order to practice, a geologist must be a member of the body recognized by the state – the Consilio Nazionale dei Geologi and the Ilustre Colegio Oficial de Geologos, respectively. In the USA, many states require a geologist to be licensed in order to practice. The National Association of State Boards of Geology coordinates this licensing system. The role of the professional body, the American Institute of Professional Geologists, then becomes the organization of a high professional standard. In Canada the Provincial Geological Bodies have come together to form the Canadian Council for Professional Geoscientists, which coordinates the professional standards organized by each state and formalizes their mutual recognition.

The Future

The profession is still changing rapidly. Will the trend towards calling ourselves Earth Scientists instead of Geologists hold sway? I hope not! It appears to be a fad with no practical basis except to persuade academic bodies that real change is going on.

The geological profession has a clearly defined role in caring for Earth resources – clean water, the environment, renewable and finite energy, geohazards, and heritage – and in observing the continually changing process that is our Earth.

See Also

Engineering Geology: Codes of Practice. **Famous Geologists:** Hutton; Lyell; Sedgwick; Smith. **Geological Maps and Their Interpretation**. **Geological Societies**. **Geological Surveys**. **History of Geology Up To 1780**.

Further Reading

Cadbury D (2000) *The Dinosaur Hunters*. London: The Fourth Estate.

Geikie A (1987) *The Founders of Geology*. London: Macmillan.

Herries Davies GL (1983) *Sheets of Many Colours*. Dublin: Royal Dublin Society.

Regueiro M and Jones GL (eds.) (2000) *European Geologist 10, Proceedings of the First International Professional Geology Conference, July 2000. Alicante* Brussels: European Federation of Geologists.

Selley RC (1996) What on Earth is a Geologist? *European Geologist*, p. 3–4.

Winchester S (2001) *The Map that Changed the World*. London: Penguin.

GEOLOGY OF BEER

S J Cribb, Carraig Associates, Inverness, UK

Introduction

The fermentation of an extract of cereal grains is probably one of the oldest processes known to man, with techniques stretching back to Ancient Egypt. In the UK and parts of western, central, and eastern Europe a wide variety of fermented barley beverages are produced, known generally in English as beer or ale. Much of the variety is a direct result of the chemical composition of the water used to extract the sugars from the barley prior to fermentation. This variation derives from the geology of the water sources.

The pleasures of beer consumption are encapsulated in the following quotation, from an unknown source.

> The enjoyment of a glass of beer may be received by many senses: the sight may be attracted first by the clarity of a pale ale or the rich creamy head of a stout. As the glass is raised to the lips the aroma of the beverage, possibly the bouquet of the essential oils of the hops, may excite the nostrils. Then, as the liquid flows over the taste buds at the back of the mouth, and further volatile products diffuse into the back of the nose, the flavour of the beverage is perceived. Finally, the beer enters the body, where the alcohol is rapidly absorbed into the bloodstream and exerts its well-known physiological and psychological effects.

In more formal terms, in any particular brew the brewer is seeking a combination of five characteristics: flavour (including both taste and aroma), alcohol content, colour, head retention, and clarity. The chemical nature of the water used in the brewing process exerts a strong control over all of these.

The Brewing Process

There are several stages in the brewing process, and these are summarized below.The first stage is malting, which is the conversion of barley grains into fermentable malt. The grains are steeped in water and spread out to germinate. This initiates enzyme processes, which start to break down complex carbohydrates into sugars. Kilning, heating to around 100°C, arrests these processes before they go to completion.

The malt is milled (ground) to form grist and then mixed (mashed) in a mashing vessel or tun with hot brewing water (liquor). Here the processes started during malting continue.

The liquor and malt (wort) are then boiled in the presence of hops, and bitter acids are extracted from them and converted into resins. The wort is then cooled by water-powered coolers, at which time proteins and tannins separate from the hot resins. Yeast is then added and fermentation takes place. Fermentation adds fruitiness and fullness to the liquid and converts the sugars into alcohol. Yeasts may ferment on the surface, something that is characteristic of most British brewing, or at the bottom of the fermentation vessel, as is more common in Europe. A problem with top fermentation has always been the susceptibility of the yeast to breeding with wild airborne yeasts. This often meant that brewing was difficult or impossible in the summer. A solution was found by brewers in Europe in the late 1880s, who developed a yeast that sank to the bottom of the fermentation vessel. Bottom fermentation is generally thought to produce a clean, soft, non-fruity beverage.

The Importance of Water

Great volumes of water are involved in all stages of the brewing process, which is why so many breweries were originally sited near major rivers, though it will be a great relief to many to know that, despite suggestions to the contrary, river water was not necessarily used in the beer itself.

It is possible to brew beer anywhere, from any type of grain and water, but historically the classical area for brewing is the Beer Belt of Europe. This belt essentially comprises six countries – Ireland, the UK, Belgium, Germany, the Czech Republic, and Slovakia – but also includes, to the south, parts of Austria (Vienna), Switzerland, and northern France and, to the north, the Netherlands, Denmark, and Sweden.

The greatest contribution to the chemistry of the brewing liquor comes from the brewing water, which forms a vital and integral part of the nature of the final product.

There are four anions that are particularly significant, of which Calcium is by far the most important. It has three major effects. It stabilizes the enzyme α-amylase and helps the breakdown of starch from the malt in the mash tun and in later processes. It precipitates phosphate and thus increases the acidity of the wort, which is important because acidity influences the strength and character of the fermentation and the microbiological stability of the enzyme processes. Lastly, it promotes flocculation (clumping together) of the yeast during fermentation. Magnesium produces a sour to bitter taste, but retards phosphate precipitation, which in turn stops the required drop in pH level. Sodium in small amounts gives a salty to sour taste, and potassium, whilst also contributing a salty flavour, can be particularly laxative above 10 ppm.

Although, as just described, it can be seen that the anions are important, it is really the cations that have the major influence, and this will be illustrated, first, using examples from the UK, before extending the logic to cover the rest of Europe.

Brewing in the UK

Brewing was established in Burton-on-Trent in the English Midlands in the sixth century, when beer was brewed by the monks of the local abbey. They drank beer rather than the more suspect river water. The brewing waters at Burton-on-Trent (Table 1) are characteristically very high in sulphate (SO_4), derived from the Triassic gypsiferous marls in the area. Sulphate is very important for the brewing of bitter beers because it helps in the degradation of proteins and

Table 1 Brewing-water analyses

TDS	*Calcium*	*Magnesium*	*Potassium*	*Sodium*	*Sulphate*	*Chloride*	*Carbonate*	*Location*
1401	283	90	0	29	725	54	171	Burton-on-Trent
670	111	38	2	83	168	62	162	Wolverhampton
428	109	5	5	29	69	37	128	London
305	100	16	0	0	17	17	150	Dublin
800	140	36	0	92	231	60	210	Edinburgh
1011	260	23	0	69	283	106	270	Dortmund
273	80	19	0	1	5	1	164	Munich
31	7	1	0	3	6	5	9	Pilsen

All figures are in parts per million.
TDS, total dissolved solids.

starch and allows the full extraction of bitter oils from the hops. It also reacts with magnesium carbonate to give magnesium sulphate, which is in itself bitter. The high proportion of hop oil acts as a preservative, as well as a flavouring, and consequently it was this type of beer that was exported to the colonies in the nineteenth century as India Pale Ale.

To the west of Burton-on-Trent, in the Birmingham and Manchester areas in particular, the predominant evaporite deposit in the Triassic red beds is halite rather than anhydrite, and the composition of the ground water reflects this difference (Table 1). The salty to sour taste provided by the sodium ion combines with particular chloride sweetness giving, at levels under 200 ppm, what brewers term palate fullness. The brews are still reasonably high in sulphate, though, as it is lower than in the Burton waters, they cannot be as highly hopped. The resulting sweeter beer, which still retains an amount of bitterness, is the classic mild ale of these areas.

In southern and eastern England, where Cretaceous chalk is the main aquifer, the groundwater has a very different composition. It is, as would be expected, dominated by carbonate (Table 1). Clearly the use of water of this composition for brewing presents problems, primarily with malting and fermentation, because of the difficulty of keeping the pH low. Additionally, the lower sulphate means that the extraction of oils from the hops is poor, giving a less flavoured brew. A brew was developed in the 1770s that was suited to these waters, and this was a sweet, dark beer in which the dominant flavour came from roasted malt rather than hops. This drink was known as porter because of its popularity in the London markets of Billingsgate and Covent Garden.

Brewing in the Rest of Europe

In Ireland at about the same time attempts to brew bitter beers had failed, though it was found that the waters used in brewing there were ideal for beers of the porter type. The reason is that most of the waters used in Dublin and Cork in particular are derived from the Carboniferous limestones in the centre of the country and are thus dominated by carbonate (Table 1). The most famous of these brews was developed by Arthur Guinness in Dublin, who produced his extra stout (meaning thick) porter, which on contraction introduced the word 'stout' into the language of brewing.

Many countries in the European Beer Belt use mineralized groundwater for their characteristic brewing, and this is highlighted by the following examples. Two of the classic brewing areas of Germany are Dortmund and Munich, where mineralized groundwater determines the brew type. In Dortmund (Table 1) much of the water is derived from the Coal Measures, where waters similar to those in Burton-on-Trent are found, though they are not as strongly mineralized and have a higher carbonate content. The classic 'Dortmunder' is bottom-fermenting, as one would expect in Europe, has enough sulphate to be well hopped, and has a high enough total mineral content to promote good fermentation and thus produce a clean, dry beer. Munich on the other hand draws water from more recent rocks, which, as can be seen from Table 1, is closer to the water of Dublin in chemical composition. The classic Munich beer is sweet, full, dark, and brown, to all intents and purposes the equivalent of stouts and porters. Denmark, particularly in Jutland, has an essentially Cretaceous geology overlain by Tertiary deposits. The naturally occurring groundwater is similarly rich in calcium carbonate and the response to this is the brewing of sweet, dark beers termed Stowts.

However, in much of Europe the groundwaters, whether derived from surface run-off, sandstones, or impermeable lower Palaeozoic or Precambrian strata, have very low levels of both calcium and sulphate. In these areas the beers that are produced are alcoholically strong but lightly hopped and texturally thin. This type of beer was developed initially in

the town of Pilsen (Table 1) in Bohemia (now in the Czech Republic). A critical process was the storage of the fermented beer in cold caves – lagering – and thus two new terms were introduced into the brewing vocabulary, pils and lager. This direct response to brewing with waters of low ion content spread rapidly in the 1880s throughout those areas of northern Europe where such waters are common. This included the Netherlands, Germany, and Denmark, where to this day lager is recognized as a major beverage.

So far, no mention has been made of another important brewing nation, and that is Belgium. The country is fascinating, as geologically it is not always very promising as far as brewing waters are concerned, but it illustrates a classic example of a unique response. Simply, in areas where there are useful mineralized waters, use them to their best effect; otherwise, ignore the rules. In East Flanders, particularly around Oudenaard, there are carbonate-rich waters, and the characteristic Brown Beers are top-fermenting, dense, and smooth. These contrast markedly with the Red Beers of West Flanders, where the less mineralized waters lead to the brewing of light and not very bitter beers. There are four monastic Trappist breweries in the Ardennes, where the waters are highly enriched in calcium carbonate and as a result the typical brews are top-fermenting, dark brown, and sweet. There is a fifth brewery at Westmalle, north of Antwerp, where the water is hard and a brew much closer to a bitter ale is produced. Where the waters are poorly mineralized there are only two options. First, there is the classical approach into pils brewing, typically around Louven, Alken, and Lindberg where the groundwater is derived mainly from superficials and surface run-off. Second, there is a much more idiosyncratic Belgian approach, which is the production of Lambic, traditional in the Brussels area. In this style wheat and malted-barley beers are produced using naturally occurring wild yeasts, which ferment completely and are very dry. Hops are used not for flavour but purely to protect the beers from unwanted infection. Fruit flavourings are often introduced, the best known being black cherry (Kriek) and strawberry (Framboise).

Scandinavian brewing is dominated by lager-style beers, as would be expected from poorly mineralized waters derived mainly from Precambrian rocks, though occasionally a Belgian style approach is attempted, as with the juniper-flavoured beers of Finland.

Modern Brewing

Many breweries are turning to the use of mains water that is deionized and then reconstituted to a given composition for a particular brew. Commercially there are advantages in this procedure, specifically related to purity and continuity of supply. Where a water grid system is in place a major disadvantage is that the composition of the water entering the brewery can change very rapidly as sources are changed, and this completely alters the nature of the brew in progress. Wells and springs are still used by some breweries, but the problems of microbiological pollution, increasing nitrate content, and, near coasts, overpumping causing seawater intrusion are gradually reducing this number.

An interesting thought provided by one brewer is that, despite all that has been written about water type, the siting of breweries initially may have been controlled not by the analysis of the water but by the presence of a source of abundant water at a constant temperature in the days before thermometers were in constant use.

See Also

Geology of Whisky. **Geology of Wine**. **Minerals:** Carbonates; Sulphates. **Palaeozoic:** Carboniferous. **Sedimentary Rocks:** Chalk; Limestones.

Further Reading

Jackson M (2000) Great Beer Guide. London: Dorling Kindersley.

GEOLOGY OF WHISKY

S J Cribb, Carraig Associates, Inverness, UK

Introduction

The distillation of liquors produced by the fermentation of cereals is a process that goes back into the mists of history. There is historical evidence from Scotland dating the distillation of malted barley from the middle of the sixteenth century. The distillate was known originally by its Gaelic name *uisge beathe* ('water of life'), which has become the present-day word 'whisky' (spelt variously in many countries with or without an 'e'). The distilling process was legalized in Scotland in the 1830s, and since then the manufacture of this special liquid, whisky, has been developed to a fine degree, as is evident in the array of unique products available today.

The manufacture of whisky occurs in many countries, but the pinnacle of success with this product is found in Scotch whisky, a specific product of Scotland. To qualify as Scotch whisky, the spirit must be derived from malted and/or unmalted cereals and be matured in oak barrels in Scotland for at least 3 years. Whiskies fall into three categories: grains, malts, and blends. Grain whisky is manufactured generally by a continuous industrial process from malted and unmalted cereals. Malt whiskies are made totally from malted barley. Blended whiskies comprise a base of between 40 and 60% grain whisky overlain by up to 30 malts. Malt whisky represents probably the pinnacle in the development of the whisky-making process and of the 100 or so distilleries in Scotland, around 75 are producing at any one time. Each of these distilleries produces a unique spirit characterized partly by the unique source of water for its production, and it is in the chemistry of this that the varying geology of the water sources comes into play.

Whisky Production

The whisky-making process essentially comprises a brewing phase, followed by the distillation of the resulting liquor. In fact, the result of the first stage is the production of the equivalent of a strong beer, though unflavoured by hops or other additives. The first part of this process is the malting of barley, whereby the grain is encouraged to germinate for several days to initiate the enzyme processes that break down starch to sugars, before the process is stopped by heating. Traditionally, where the malting takes place at a distillery site, a certain amount of peat smoke is allowed to permeate the malt during the heating process. The level of permeation is reflected in the smokiness of the final product. Nowadays there are larger scale centralized maltings serving several distilleries, but the process of peating is essentially the same. The second stage is the mixing of the malted barley with fresh potable water to extract the sugars and produce a liquor ready for fermentation. Fermentation takes place using a combination of beer and wine yeasts to raise the alcohol content to around 8%. The resulting liquid is then heated in the distinctive onion-shaped copper stills, normally in a two-stage process that raises the alcohol content, first to 25% and second to the final spirit alcohol content of between 55 and 65% alcohol. The raw spirit is stored in oak barrels of various types: new wood, European oak barrels generally from the sherry industry in Spain, and American oak bourbon barrels. Thus the four crucial factors that combine to produce the wide variety of malt whiskies are the degree of peating of the malt, the shape and type of still, the types of barrels used for maturation, and the chemistry of the process waters. It is with the latter step that the geological influence on whisky type is exerted.

Water chemistry affects three areas of the production process: the extraction of sugars from the malt, the nature and type of fermentation, and the chemistry of the distillation process (which is largely unstudied). There is an oft-quoted maxim that the best type of water for the production of malt whisky is soft water through peat over granite. Even a cursory comparison of the siting of malt whisky distilleries with the geological map of Scotland will show that only in comparatively rare instances is it likely that such water will be available. In fact, only 20 distilleries, most of which are centred in the Speyside area of north-east Scotland, have such water. The remaining rock types from which water is sourced cover the whole spectrum from sandstones and schists to quartzites and shales. Clearly the chemistry of the waters derived from such a range of sources will show significant variations up to and including significantly heavily mineralized (hard) waters.

Distilleries

Distilleries come in all shapes and sizes and they are situated on the coast, inland, on hills, in valleys, by mighty rivers, near small streams, and on islands

Figure 1 Clynelish Distillery at Brora in Sutherland.

Figure 2 The Tarlogie Spring; the source of hard water used by the Glenmorangie Distillery near Tain in Ross-shire.

(Figure 1). They use water from boreholes, rivers, springs, and lochans (small lakes), and every water source is unique in terms of its chemical make-up; distillery waters vary significantly not only in the chemistry of the major elements, but even apparently similar waters have markedly different trace element profiles (Figure 2). The study of malt whisky is in a very real sense a study of the geology of Scotland.

Starting with the Speyside district of north-eastern Scotland, centred around the villages of Dufftown and Rothes, there are over 20 distilleries in an area of approximately 100 miles^2. In this area are some of the most famous distilleries of all. The whiskies fall easily into three groups, based on taste and aroma, and these groups correspond very closely with the geology of the water sources. The first group taking its water from the granites to the west of Dufftown include Glenfiddich, Balvenie, and Aberlour. The second group, in the west and north-west of the region, draw spring waters from the well-bedded, light grey, Dalradian quartzites of the Grampian group, and include Glen Grant and Knockando. The third group takes water from a wider range of Dalradian quartzites, phyllites, and schists. Among these are Glenlivet and Strathisla.

Moving to the west of Speyside, in the area called the Grampian Highlands, which lies between the Highland Boundary Fault to the south and the Great Glen Fault to the north, is an area that stretches from the islands of Islay and Jura in the south-west to the Buchan coast in the north-east. Within the area there are many igneous intrusions of granitic, basic, and ultrabasic composition. The mountain of Lochnagar provides water for the Royal Lochnagar distillery, which is situated behind Balmoral Castle. At Knock Hill, east of Keith, the water seeps out between the Dalradian schists and the underlying gabbro. The whisky produced has the Gaelic name An Cnoc. To the south around Insch, 17 springs provide the ultrabasic water source for Ardmore, an unusual whisky.

A drive along the main road north from Perth to Inverness highlights not only the change in geology but also the variation of spirit and water source, in a comparison of five distilleries. Starting in the rich agricultural land just north of Perth, underlain by the Old Red Sandstone, the Highland Boundary Fault is soon crossed at Dunkeld and immediately the scenery becomes mountainous as the Dalradian succession is encountered. North of Dunkeld is the town of Pitlochry, home to the Blair Atholl distillery, and close by is the small distillery of Edradour. Both of these draw their water from high in the schistose hills around Ben Vrackie. Northwards through the pass of Drumochter, the highest distillery in Scotland is reached at Dalwhinnie; the water source is the clear blue sparkling mountain waters of the Lochan na Doire-uaine. Further north, near the town of Newtonmore, is the distillery of Drumguish, and further still, near Carrbridge, is the huge distillery of Tomatin, taking its water from the Allt na Frith, a stream that bounds down a rocky juniper-filled valley, flowing through the central Highland migmatites.

To the far south-west of Scotland lie the Argyll islands of Islay and Jura, known particularly for the very heavily peated malts of Laphroig, Lagavulin, and Ardbeg. However, there are other whiskies that, being less peated, reflect more closely the differences in geology of their water sources. The furthest north is Bunnahabhain, the only distillery on the island to take its water from underground springs rather than from surface streams. These waters, rising as they do through Dalradian dolomites, are hard, containing substantial amounts of calcium and magnesium carbonates. This type of hard water, combined with the use of totally unpeated malt, produces a markedly unique whisky, and the water composition is considered a major influencing factor. In the far west of Islay is Bruichladdich, the most westerly of all distilleries in the United Kingdom, taking its water from the dark brown sandstones of the Colonsay group, a group for which the affinities are not yet completely understood.

An area of Old Red Sandstone that runs along the north-east coast of Scotland, taking in Inverness, Dingwall, and Dornoch, includes the Caithness towns of Wick and Thurso and the Orkney Islands in the far north. The waters from these sandstones generally have a high calcium carbonate hardness, something that the distilleries in the area claim has a significant effect on the nature of the whisky that they make. The red sandstones can, however, create problems, because they break down to form excellent farmland, and where there is intensive agriculture there is often a problem with the high nitrate content of groundwaters. Indeed, several distilleries within the area have closed particularly because of water problems. The most famous distilleries in the region are Glenmorangie in Tain, Clyneleish in Brora, Old Pulteney at Wick, and the most northerly distilleries of Highland Park and Scapa on Orkney.

Far to the south, beyond the Highland Boundary Fault, lies the Midland Valley of Scotland. At the northern edge, close to the Highland Boundary Fault, there are Old Red Sandstones similar to those in the north of Scotland. The distilleries in this area may be divided into two groups, based on their water sources. First, there are those that are situated in the Lowlands, taking their water from either across the Highland line or from the red sandstones just to the south, such as Fettercairn. Second, there are those that utilize a variety of water sources from rocks within the Midland Valley, such as Auchentoshan on the Clyde (Carboniferous volcanics), Glengoyne (Devonian volcanics), and Glenkinchie south of Edinburgh, which takes its water from south of the Southern Uplands Fault from a reservoir in Ordovician volcanics. In western and north-western Scotland, the

lavas of the Tertiary Volcanic Province provide surprisingly hard waters; these are seen in Talisker on Skye and Tobermory on Mull and across the water at Bushmills in the plains of Antrim in Northern Ireland. Rarely does a new distillery open, but the latest to do so was built in 1995 at Lochranza in the north of the Isle of Arran. The Isle of Arran single-malt whisky is well into production and in one sense closes the circle with regards to a water source, in that it utilizes soft waters from the Easan Biorach, which flows over granite, having passed through thick peat beds.

Waters used in whisky making, though more esoteric than brewing waters used for beer, are certainly less studied, and a full understanding of the effects of water chemistry on whisky making has still to be achieved, but there can be no doubt that the effects are significant and demonstrable.

See Also

Engineering Geology: Ground Water Monitoring at Solid Waste Landfills. **Geology of Beer**. **Geology of Wine**.

Further Reading

Cribb S and Cribb J (1998) *Whisky on the Rocks*. Keyworth, Nottingham: British Geological Survey.

GEOLOGY OF WINE

J M Hancock[†], Formerly Imperial College London, London, UK

Introduction

Geology can be important for the growers of grapes for wines; it can be ignored by the consumer. Most of the arguments for the importance of '*terroir*', which includes geology, have been along the lines: "This wine has one flavour, that wine has another flavour, yet they are made from the same grape variety". The difference must be something in the ground, i.e., it depends on the geology. Geologists have known for a long time that this is a misguided simplification. In the mid-nineteenth century, a distinguished French geologist, Henri Coquand, played a joke on the wine producers by stating that the quality of cognac (which is a distilled product anyway) was directly related to the quantity of chalk in the ground. The zones of quality for cognac are arranged in circles centred on Cognac: the strata have a roughly linear strike NW–SE. The best cognac comes from an area in which chalk is largely absent. In spite of the obvious nonsense of a relationship between chalk and the quality of cognac, the idea is still being quoted today in some books.

As with any agricultural product, there are many factors that control the quality of wine. Geology can play a part in three of these: the temperature around the vines in general and the bunches of grapes in particular; variations in porosity and permeability around the roots of the vine, which will affect both the supply of moisture and the rate at which the vine can take up nourishment through its roots; and variations in the composition of the ground, which will control the availability of nourishment supplied through the roots.

Temperature

It was known to Lucius Columella in the first century AD that the quality of grapes depends on the temperature around the vines. All modern work on the relationship between heat and grapes originates from Amerine and Winkler in California, who showed that each grape variety requires its own heat regime to bring out the best of its qualities. Amerine and Winkler worked in California where, at that time, vines were kept around 1.1 m above nearly level ground. This meant that it was the general ambient temperature of the district which was the controlling factor. However, in many vineyard regions, the geology affects the mesoclimate, principally by its control of the topography.

1. *Shelter from cold winds*. In more northerly vineyards (or more southerly vineyards in the southern hemisphere), the better vineyards are located on slopes which face south or south-west. Vineyards are not extended to the top of the hill, so that the summit can act as a break from cold winds from the north and north-east. It is even better if there is a clump of trees on the summit. This sheltering effect is of value even if the local slopes in the vineyard are to the east or north.
2. *Thermal belt on slopes*. With a layer of cold air at the bottom of a valley and a layer of cold air near the ground over the plateau above the slope, an

[†]Deceased

intermediate zone, known as the thermal belt, is developed on the slope. Although somewhere in the middle of the slope, its actual height is highest in winter and lowest in summer. Its position on any one hillside is remarkably constant. Sites only 2 or 3 km apart horizontally may show differences of 8°C over height differences of less than 100 m. In vineyards, the main importance of a thermal belt is often the maintenance of slightly higher temperatures during the night.

3. *Radiant heat from the sun* (Figure 1). The radiant heat that a plot of ground receives from the sun depends on the angle it presents to the sun. In simplified form, the relationship is:

$$I = k \sin(\alpha + \beta)$$

where I is the intensity of radiation received on the slope, k is a constant, α is the angular elevation of the sun, and β is the angle of inclination of the slope to the horizontal along a meridian, i.e., to the south in the northern hemisphere, and to the north in the southern hemisphere.

Thus, the steeper the slope, the more radiant heat the vineyard will receive (Figure 2). As the direction of slope is angled away from the south (in the northern hemisphere), the less radiant heat the vineyard will receive. On a cloudy day in January in central Europe, a slope of 20° facing south receives roughly twice as much radiant heat as a flat surface.

This effect is well illustrated in the Mittelmosel. Thus, near the village of Piesport, the best vineyards on the steep, south-facing slopes of up to 30–35° on phyllites are on the north flank of the river. In an area of less than 3 km × 1 km, there are seven Einzellagen (individually named vineyards and wines), distinguished by modest differences of angle and direction of slope, but each of high quality (the most famous Einzellage – 'Goldträpfchen' – is not on the steepest part of the slope because of the position of the thermal belt). The whole of the flat area south of the river is lumped into a single Einzellage, Treppchen, producing quantities of undistinguished wine, with lower concentrations of alcohol and less flavour.

4. *Re-radiation of heat from the ground*. Because the ground is kept clear of weeds in most vineyards, the effects of re-radiation of heat can be considered in terms of bare soil. The albedos (diffuse reflectivity) of most rocks, including soils, are in the range 0.1–0.3 (i.e., 10–30% of radiation from the sun is reflected back). Sandy soils have higher albedos than clays; dry ground has a higher albedo than damp ground. Hence, a dry sandy soil reflects back about twice as much radiation as damp clay. Dark rocks absorb radiation more easily than light-coloured rocks, and then re-radiate when the sun sets or is hidden by cloud. The importance of this re-radiation process is enhanced by the fact that the re-radiation has a longer wavelength, i.e., is more warming, than the incoming radiation.

The value of re-radiation is well known to vineyard managers. In the Schlossböckelheim district of the Nahe, the growers pile pebbles and boulders of the feldspar porphyry beneath the rows of vines. In the Mosel, the dark phyllite, with a particularly high albedo, has long been used to enhance this effect.

However, it should be noted that the strength of re-radiation is inversely proportional to the square of height above the ground, i.e., doubling the height of the grapes quarters the energy received by them. In Chateauneuf du Pape, where many vineyards are covered with pebbles of quartz, many of the bunches of grapes are no more than 0.2 m above the ground. This practice is even more marked for the sugar-rich grapes in the Montilla-Morilles district of south-west Spain.

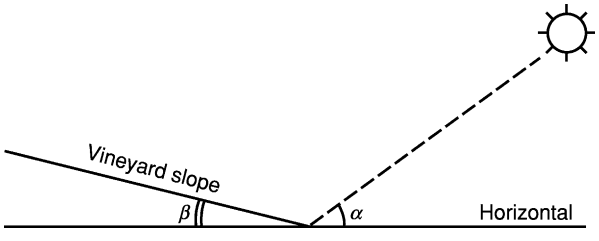


Figure 1 $I = k \sin(\alpha + \beta)$, where I is the intensity of radiation received in the vineyard, α is the angular elevation of the sun, β is the angle of inclination of the vineyard to the horizontal along a meridian, and k is a constant. © Jake Hancock.

Figure 2 Vineyard of Schloss Böckelheimer Kupfergrube, north of Oberhausen, southern Nahe, Germany. An example of a European planned vineyard, only developed in the early twentieth century, on a steep south-facing slope to gain maximum advantage of direct radiation from the Sun in a relatively cool climate. © Jake Hancock.

Factors 1–4 can enhance the temperature during the growing season. There is also a slope factor during spring budding.

5. *Drainage of cold air.* The woody stems of vines can withstand extremely low temperatures but, once buds start to break from the stem, the vine is vulnerable to damage by even modest frosts. Bud-break is usually taken to be at 10°C but, depending on the variety of vine and the amount of nutriment stored from the previous season, budding may start at 3.5–7°C. Keeping in mind that the dates of the 'ice-saints', 11–14 May, are long after bud-break in northern Europe, late frosts can be a serious problem. The slope needed to provide a natural protection from still cold air depends on how severe the frost will be. In the Napa Valley in California, a slope as little as 2.5° may be sufficient to allow the freezing air to drain away. In Chablis, in northern France, a slope twice as steep as this may not be adequate to cope with the more severe frosts.

Water Balance

Like all plants, vines need water in the ground. However, without actually being a species of arid climates, its need for water is small. As a result, vines will grow almost anywhere. Nevertheless, to produce a good crop of grapes, the plant needs a small steady supply of moisture from the ground.

Without artificial supplies of water, e.g., by irrigation, the ideal drainage involves a high porosity for storing water, a low matrix permeability to stop it draining away, and a high mass permeability to ensure that excess rainwater drains away fast. A facies with ideal drainage is chalk, a distinctive, very fine-grained limestone, which occurs in much of Champagne. Typical chalks have a porosity of 35–45%, a matrix permeability of 2–6 mD, and a very high, but variable, mass permeability (normally more than 150 mD and can be several thousand millidarcy, i.e., more than a darcy).

For the vine, such a high porosity and low matrix permeability mean that plenty of water is held in the pores of the sediment, which cannot escape under gravity because the pores are less than 30–60 μm. However, plants can exert suctions equivalent to 100 m or more of water, i.e., the vines can extract water which cannot escape under gravity. In addition, the low matrix permeability means that the chalk does not easily dry out.

However, when there is very heavy rain, the mass permeability becomes valuable: it ensures that the large volume of rainwater drains away without leaving the sediment, and hence the roots of the vine, waterlogged (Figure 3).

For simple survival of a vine, waterlogging is a greater danger than a shortage of water, because the limited root surface area can be poisoned from a shortage of oxygen in solution. If the vine has used all the oxygen in stagnant water, the root metabolism starts to form poisonous alcohols. Clearly, it is essential that even the lowest roots of the vine are well above the top of the water table. As the roots normally penetrate downwards to several metres, and can penetrate to 20 m, vineyards can only be successfully established many metres above the normal local water table. In the Médoc, land drainage is used in vineyards which are closer to the Gironde Estuary. In the Barossa, in South Australia, where low rainfall makes drip-irrigation helpful, it is essential that open reservoirs are many metres below the height of the vineyards themselves.

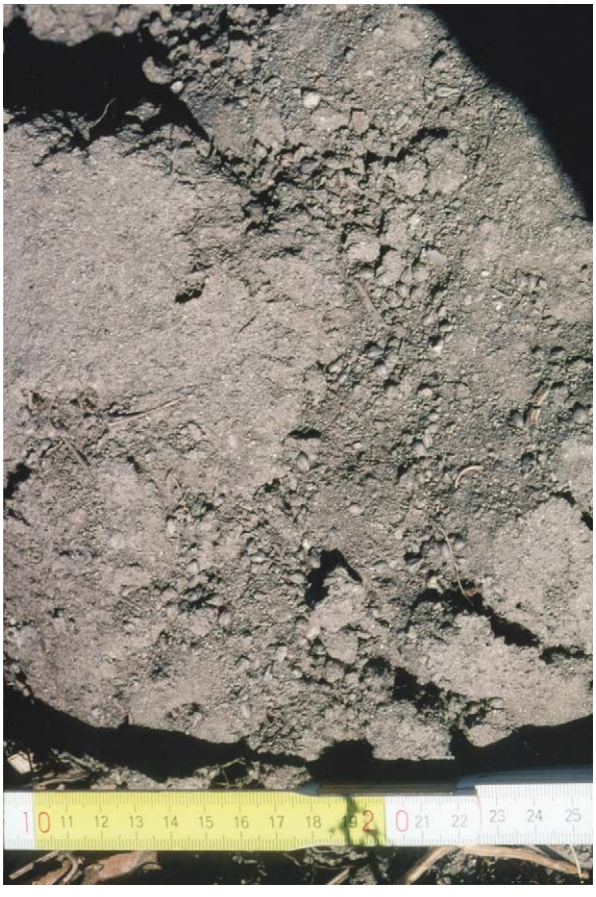


Figure 3 Muscovite-rich silty sand in a vineyard of Michel Torino, Cafayte, south-south-west of Salta, Argentina. A reliable soil for vines: medium porosity with relatively high permeability, holding water for roots, but allowing it to drain away before all dissolved oxygen has been lost. It will need some nitrogen for optimal growth. The muscovite, although not a direct source of potassium itself, probably indicates that adequate potassium is present from the weathered muscovite and illite. © Jake Hancock.

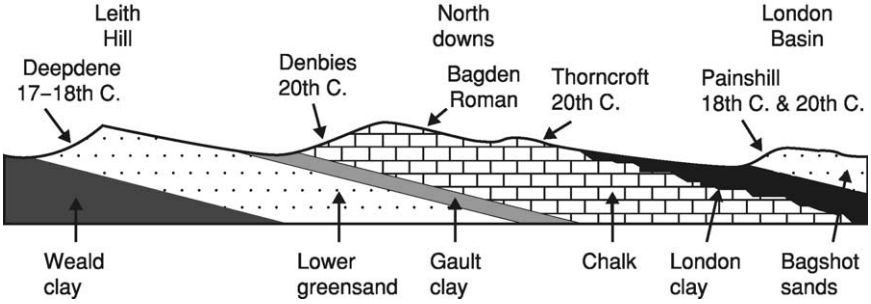


Figure 4 South-north cross-section from the Wealden anticline into the London basin of Surrey, south-east England, showing the location of vineyards ancient and modern. Note that though they planted on rocks of different ages and types, vineyards have been located on well-drained sunny south-facing slopes. From Selley (2004). Courtesy of Petravin Press.

There is another complication with the drainage for vines because of their long roots. In many vineyards, most of the nourishment is obtained in the top 0.5 m, i.e., from the soil. Moisture for everyday metabolism can be obtained from much lower levels, possibly from superficial sediments, but commonly also from bedrock. Each of these three may have their own values of porosity and permeability. Of course, the roots take in nourishment and water from all levels, but the soil is likely to have the most nourishment available, and the bedrock may have the most steady and reliable supply of simple water. The control of geology on vineyard location is demonstrated in northern Surrey, England. Here vineyards have been planted over two millennia on rocks of various ages and types, but always on well-drained sunny south-facing slopes (**Figure 4**).

Vine Nourishment

Many of those who believe in geological '*terroir*' feel that it must be something in the nourishment from the ground which gives character to the wine. However, of the three geological influences, nourishment is the least important. Only the three major elements required by plants will be considered here.

1. *Nitrogen*. A nitrogen-rich soil makes for a flabby wine. In practice, a deficiency in nitrogen is the most widespread problem in the nourishment of vineyards. However, geology has little to contribute. There is some nitrogen in primary rocks, typically 10–12 $\mu g\ g^{-1}$, but almost all the nitrogen for plants is supplied as NH_4^+, NO_2^-, and NO_3^-, formed in a complex series of actions by heterotrophic bacteria and fungi in the soil.
2. *Phosphorus*. The most common phosphorus mineral is fluorapatite, usually just named apatite: $Ca_5(PO_4)_3F$. Biogenic apatite usually contains some OH^- and/or CO_3^{2-}. The solubility is very low: a typical soil contains 0.3 mg of phosphorus per litre of soil solution, most effectively as $H_2PO_4^-$; the double negative charge of HPO_4^{2-} makes it less effective because of the negative charge on the roots.

   Vines need very little phosphorus and a phosphorus deficiency is almost unknown. Nevertheless, the relatively small surface area of the roots means that vines would have difficulty in obtaining sufficient phosphorus, but for the fact that they belong to the group of plants which live symbiotically with mycorrhizal fungi. The numerous hyphae of the fungus penetrate the cortical cells (but not the protoplast) and form complex branches which increase the absorption area of the plasma membrane. This allows the root to absorb relatively large amounts of phosphorus. In return, the vine gives the fungus carbohydrates and vitamins. The effect is enhanced still further by the concentration of phosphorus in the fungi themselves, which is about three times the amount of phosphorus in the vines.
3. *Potassium*. Geology is a major controlling factor for potassium in vines. The plant takes up the element as the ion K^+. The vine seems to have no control on how much K^+ it absorbs: if there is a shortage of K^+, the vine suffers; if there is an excess of K^+ available, the roots simply absorb it.

   The most common potassium minerals are: muscovite, $K_2Al_4[Si_6Al_2O_{20}](OH,F)_4$; biotite, which differs in basic composition from muscovite only in containing Mg_6 instead of Al_4; K-feldspar (e.g., orthoclase and microcline), $K(AlSi_3O_8)$; and the clay mineral illite, $K_yAl_4(Si_{8-y}Al_y)O_{20}(OH)_4$ where $y < 2$.

   Muscovite is resistant to weathering, biotite is quantitatively insignificant, K-feldspar is important in some vineyards (see below), and illite is widespread and can yield K^+ by direct proton replacement. However, in most vineyards, as with

most agricultural products, K^+ comes not from the weathering of primary minerals, but from ions held by electrostatic attraction to the negatively charged colloidal-sized particles of clay minerals. The ease with which a clay mineral can release cations back into solution for plant roots is known as the cation exchange capacity (CEC). The CEC of the smectite group clay minerals is high (47–162 $cmol_c\ kg^{-1}$), and this can be valuable in some vineyards, e.g., on the weathered microgranites of Beaujolais, and possibly in soils derived from weathered volcanic rocks in California. The CEC of the illite group is only 20–40 $cmol_c\ kg^{-1}$, but illite is more important in most vineyards because of the high illite content in the clays and marls of districts such as Sancerre and the Côte d'Or of Burgundy. Moreover, as the surface K^+ is removed by the roots, some K^+ can be replaced from the illite itself.

There is a common myth that, because K^+ has a high solubility, it is easily washed out of a soil. This ignores the power of illite and smectite particles to seize K^+. A single dose of a potassium fertilizer can affect the fertility of the ground 50 years after it has been applied.

Although cationic exchange from illite is probably the main source of K^+ in vineyards, a remarkable number of high-quality wines come from vineyards which may obtain some K^+ from the weathering of K-feldspar: the Méric Conglomerate in the Médoc; the granite of Hermitage; the feldspathic sandstones of the Permian Rotliegend west of Nierstein in the Rheinhessen, and some of the grands crus vineyards in Alsace; the feldspar porphyry of Schlossböckelheim in the Nahe; the alluvial sediments in the Napa valley derived from volcanic and pyroclastic rocks lining the sides of the valley; and the lateritic soils developed on granite–gneiss in the Margaret River country of western Australia.

Summary

The fact that vines grow on rocks of all ages and types may superficially suggest that geology has no role to play in viticulture. Nothing could be further from the truth. The interplay of geology and climate determines the landscape within which a vineyard stands, and the soil on which it grows (Figure 5).

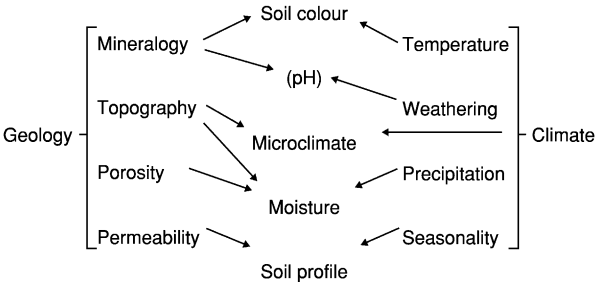


Figure 5 Soil, referred to as 'the soul of the vine' (Wilson, 1998) results from the complex interplay of geology and climate. Note that in many vineyards it is the superficial deposits, such as alluvium, eolian sand or boulder clay that determine the character of the soil, as much as the solid rock beneath. From Selley (2004). Courtesy of Petravin Press.

See Also

Geology of Beer. **Geology of Whisky**. **Sedimentary Rocks:** Chalk; Clays and Their Diagenesis; Phosphates; Sandstones, Diagenesis and Porosity Evolution. **Soils:** Modern.

Further Reading

Amerine MA and Winkler AJ (1944) Composition and quality of musts and wines of Californian grapes. *Hilgardia* 15: 493–675.

Asselin C, Flanzy C, Sapis JC, and Martin JC (eds.) (1997) *Les Terroirs Viticoles – Concept, Produit, Valorisation*. Angers: Institut National Recherches Agronomiques.

Coomb BG and Dry FR (eds.) (1988) *Viticulture, 1. Resources*. Adelaide: Winetitles.

Geiger R (1961) *Das Klima der Bodennahen Luftschicht* (translated as *The Climate Near the Ground*). Cambridge: Harvard University Press.

Gladstones J (1992) *Viticulture and Environment*. Adelaide: Winetitles.

Hancock JH and Selley RC (2003) *Coquand's joke*. London: Geoscientist. 13.10.17.

Jackson RS (2000) *Wine Science*, 2nd edn. San Diego: Academic Press.

Pomerol C (1989) *Terroirs et Vins de France*, 2nd edn. (translated as *The Wines and Winelands of France*). London: Robertson McCarta.

Selley RC (2004) *The Winelands of Britain: past, present & prospective*. Dorking: Petravin Press.

Unwin T (1991) *Wine and the Vine*. London: Routledge.

Wild A (ed.) (1988) *Russell's Soil Conditions and Plant Growth*, 11th edn. Harlow: Longman.

Wilson JE (1998) *Terroir*. London: Mitchell Beazley.

GEOMORPHOLOGY

P H Rahn, South Dakota School of Mines and Technology, Rapid City, SD, USA

Introduction

Geomorphology is the study of landforms and of the processes that act on Earth's surface to produce these landforms. This article emphasizes the geomorphologic processes that form different types of surficial deposits. For the current purposes, the range of surface processes can be considered with reference to the following four topics: fluvial processes/floods, mass wasting/landslides, tectonic terrains, and glaciation.

Fluvial Processes/Floods

Streams play an important role in erosional processes. The surface of Earth, over geological time, is continually and gradually eroded, and many landforms are a result of stream erosion or deposition. During the erosion process, when a stream reaches base level it begins to meander (Figure 1) and carves out a flood plain, an area that is subject to flooding (Figure 2) (*see* **Sedimentary Processes:** Fluvial Geomorphology). Only a small percentage of Earth's surface is subject to flooding. The degree of inundation depends on subtle topographic features such as natural levees, oxbows, local alluvial fans, low terraces, and mass-wasting accumulations from adjacent hill slopes. Identification of these features requires detailed geomorphological scrutiny using aerial photographs and topographic maps.

A stream terrace is an abandoned flood plain that results from a lowered base level. Terraces may be paired or unpaired. The continuity of a terrace along a valley, paired on both sides, may be correlated to a former graded condition. On the other hand, progressive stream down-cutting will leave isolated unpaired terraces along the valley sides. The surface elevation of terraces cannot be correlated because the terraces were never portions of a single, continuous surface. Unpaired terraces imply continuous down-cutting accompanied by lateral erosion. The stream shifts back and forth from one side of the valley to the other, and the valley floor is gradually lowered.

Figure 1 Meanders and flood plain of the Wind River near Dubois, Wyoming.

Terraces are of special interest as flood hazards because some very low terraces may be inundated during floods. For example, alluvial stratigraphy and geomorphology were used to evaluate the inundation of low terraces along the Connecticut River during a flood that occurred in 1936. It was shown that the riverine terraces were hazardous but that the higher kame terraces were not inundated. Another example is the Missouri River. A Landsat image of Glasgow, MO in September 1992 (Figure 3A) shows the Missouri River meandering on the flood plain and the terraces between the confining bluffs. A later image (Figure 3B) was taken in September 1993, during severe flooding. The entire flood plain and some low terraces are inundated; though one terrace on the north side was above the water (*see* **Sedimentary Processes:** Catastrophic Floods).

The identification of flood-prone terraces may be complicated by changes in aggradation or degradation. Such changes can be natural processes or can be caused by humans. Cyclic episodes of aggradation or degradation can be related to overgrazing or climatic change. A low terrace may be practically indiscernible from the flood plain, and, unfortunately, geological maps rarely discern these geomorphological features. Instead, the maps simply classify bottomland as Quaternary alluvium (Q_{al}). Rivers can have complex terrace systems. For example, the active flood plain width of the lower Mississippi River is approximately 16 km, lying within a terraced valley floor varying from 40 to 200 km wide. The terraces along the lower Mississippi River were formed by degradation accompanying sea-level lowering during glacial maximum, whereas the terraces along the upper Mississippi river were formed by degradation that accompanied reduced sediment loads during interglacial time.

Planners and engineers responsible for construction projects often overlook the geomorphology and interpretation of surficial deposits. For example, the

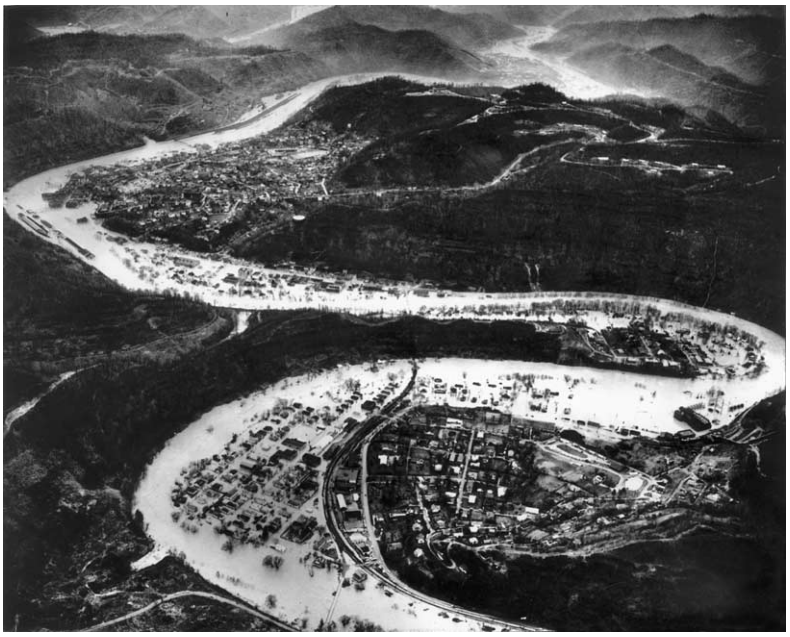


Figure 2 Oblique aerial photograph of the Kentucky River during the 1963 flood at Hazard, Kentucky. There is a clear division between the flood plain inundated by the 1963 flood and the adjacent hills. Courtesy of the Geological Society of America.

Figure 3 Landsat images of Glasgow, Missouri. (A) Image taken in September 1992. (B) Image taken in September 1993, during the flood peak. Courtesy of the United States Geological Survey.

United States Federal Emergency Management Agency requires communities to identify the so-called floodway based on the estimated 100-year flood discharge. Engineering firms use computer programs to estimate roughness and stage for this flood, and superimpose this stage ('base flood elevation') on topographic maps. The estimate of the 100-year flood discharge, the quality of the topographic maps, and the other assumptions of the computer programs limit this hazard evaluation. Further, this process does not protect flood-prone areas just above the 100-year flood stage.

An example of the usefulness of geological maps showing surficial deposits can be seen in Figure 4. In

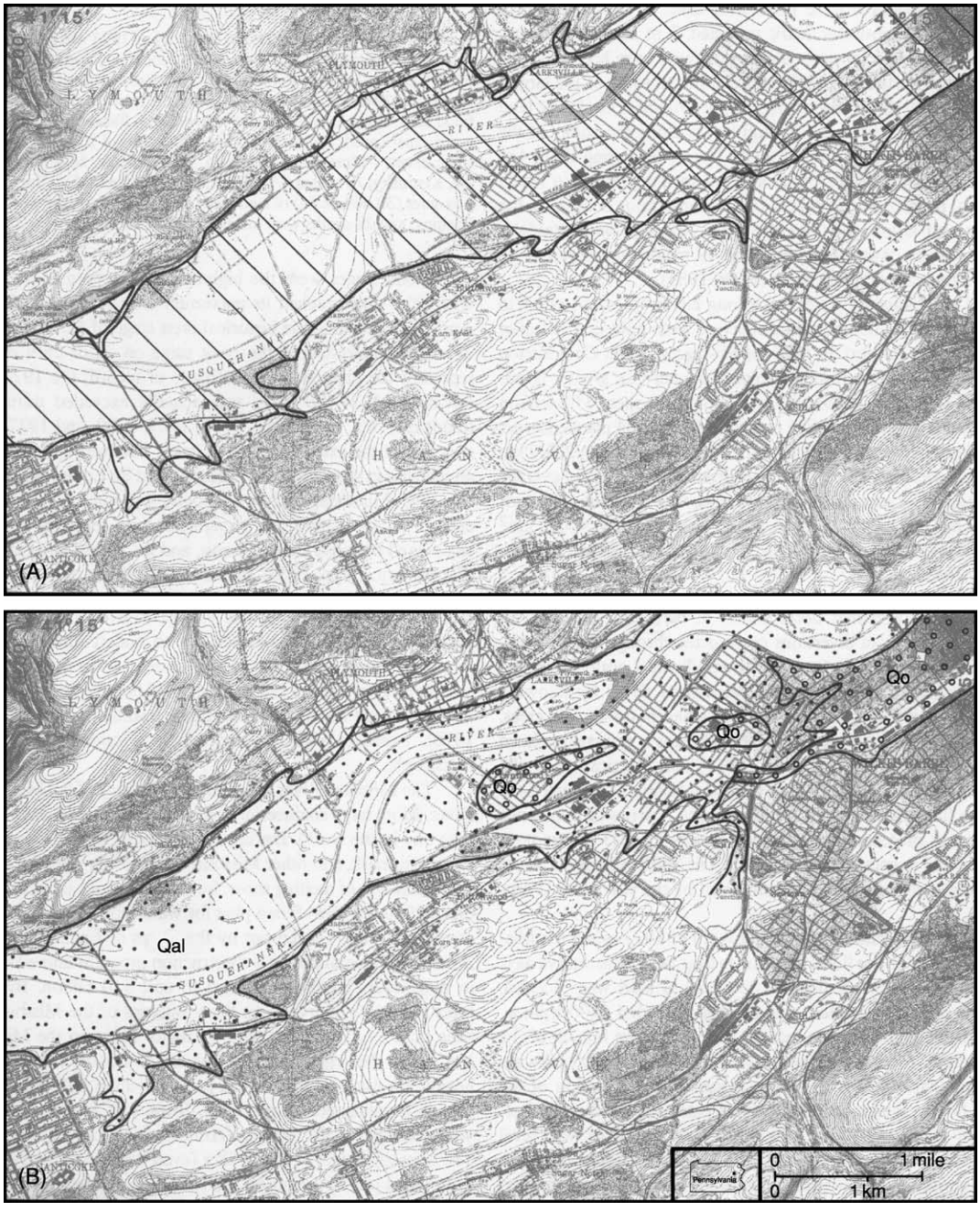


Figure 4 Maps of the area around Wilkes-Barre, Pennsylvania. (A) Portion of a Wilkes-Barre topographic map. The topography of the area has been affected by coal mining; culm banks more than 30 m high and subsidence pits are abundant. The area inundated by the June 1972 flood is indicated by diagonal lines. Courtesy of the United States Geological Survey. (B) Geological map of the same area. Quaternary alluvium (Q_{al}) and outwash (Q_o) are indicated. Courtesy of the Pennsylvania Geological Survey.

1972, at Wilkes-Barre, PA, the Susquehanna River crested at 12.4 m and the downtown area (Figure 4A) was inundated. The 1972 floods, resulting from Hurricane Agnes, caused over $3 billion damage in the northern Appalachian Mountains region. A geological map of Wilkes-Barre (Figure 4B) shows glacial outwash overlain by Holocene alluvium within the alluvial plain of the Susquehanna River valley. The coincidence of the two maps shows the utility of good topographic and surficial geological maps.

Mass Wasting/Landslides

Mass wasting includes several gravity-driven processes that act on Earth's surface. These hill-slope processes deliver material to a stream that ultimately carries the eroded debris to the sea. Mass-wasting processes include debris flow, landslides, and creep. Debris flows typically originate in semi-arid areas where thick soils or surficial deposits begin moving as a landslide and become more mobilized as they move into a stream channel. Eventually, they flow like wet concrete. Many alluvial fans include areas covered by prehistoric debris-flow deposits and hence are susceptible to continued activity. The levees associated with these deposits may get overtopped and exceed the stage of a 100-year flood as determined by stream discharge probability. Areas susceptible to debris flow require detailed geomorphologic analysis to predict the degree of hazard. A rock glacier is a form of debris flow; it is typically found in a glacial cirque and may have an ice core. Figure 5 is a radar image 20 km east of Seattle, WA. High-resolution light-detection and ranging (lidar) topography is a technology utilizing aircraft rangefinders to determine distance with a laser pulse. Topographic features are enhanced, even in forested terrain.

Landslides can occur within minutes, but many move very slowly, even over a period of years. Figure 6 shows some of the many landslides that occurred 130 km from the epicentre of the 1964 Alaska earthquake (M = 8.6). The landslides developed in the Bootlegger Cove Clay, a Quaternary marine clay. Careful mapping of surficial deposits is an important job for the engineering geologist in earthquake areas. Landslides are particularly common on steep hill slopes with clayey soils, or where a hill slope is oversteepened by shoreline processes or a river meander. Above a wave-cut cliff, an entire hill slope may be a slowly moving complex of landslides undergoing progressive failure. Translational landslides typically occur along bedding planes of moderately to steeply dipping sedimentary rocks. Geological maps, aerial photographs, and detailed fieldwork are needed to identify landslide hazards (*see* **Sedimentary Processes:** Landslides).

Gravity continuously acts on a hill slope, and weathered debris is dragged down slowly, forming colluvium. In humid areas, colluvium practically covers the whole land surface, obscured only where an occasional outcrop protrudes or a surficial deposit is present. Some landslides develop in colluvium, so the processes of creep and landslide are intermingled. Engineering geologists can help evaluate the stability of hill slopes for catastrophic landslides or minor damage from creep. The classic landslide is spoon shaped and has a bulging toe and a scarp at the top. Ancient landslides may still show this distinct topographic pattern.

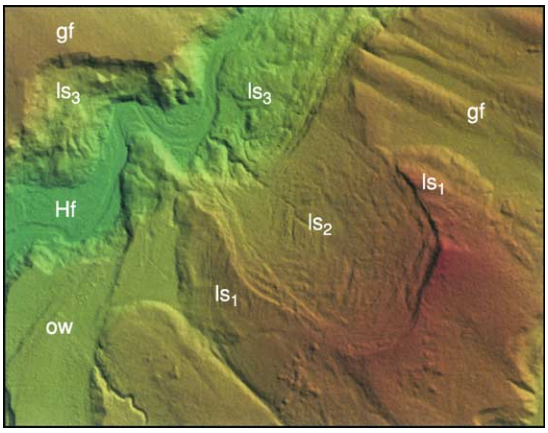


Figure 5 Lidar topography showing landslides near Seattle, Washington. The view is 3.5 km wide. Symbols: ow, outwash; ls, landslides (1 = youngest, 3 = oldest); Hf, Holocene flood plain; gf, glacially fluted surface. Courtesy of the Geological Society of America (see *GSA Today*, November 2003, pp. 4–10).

Figure 6 Landslides at Anchorage, Alaska, caused by the 1962 Alaska earthquake. Courtesy of the United States Geological Survey.

Residual soils develop *in situ* from the weathering of the parent bedrock (*see* **Soils:** Modern). Saprolitic deposits develop soil-like attributes and can accumulate to great thickness in humid areas such as the south-eastern United States. Transported soils, on the other hand, are simply surficial deposits that were moved and deposited by glaciers, running water, wind, or coastal processes. These geomorphologic processes produce deposits that have distinct attributes. For example, loess (wind-blown silt) is noted for its property of standing in vertical excavations. Till, unsorted debris deposited directly by a glacier, covers vast areas of northern Europe and North America. Till is noted for its poor drainage and often forms swamps. Its low permeability is well known, making it useful for ameliorating groundwater contamination plumes. Engineering geologists who work in glacial terrain need to understand the origin and characteristics of glacial deposits. A good working knowledge of the Quaternary deposits in a glaciated region is important in engineering endeavors and a valuable asset to a geotechnical firm.

Tectonic Terrains

Much of Earth's surface is the result of active tectonism. Mountains such as the Alps and Himalayas have been formed by tectonic uplifts . Normal faults in the western USA have produced a distinctive 'basin and range' topography; alluvial fan complexes form in the grabens adjacent to the uplifted mountains. Active strike–slip faults in California have resulted in diverse topographic features. Figure 7 is an oblique aerial photograph showing "basin and range" topography at the Wasatch Mountains near Salt Lake City, UT. A geomorphologist would quickly grasp the significance of this remarkable photograph. It shows that the terminal moraine of a mountain glacier has been offset by a fault. The moraine obviously is late Pleistocene in age, probably deposited around 20 000 years ago. Thus the fault displacement is relatively recent, and the fault must be considered active. The area along the fault will be offset again. The damage from the earthquake will not be limited by the shearing of any road or structure directly on the fault trace, but will also affect nearby areas. The response of earth materials to ground shaking varies. Figure 6, for example, shows the response of clay deposits at a great distance from the epicentre of the 1964 Alaska earthquake: the marine clay broke into hundreds of landslides.

Figure 7 Oblique aerial photograph of a faulted terminal moraine, Salt Lake City, Utah. Courtesy of the Utah Geological Survey.

Geological deposits have a wide range of responses to earthquakes. Surficial deposits are especially susceptible to shaking. For example, in the 1989 Loma Prieta earthquake (M = 7.1) in the San Francisco Bay area, damage to houses was widespread where thick residual soils occur in the mountainous terrain. Damage to the Oakland waterfront was related to infilling along the wharves. Condominiums collapsed in the Marin District of San Francisco in areas underlain by artificial fill. Another example is at Kobe, Japan, where liquefaction of alluvium and lagoonal clay occurred during the 1995 earthquake (M = 7.2). Clearly, an important aspect of engineering geology in earthquake-prone regions is the identification of sensitive surficial deposits.

Glaciation

There are two general types of glaciers: mountain glaciers and continental glaciers (*see* **Sedimentary Processes:** Glaciers). Mountain glaciers form distinctive erosional and depositional topographic features. Terminal moraines mark the terminus of Pleistocene mountain glaciers (Figure 7). Vast continental icesheets covered much of the land of the northern hemisphere and have left surficial deposits, including till and stratified drift. The best aquifers in glaciated terrain are stratified drift. Figure 8 shows an area of eastern Connecticut where stratified drift primarily occurs in the form of kame terraces. Where saturated, these constitute excellent aquifers, typically yielding over 500 gallons minute^{-1} (2500 m^3 day^{-1}) to wells. Much of the uplands in Figure 8 are bedrock (schist and gneiss) that is thinly veneered with glacial till. Drumlins are found throughout much of the till-covered areas in New England. The yield of wells in these upland areas is very low, on the order of 1 gallon minute^{-1} (5 m^3 day^{-1}). The kame terraces, primarily those along the Willimantic River valley, are excellent aquifers.

Other surficial deposits are associated with glaciation, including loess (wind-blown silt) and deltaic deposits. Figure 6 shows a large mass of varved clay that was deposited below sea-level during the

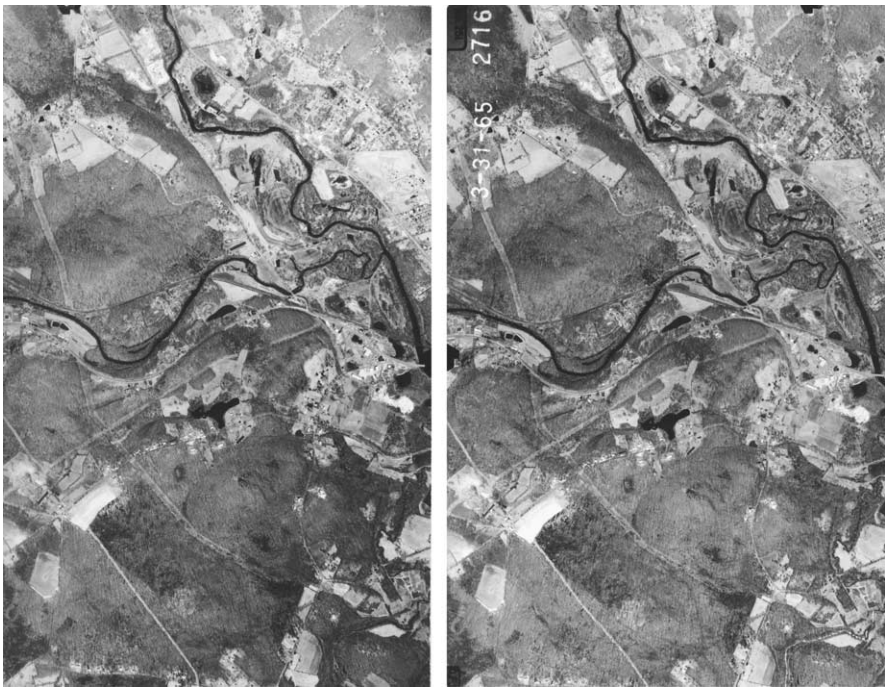


Figure 8 Stereo aerial photographs of glaciated terrain near Willimantic, Connecticut. Stratified drift, primarily kame terraces, are found in the Willimantic River valley; some are being quarried for sand and gravel. Best viewed with a pocket stereoscope.

Pleistocene Epoch. Much of Anchorage, AK, was destroyed when the 1964 Alaskan earthquake (M = 8.3) shook these deposits and they broke into hundred of landslides.

See Also

Engineering Geology: Geomorphology; Liquefaction. **Sedimentary Processes:** Catastrophic Floods; Fluvial Geomorphology; Glaciers; Landslides. **Soils:** Modern. **Tectonics:** Earthquakes.

Further Reading

Coates DR (1976) *Geomorphology and Engineering*. Stroudsburg, PA: Dowden, Hutchinson and Ross.

Costa JE and Baker VR (1981) *Surficial Geology, Building with the Earth*. New York: John Wiley & Sons.

Eckel EB (1970) *The Alaskan Earthquake, March 27, 1964, Lessons and Conclusions. US Geological Survey, Professional Paper 546*. Washington, DC: US Geological Survey.

Fookes PG and Vaughan PR (eds.) (1986) *A handbook of Engineering Geomorphology*. New York: Chapman and Hall.

Jahns RH (1947) *Geologic Features of the Connecticut Valley, Massachusetts, as Related to Recent Floods. US Geological Survey, Water-Supply Paper 996*. Washington, DC: US Geological Survey.

Kiersch GA (ed.) (1991) *The Heritage of Engineering Geology; the First Hundred Years. Geological Society of America, The Decade of North American Geology, Centennial Special Volume 3*. Boulder, CO: Geological Society of America.

Leopold LB (1994) *A View of the River*. Cambridge, MA: Harvard University Press.

McNutt SR and Sydnor RH (1990) *The Loma Prieta (Santa Cruz Mountains), California, Earthquake of 17 October 1989. California Division of Mines and Geology, Special Publication 104*. Sacramento, CA: California Geological Survey.

Rahn PH (1996) *Engineering Geology, an Environmental Approach*. Upper Saddle River, NJ: Prentice Hall.

Ritter DF, Kochel RC, and Miller JR (1995) *Process Geomorphology*. Dubuque, IA: William C Brown.

Royster DL (1973) *Highland Landslide Problems along the Cumberland Plateau in Tennessee. Association of Engineering Geologists, Bulletin 10*, pp. 255–287. Boulder, CO: Association of Engineering Geologists.

Underwood JR and Guth PL (eds.) (1998) *Military Geology in War and Peace. Reviews in Engineering Geology*, vol. 13. Boulder, CO: Geological Society of America.

Verstappen HT (1983) *Applied Geomorphology*. Amsterdam: Elsevier.

Waltham AC (1994) *Foundations of Engineering Geology*. London: Blackie Academic and Professional.

West TR (1995) *Geology Applied to Engineering*. Englewood Cliffs, NJ: Prentice-Hall.

GEOMYTHOLOGY

A Mayor, Princeton, USA

Introduction

Geomythology (also called 'legends of the Earth', 'myths of observation', 'natural knowledge', and 'physico-mythology') is the study of etiological oral traditions created by pre-scientific cultures to explain, in poetic metaphor and mythological imagery, geological phenomena such as volcanoes, earthquakes, floods, fossils, and other natural features of the landscape. (A related field, 'archaeoastronomy', studies pre-scientific knowledge of sky phenomena.)

Two types of geomyth have been identified: folk explanations of notable geological features, and often-garbled descriptions of catastrophic geological events that were witnessed in antiquity. In the case of geomorphic events that occurred in the pre-human past, observation and imagination led to mythic explanations that were handed down over millennia. In the case of natural cataclysms within living human history, descriptions were transmitted over generations, often accruing supernatural details. Because of the mythological language of oral folklore, scientists and historians have often missed the kernels of truth and rational concepts embedded in geomythological narratives. Some geomyths are simply fanciful stories based on imagination or popular misconceptions, such as tales of creatures or humans that were magically transformed into rock to explain the shapes of landforms. Many geomyths, however, contain surprisingly accurate insights into geological processes, as well as important eyewitness data from the distant past. Modern scientific investigations have revealed that much ancient folklore about the Earth was based on rational speculation and understandings grounded in careful observations of genuine, but extraordinary physical evidence over time.

Geomythology in Classical Antiquity

Although the term 'geomythology' was coined in 1968 by the geologist Dorothy Vitaliano, and it is often considered a new field of study, the concept was known and applied since antiquity. Euhemerus, a Greek philosopher (*ca.* 300 BC), held that myths about divinities and their activities were poetic accounts of real people and events. His approach, later called 'euhemerism', was taken up by other classical scholars who rationalized myths by stripping away supernatural and impossible details to reveal an underlying core of facts. Some of the rationalizing deconstructions of hero and monster myths by the Greek euhemerist Palaephatus (fourth century BC) may seem contrived, but others, such as his interpretation of the myth of Cadmus sowing the dragon teeth, are quite sophisticated. Palaephatus suggested that the tale represented an ancient misunderstanding of fossil elephant molars, which were frequently found in the ground and treasured by kings in archaic Greece before knowledge of elephants was brought back from India by Alexander the Great in the fourth century BC.

Many other classical Greek and Latin writers identified archaic myths about creation and primal creatures, first written down in the eighth century BC by the epic poets Homer and Hesiod, as symbolic ways of describing actual events and processes in deep time or within human memory. In his account of Atlantis and in other works, for example, the philosopher Plato (fourth century BC) correctly described large-scale changes in prehistoric land masses and coastlines in the Aegean. In the first century BC, the Latin poet Ovid expressed accurate conceptions of geomorphology and the process of petrifaction in his *Metamorphoses*; verses about the transformations of mythic beings. Around the same time, the Greek geographer Strabo regarded traditional geomyths as cryptic historical records, observing that "the ancients expressed physical notions and facts enigmatically by adding mythical elements".

In his fifth century BC tragedy *Prometheus Bound*, the Greek dramatist Aeschylus recounted the myth of Zeus burying Typhoeus (a monstrous, many-headed dragon with many voices that embodied primal chaos) under Europe's largest and most active volcano, Mount Etna in Sicily. The etiological account explains Etna's eruptions as Typhoeus's struggles to escape the subterranean prison: his roars and hisses were the auditory features of the volcano, and his fiery breath was supposed to melt rock, creating the periodic lava flows that endangered towns on the volcano's slopes. In the first centuries BC and AD, the Latin epic poet Virgil and the natural historian Pliny the Elder suggested that the one-eyed giant Cyclopes described by Hesiod in his *Theogony* (*ca.* 750 BC) personified other active volcanoes in the Mediterranean.

Contributions of Geomythology to Modern Science

Robert Hooke was the first modern scientist known to have employed classical geomythology as historical evidence. To support his theory of vast geological alterations over time, in lectures presented in the period 1667–1688 Hooke drew on geomythology of landform changes and earthquake-related lore recorded by Plato, Pliny, Strabo, Virgil, and Ovid (as well as biblical scriptures). Hooke sought the empirical evidence of mythic narratives to show that great floods and earthquakes had repeatedly transformed the Earth in the deep past, and could justify the presence of petrified marine remains on mountaintops. His explanatory hypothesis was based on the idea of pole wandering (to be tested astronomically), which was not well received by some of his contemporaries. It was for this reason that he collected evidence from myths in support of his ideas about Earth history.

Another early geomythologist was the founder of comparative anatomy, Georges Cuvier (*see* **Famous Geologists:** Cuvier). He compiled a collection of ancient Greek accounts and North and South American Indian traditions about the discoveries of petrified bones of remarkable size to demonstrate the worldwide distribution and longstanding observations of the fossilized remains of immense creatures, which Cuvier identified as extinct elephant-like creatures. Edward Burnet Tylor (1865) was another early pioneer of geomythology. He called traditional legends about natural history 'myths of observation' to emphasize that they were reasonable efforts to account for mysterious physical evidence.

Vitaliano was the first modern geologist to systematically match the insights contained in various cultures' myths about geology to modern scientific knowledge, thereby giving the study of folklore a disciplinary status. As the first stirrings of geological observation and hypothesis-forming that would later evolve into the Earth sciences, geomyths are significant milestones in the history of science. As Vitaliano and others have noted, scientific theories themselves are analogous to etiological geomyths in that both are efforts to explain mysterious observed facts. The ability to link traditional descriptions with present-day science is a notable contribution to scientific knowledge. Geomyths can provide previously unknown additional data for studying geological events that were actually witnessed in the pre-scientific past. The study of geomythology also helps reveal the processes of transmitting cultural memories over many generations and the origins and functions of oral mythopoesis.

Scientific analysis of geomyths can verify the historical foundations of many myths previously viewed as imaginary products of creative storytelling. For example, in 1999, the frozen mummy of Kwaday Dan Sinchi (meaning 'long ago person found') was discovered in a melting glacier between Yukon Territory and north-western British Columbia, Canada. Radiocarbon dating showed that the young man had lived in the 1400s. He carried a waterproof hat woven of roots from the Pacific coast, a leather bag of dried fish and plants, tools made from both coastal and inland trees, and a cloak made from gophers that live far inland. These finds confirmed the ancient oral traditions of the local Champagne and Aishihik First Nations, which describe their ancestors using the glaciers as trade routes to travel between the interior and coast.

Examples of Geomythology

Geomyths from around the world explicate the gamut of geological phenomena, from seismic and volcanic events to fossil deposits, such as shells and marine creatures stranded far from the sea and strange, oversized, unfamiliar skeletons embedded in rock. Extraordinary landmarks and the sudden disappearance or appearance of islands; climatic changes; great floods and changing watercourses; natural petroleum fires and deadly gases emitted from the Earth; the formation of minerals and gems underground; and myriad other large and minute natural features of the landscape are all featured in geomythology.

For example, legends of deadly miasmas and 'birdless places' often arose in regions where toxic natural gases are released from vents in the Earth, affecting plants and wildlife. The ancient image of the cave-dwelling, fire-breathing monster, the Chimera, which supposedly dwelled in what is now modern Turkey was no doubt influenced by observations of spontaneously burning natural gas wells in Asia Minor. In 2002, a team of archaeologists and geologists confirmed a long-discounted classical Greek tradition that the priestess possessed by the god Apollo at the Oracle of Delphi was inspired by fumes emanating from a crack in the Earth. The team discovered that intoxicating methane and other gases escape from fissures at the ancient site of the Oracle.

Volcanoes and earthquakes are well represented in geomyths. The spectacular volcanoes of Hawaii inspired legends of the fire goddess Pele digging a series of great fire-pits as she traveled across the islands. Geologists point out that the legend reflects an ancient awareness that the volcanic activity from northwest to south-east was progressively younger.

The Greek myths of the cosmic wars between Zeus and the Titans, Cyclopes, and Typhoeus as described in Hesiod's *Theogony* were scientifically analyzed in 1992 by geology historian Mott T. Greene. Hesiod's poem contains some ancient oral stories that date to the second millennium BC. Greene demonstrated that the violent battle with the one-eyed Cyclops can be matched to volcanic phenomena associated with the solitary Mount Vesuvius and the solfataric gas emissions in the 'fields of fire' near Naples, Italy. In contrast, the god's conflict with Typhoeus and the Titans represents Mount Etna's multiple cones, and the hissing and roaring features, lava flows, and deep tectonic earthquakes. Moreover, details of Hesiod's poem suggest that it forms a chronological record of datable major eruptions of Etna in about 1500 BC and in 735 BC, and the Plinian eruption in 1470 BC on the island of Thera–Santorini, which destroyed the Minoan civilization.

Ancient Greek earthquake lore distinguished between local, weak volcanic tremors, attributed to struggling giants imprisoned by Zeus in the Earth, and large-magnitude tectonic quakes with associated tsunamis, which were attributed to Poseidon, the Earth-shaking god of the sea. In West Africa, where tectonic shocks typically emanate from the west, the natives imagined a giant who tires of facing east clumsily shifting position. Highly seismic Japan has elaborate quake myths, with some tales attributing the earthquakes and tsunamis to the movements of a colossal serpent-dragon surrounding the island. According to traditions in India and many other cultures, earthquakes were caused by giant creatures burrowing underground.

In Samos, an Aegean island with rich Miocene mammal fossils and subject to severe earthquakes, ancient Greeks devised an ingenious myth (fifth century BC) to explain both features. Before the era of humans, it was said that the island was populated by enormous monsters called Neades, whose deafening shrieks caused the very Earth to collapse upon them. According to ancient writers, their huge bones were displayed *in situ*, and archaeologists have found fossil relics in the ruins of the Temple of Hera on the island. By the first century AD the great bones were identified as the remains of Indian elephants brought to Greece by the god Dionysus. Not only are earthquakes on Samos distinguished by loud roaring, but many of the fossils on Samos are those of large, extinct elephant ancestors and the bone deposits are often found trapped underneath earthquake-faulted blocks.

Fossil remains generated a variety of geomyths speculating on the creatures' identity and cause of their destruction. Many ancient cultures, from China and India to Greece, America, and Australia, told tales of dragons, monsters, and giant heroes to account for fossils of animals that they had never seen alive. Some scenarios of their destruction in 'deep time' anticipated catastrophic extinction theories, first suggested scientifically by Cuvier, while other ancient accounts leaned toward more gradualist theories.

For the ancient Greeks, the Gigantomachy, the cosmic wars in which the gods destroyed giants and monsters and buried them underground, accounted for abundant deposits of the fossil skeletons of enormous, extinct Tertiary mammals found in 'giants battle fields' all around the Mediterranean. In the Siwalik foothills of the Himalayas, ancient Greek travellers reported that Indians displayed bizarrely horned dragons with sparkling gems embedded in their skulls. The origin of the myth came to light when nineteenth-century palaeontologists discovered rich fossils of giant giraffids and curiously tusked proboscids encrusted with calcite crystals. In China, the 'dragon' bones collected and ground into medicine turned out to be the fossils of extinct mammals and dinosaurs. In central Asia, the legend of the gold-guarding griffin, a creature with the body of a lion and beak of a raptor, arose as nomadic prospectors on their way to gold deposits came upon conspicuous fossils of beaked quadruped dinosaurs in the Gobi desert.

In Europe, observation of dinosaur tracks in Triassic sandstones in the Rhine Valley probably influenced the legend of the slaying of the dragon Fafnir there by the Germanic hero Siegfried. For Aborigines near Broome, north-west Australia, the footprints of Cretaceous carnosaur and stegosaur dinosaurs in sandstone near Broome, form a 'song-line' from dream-time. The trackways are considered the trail of a giant 'Emu-man' of myth. Where tracks head out to sea and back to shore, legend tells of him wading into the ocean and returning. Wherever he rested, Emu-man's feathers made impressions in the mud, a logical interpretation of fern fossils in sandstone, which resemble large feathers. According to an Aztec legend preserved by sixteenth-century Spanish explorers, the great feathered serpent-god Quetzalcoatl left his hand and seat prints in stone near Mexico City, one of the earliest geomyths recorded in America. Pleistocene fossils of large mammals are in fact abundant around ancient Aztec sites near Mexico City and the tracks of these probably account for the myth.

Myths of a devastating flood are nearly universal. Recently, a team of geologists studied sediments in the Black Sea region and concluded that Mesopotamian and Biblical flood myths originated when the rising Mediterranean suddenly broke through the Bosphorus,

inundating the populous farmlands of the Black Sea basin about 6000 years ago. The awareness that a vast sea once covered the American south-west in Cretaceous times is evident in Zuni Indian traditions about the bizarre huge marine monsters of a long-past era before humans evolved, whose remains are found along with shells and ripple marks in the desert bedrock. The Zuni creation myth describes how the sea was dried out by a great conflagration. Indeed, giant marine reptiles of the Cretaceous are found in Zuni lands, along with the burned stumps of great prehistoric forests in the desert.

Remarkable landforms have long elicited folk etiologies to explain their origins and notable features. Devil's Tower in Wyoming, a prominent volcanic formation in the American West with distinctive grooves and facets, was said by Native Americans to have been formed when a gigantic bear clawed at the rock in an attempt to reach children trapped on the top. Notably very large dinosaur claws are found in the region. The unusual Cuillin Mountains of Skye, Scotland, were fabled to have been formed when the Sun hurled his fiery spear into the ground. Where it struck, a huge blister or boil appeared and grew, swelling until it burst and discharged molten, glowing material that congealed to form mountains perpetually covered in snow. Geologists have remarked that the legend accurately recounts the formation of a volcanic dome, which grows, bursts, and spews glowing-hot magma. The Cuillins consist of gabbro, crystallized molten matter, and the adjacent mountains of granite, the Red Hills, are indeed snow-capped in contrast to the steeper Cuillins. However, the Cuillins are much too old to have been observed being formed by humans.

Controversies and Future Directions

In recent years, the horizons of geomythology have been expanded by Native American scholars who relate Amerindian geological traditions to modern scientific knowledge. These scholars also grapple with the controversial questions raised by geomyths: how far can human memory, perpetuated in spoken traditions over generations, extend back in time? Strabo was the first to address this issue in the first century BC. The very magnitude of time encompassed in folk memories makes legends seem incredible, he wrote, yet it is worth trying to decipher what the ancients understood and witnessed.

In the Renaissance, naturalistic approaches to the meaning of ancient geomyths vied with moralistic interpretations; and debates over the validity of traditional folk knowledge continued into the scientific era. For example, debate raged in the nineteenth century over Native American legends that seemed to contain ancestral memories of mastodons hunted to extinction in the last Ice-Age about 10 000 years ago. The possibility was supported by twentieth-century archaeological discoveries of mammoth kill sites that matched local tribal lore about elephant-like monsters.

One method of testing the reliability of oral geology traditions is to examine traditions about geomorphic events in specific geographic areas with datable chronologies. Greene's analysis of Hesiod's ancient volcano data provides an example of this approach. In 2003, geologists found that Homer's description of the landforms around Troy, now radically changed, is consistent with the way the region probably looked 3 millennia ago. Another convincing geomyth of surprising antiquity is the Klamath Indians' oral tradition about the largest Holocene eruption in North America, the volcanic explosion of Mount Mazama in the Cascades Range of southern Oregon. About 7500 years ago, the spectacular eruption blew off the top of the mountain and rained ash over a half million square miles. The resulting caldera formed Crater Lake. Surviving palaeo-Indian witnesses created a detailed oral tradition of the violent event, expressed in a mythological story that has been transmitted in the original Native American language over some 250 generations. The Klamath myth contains geological facts about the eruption and collapse of the mountain that were unknown to scientists until the early twentieth century.

Future directions in geomythology will continue the search for evidence of very early geological knowledge, based on logical reasoning or firsthand observations of natural phenomena, embedded in ancient mythologies. Some of the most interesting new research trends in geomythology are the investigations by anthropologists, psychologists, and folklorists into the evolutionary psychology of the oral mythmaking process itself, to learn more about the mechanisms of preserving and perpetuating ancestral human memories over millennia. Such studies may help solve an urgent geological dilemma that requires today's scientists to think geomythologically: the permanent and safe geological disposal of thousands of tons of highly radioactive nuclear waste from reactors and weapons production. The transuranic waste is expected to remain radioactive for 100 000 years. The plan is to bury the dangerous materials very deep in the Earth and guarantee that the sites remain undisturbed for at least 10 000 years. The vast geological and chronological scale of the project means that warnings to succeeding generations must survive in meaningful form until the year AD 12 000.

Scientific proposals for ensuring that inadvertent human intrusion will not occur at such burial sites in the far distant future have called for the creation of new, long-lasting geomythological 'traditions', with written and visual markers of menacing design, to indicate to future generations the grave perils of what lies buried underground.

See Also

Biblical Geology. **Famous Geologists:** Cuvier. **History of Geology Up To 1780**. **Tectonics:** Earthquakes. **Volcanoes**.

Further Reading

Barber E and Barber P (2005) *When They Severed Earth from Sky: How the Human Mind Shapes Myth*. Princeton: Princeton University Press.

Birkett K and Oldroyd D (1991) Robert Hooke, physicomythology, knowledge of the world of the ancients and knowledge of the ancient world. In: Gaukroger S (ed.) *Uses of Antiquity*, pp. 145–170. Dordrecht: Kluwer Academic Publishers.

Cataldi R, Hodgson SF, and Lund JW (eds.) (1999) *Stories from a Heated Earth: Our Geothermal Heritage*. Sacramento: Geothermal Resources Council, International Geothermal Association.

Clark E (1952) *Indian Legends of the Pacific Northwest*. Berkeley: University of California Press.

Deloria V (1997) *Red Earth, White Lies*. Golden (CO): Fulcrum.

Echo-Hawk RC (2000) Ancient history in the New World: integrating oral traditions and the archeological record in deep time. *American Antiquity* 65: 267–290.

Greene MT (1992) *Natural Knowledge in Preclassical Antiquity*. Baltimore: Johns Hopkins University Press.

Hale JR, de Boer JZ, and Chanton J (2001) New evidence for the geological origins of the ancient Delphic oracle (Greece). *Geology* 29: 707–710.

Kraft J C, Rapp G, Kayan I, and Luce JV (2003) Harbor areas at ancient Troy: sedimentology and geomorphology complement Homer's Iliad. *Geology* 31: 163–166.

Mayor A (2005) *Fossil Legends of the First Americans*. Princeton: Princeton University Press.

Mayor A (2000) *The First Fossil Hunters: Paleontology in Greek and Roman Times*. Princeton: Princeton University Press.

Mayor A and Sarjeant WAS (2001) The folklore of footprints in stone: from classical antiquity to the present. *Ichnos* 8(2): 1–22.

Rappaport R (1997) *When Geologists were Historians, 1665–1750*. Ithaca: Cornell University Press.

Ryan W and Pitman W (1998) *Noah's Flood: The New Scientific Discoveries about the Event that Changed History*. New York: Simon & Schuster.

Tylor EB (1865) *Researches into the Early History of Mankind*. London: John Murray (reprinted 1964, Chicago: University of Chicago Press).

Vitaliano D (1973) *Legends of the Earth: Their Geological Origins*. Bloomington: Indiana University Press.

GEOPHYSICS

See **EARTH: Orbital Variation (Including Milankovitch Cycles); EARTH SYSTEM SCIENCE; ENGINEERING GEOLOGY: Seismology; MAGNETOSTRATIGRAPHY; MOHO DISCONTINUITY; PALAEOMAGNETISM; PETROLEUM GEOLOGY: Exploration; REMOTE SENSING: Active Sensors; GIS; Passive Sensors; SEISMIC SURVEYS; TECTONICS: Seismic Structure At Mid-Ocean Ridges**

GEOTECHNICAL ENGINEERING

D P Giles, University of Portsmouth, Portsmouth, UK

Introduction

Geotechnical engineering has recently (1999) been formally defined in a Memorandum of Understanding on the proposed unification of the British Institution of Civil Engineers Ground Board and the British Geotechnical Society. Appendix A of that memorandum, establishing the British Geotechnical Association, sets out the following definition:

> Geotechnical engineering is the application of the sciences of soil mechanics and rock mechanics, engineering geology and other related disciplines to civil engineering construction, the extractive industries and the preservation and enhancement of the environment.

Geotechnical engineering plays a key role in all civil engineering projects, since all construction is built on or in the ground. In addition it forms an important part of extractive industries, such as open cast and underground mining and hydrocarbon extraction, and is essential in evaluating natural hazards such as earthquakes and landslides.

The use of natural soil and rock makes geotechnical engineering different from many other branches of engineering: whereas most engineers specify the materials they use, the geotechnical engineer must use the material existing in the ground and in general cannot control its properties.

In most cases the complexity of the geology means that the geotechnical engineer is dealing with particularly complicated and variable materials; their mechanical properties usually vary with time and are critically dependent on the water pressures in the ground, which can often change.

The geotechnical engineer does sometimes have the opportunity to specify certain properties or treatment of soils, rocks and other materials used in construction.

Geotechnics can thus be primarily considered as the science of the engineering properties and behaviour of rocks and soils. Geotechnical engineering can be considered as the professional practice and implementation of that knowledge contributing principally to the design of engineered structures in and on the ground (Figure 1).

Fundamental to geotechnical engineering are the study and practice of engineering geology, geomechanics (rock mechanics and soil mechanics), the design of foundations, the stabilization of slopes, the improvement of ground conditions, the excavation of tunnels and other underground openings, the analysis of ground behaviour, and the assessment of ground movements.

Soil Mechanics and Rock Mechanics

Soil mechanics (*see* **Soil Mechanics**) and Rock Mechanics (*see* **Rock Mechanics**), together known as 'geomechanics', involve the study and understanding of the physical properties and behaviour of rocks and soils. These properties will include material strength (in tension, compression, and shear), moisture content, porosity and permeability, and a description in engineering terms (including a description of the weathered state and of the rock and soil

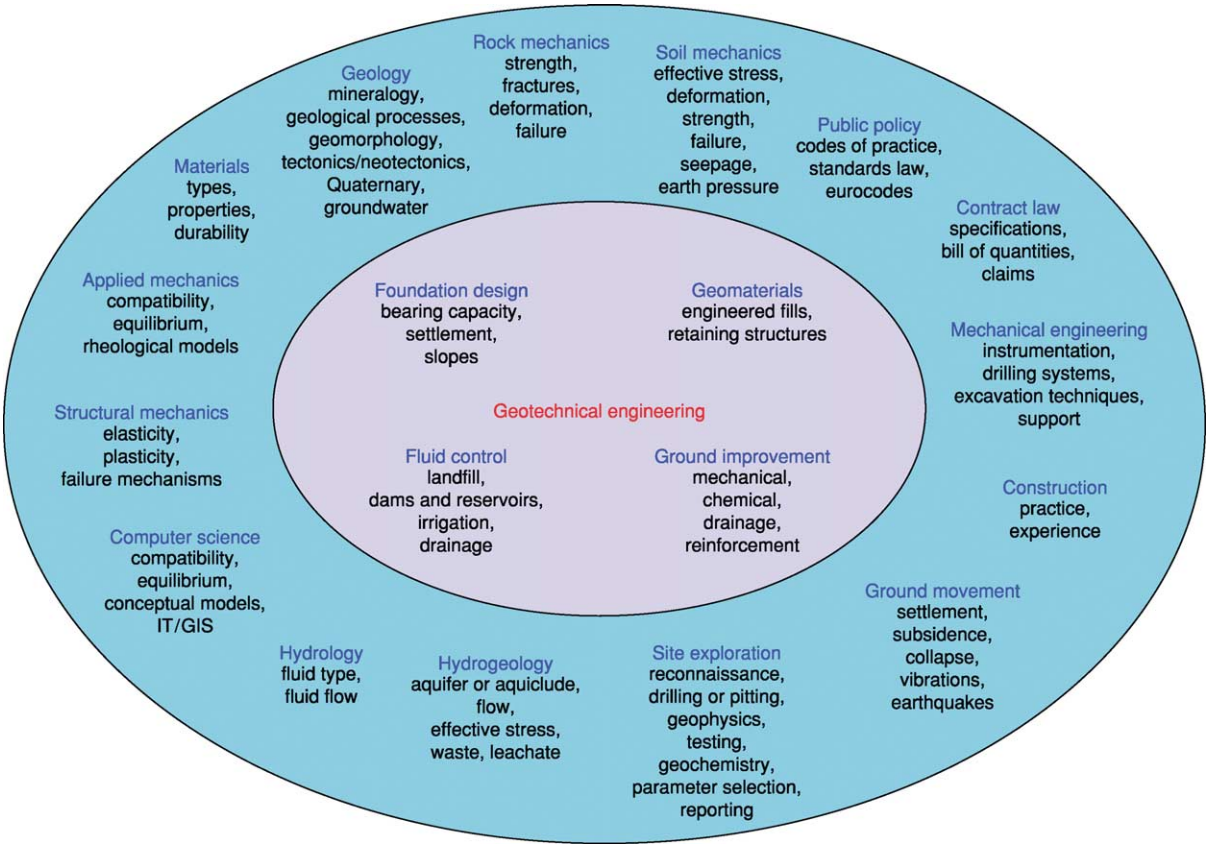


Figure 1 The practice of geotechnical engineering encompasses a wide variety of skills. Modified from the British Geotechnical Association (1999) Memorandum of Understanding, Appendix A, Definition of Geotechnical Engineering. *Ground Engineering*, Nov. p. 39 EMAP, London.

Table 1 Some typical geotechnical properties of engineering rocks

Rock type	*Unit weight (kN m^{-3})*	*Porosity (%)*	*Dry UCS range (MPa)*	*Dry UCS mean (MPa)*	*Saturated UCS (MPa)*	*Modulus of elasticity (GPa)*	*Tensile strength (MPa)*	*Shear strength (MPa)*	*Friction angle (φ°)*
Granite	27	1	50–135	200		75	15	35	55
Basalt	29	2	100–350	250		90	15	40	50
Greywacke – Ordovician	26	3	100–200	180	160	60	15	30	45
Sandstone – Carboniferous	22	12	40–100	70	50	30	5	15	45
Sandstone – Triassic	19	25	5–40	20	10	4	1	4	40
Limestone – Carboniferous	26	3	50–150	100	90	60	10	30	35
Limestone – Jurassic	23	15	15–70	25	15	15	2	5	35
Chalk – Cretaceous	18	30	5–30	15	5	6	0.3	3	25
Mudstone – Carboniferous	23	10	10–50	40	20	10	1		30
Shale – Carboniferous	23	15	5–30	20	5	2	0.5		25
Clay – Cretaceous	18	30	1–4	2		0.2	2	0.7	20
Coal – Carboniferous	14	10	2–100	30		10	2		
Gypsum – Triassic	22	5	20–30	25		20	1		30
Salt – Triassic	21	5	5–20	12		5			
Hornfels	27	1	200–350	250		80			40
Marble	26	1	60–200	100		60	10	32	35
Gneiss	27	1	50–200	150		45	10	30	30
Schist	27	3	20–100	60		20	2		25
Slate	27	1	20–250	90		30	10		25

UCS, uniaxial compressive strength.

Courtesy of Waltham AC (2001) Foundations of Engineering Geology. 2nd Edition. Spon Press. London.

en masse as well as material, a description of any discontinuities – such as fissures, fractures, joints, faults, and shears – and any other geological aspect of the mass fabric).

Geotechnical Engineering

Geotechnical engineering encompasses the civil engineering of the ground and deals with the interaction of that ground with engineered structures. The principal works undertaken are foundations, slopes, ground improvement, and underground excavations.

Foundations

The design and construction of foundations involves the calculation of bearing loads and capacities, the selection of the optimum foundation type (e.g. footing or pile), embankment design, retaining wall design, and seepage analysis. The geomechanical investigations required for foundation design will include evaluating soil stiffness, lateral earth pressures, consolidation coefficients, and material strengths, together with an analysis of the effective stress conditions.

Slopes

The study of slopes and landslides requires the design of cuttings, analysis of slope stability (in rock, soil, or debris), and consideration of slope stabilization measures (which may include slope drainage schemes, slope reprofiling, ground anchor and pinning systems, and other physical support methods).

Ground Improvement

Ground improvement includes mechanical and reinforcement measures, chemical treatment, drainage, bioengineering, and the use of geotextiles.

Underground Excavations

The design of tunnels and underground excavations involves the consideration of rock mass strength, the design of suitable excavation techniques (such as drill and blast, use of tunnel boring machines, use of roadheaders), and the design of suitable support systems (such as sprayed concrete (shotcrete), steel sets or supports, concrete segments, rock bolts). Arrangements must also be made to monitor any ground settlement, closure, or overstressing during the excavation operations.

Risk Analysis

The identification of geohazards and their incorporation in the risk management process requires hazard assessment and risk analysis (*see* **Engineering Geology:** Natural and Anthropogenic Geohazards). This work underpins all aspects of modern geotechnical engineering practice. The geotechnical engineer will be responsible for compiling a risk register, which includes an analysis and assessment of all ground-related risks that could affect the engineering project together with a programme of suitable responses to prevent such risks occurring or, should such risks be realized, suitable mitigation and management measures that should be implemented if the event occurs.

Ground Investigation and Characterization

Geotechnical engineering is underpinned by the assessment and characterization of ground conditions using site and ground investigation data (*see* **Engineering Geology:** Site and Ground Investigation). Geotechnical engineers therefore need to work closely with engineering geologists to develop conceptual models appropriate to the ground conditions, including the geological setting of the site and the geotechnical variability (Tables 1–4). Detail is provided by the site and ground investigation process, building on the background knowledge provided by earlier investigations, research, and the records held by local and regional authorities (notably the British Geological Survey). Site-specific data is used to generate factual and interpretive accounts of the ground conditions, using both *in situ* data from intrusive test programmes and tests conducted on sample data obtained from extensive laboratory testing programmes. These data are used to create a conceptual model from which expected ground conditions can be anticipated. This data analysis provides the key geotechnical parameters and values that are used in the subsequent geotechnical design.

Table 2 Some typical geotechnical properties of engineering soils

Soil type	*Grain size (mm)*	*Liquid limit (%)*	*Plasticity index (%)*	*Friction angle (φ°)*
Gravel	2–60	N/A	N/A	>32
Sand	0.06–2	N/A	N/A	>32
Silt	0.002–0.006	30	5	32
Clayey silt	0.002–0.06	70	30	25
Clay	<0.002	35	20	28
Plastic clay	<0.002	70	45	19
Organic	Generally amorphous	>100	>100	<10

N/A, not appropriate.

Courtesy of Waltham AC (2001) Foundations of Engineering Geology. 2nd Edition. Spon Press. London.

Table 3 Some typical geotechnical properties of typical clay soils

State	*LI (%)*	*SPT (N)*	*CPT (MPa)*	*C (kPa)*	*m_v ($m^2 MN^{-1}$)*	*ABP (kPa)*
Soft	>0.5	2–4	0.3–0.5	20–40	>1.0	<75
Firm	0.2 to 0.5	4–8	0.5–1	40–75	0.3–1.0	75–150
Stiff	−0.1 to 0.2	8–15	1–2	75–150	0.1–0.3	150–300
Very stiff	−0.4 to −0.1	15–30	2–4	150–300	0.05–0.1	300–600
Hard	<0.4	>30	>4	>300	<0.005	>600

LI, liquidity index; SPT, standard penetration test; CPT, cone penetration test; C, cohesion; m_v, compression coefficient; ABP, acceptable bearing pressure.
Courtesy of Waltham AC (2001) Foundations of Engineering Geology. 2nd Edition. Spon Press. London.

Table 4 Some typical geotechnical properties of typical sand soils

Packing	*Relative density*	*SPT (N)*	*CPT (MPa)*	*Friction angle (φ°)*	*SBP (kPa)*
Very loose	<0.2	<5	<2	<30	<30
Loose	0.2–0.4	5–10	2–4	3–32	3–80
Medium dense	0.4–0.6	11–30	4–12	32–36	8–300
Dense	0.6–0.8	31–50	12–20	36–40	3–500
Very dense	>0.8	>50	≥20	≥40	≥500

SBP, safe bearing pressure; CPT, cone penetration test end resistance; SPT, standard penetration test corrected N value.
Courtesy of Waltham AC (2001) Foundations of Engineering Geology. 2nd Edition. Spon Press. London.

Hydrology and Hydrogeology

Geotechnical engineering requires an understanding of and the ability to control fluids. Knowledge of both fluid type and fluid flow (hydrology) is necessary and builds on a knowledge of the geological controls (hydrogeology (*see* **Engineering Geology:** Ground Water Monitoring at Solid Waste Landfills)). This enables the design and construction of landfills, dams for impounding reservoirs, irrigation, and abstraction for water supply.

Geotechnical Modelling

Geotechnical modelling generally involves the construction of numerical and computational models of the ground conditions, permitting an evaluation of alternative engineering solutions for the design and construction of the proposed works. Computer-based techniques may use finite element, finite difference, distinct element, or particulate codes. These techniques are used to model, analyse, interpret, and visualize the variety of geotechnical processes and designs.

Geotechnical engineering is therefore a multidisciplinary subject that critically involves the understanding and application of soil mechanics, rock mechanics, and hydrogeology and is linked to the study of engineering geology. It is applied to civil engineering construction, mineral extraction, and the improvement of our environment.

See Also

Engineering Geology: Natural and Anthropogenic Geohazards; Site and Ground Investigation; Ground Water Monitoring at Solid Waste Landfills. **Environmental Geology**. **Rock Mechanics**. **Soil Mechanics**.

Further Reading

Anon (1999) Memorandum of Understanding, Appendix A, definition of geotechnical engineering. *Ground Engineering* 32: 35–39.

Attewell PB and Farmer IW (1976) *Principles of Engineering Geology*. London: Chapman & Hall.

Barnes EE (2000) *Soil Mechanics: Principles and Practice*. Basingstoke: Palgrave Macmillan.

Bell FG (ed.) (1987) *Ground Engineer's Reference Book*. London: Butterworth-Heinemann.

Brady B and Brown ET (1992) *Rock Mechanics for Underground Mining*. London: Chapman & Hall.

British Geotechnical Association (1999) Memorandum of Understanding, Appendix A, Definition of Geotechnical Engineering. *Ground Engineering*, Nov. p. 39. Emap, London.

Bromhead EN (1992) *The Stability of Slopes*. Glasgow: Blackie.

Clayton CRI (2001) *Managing Geotechnical Risk: Improving Productivity in UK Building and Construction*. London: Thomas Telford.

Clayton C, Matthews M, and Simons N (1995) *Site Investigation: A Handbook for Engineers*. UK: Blackwell.

Craig RF (1997) *Soil Mechanics*. London: Spon and Chapman & Hall.

Fetter CW (2001) *Applied Hydrogeology*. New Jersey: Prentice Hall.
Goodman RE (1989) *Introduction to Rock Mechanics*. New York: John Wiley & Sons.
Goodman RE (1993) *Engineering Geology*. New York: John Wiley and Sons.
Hoek E and Bray J (1981) *Rock Slope Engineering*. London: Spon Press.
Hoek E and Brown ET (1981) *Underground Excavations in Rock*. London: Institution of Mining and Metallurgy.
Hoek E, Kaiser PK, and Banden WF (1995) *Support of Underground Excavations in Hard Rock*. Rotterdam: Balkema.
Moseley MP (1992) *Ground Improvement*. London: Spon Press.
Muir-Wood D (2004) *Geotechnical Modelling*. London: Spon Press.
Potts DM and Zdravkovic L (2001) *Finite Element Analysis in Geotechnical Engineering: Theory and Application*. London: Thomas Telford.
Smith GN (1998) *Elements of Soil Mechanics*. London: Blackwell.
Smith IM (1994) *Numerical Methods in Geotechnical Engineering*. London: Thomas Telford.
Tomlinson MJ (2001) *Foundation Design and Construction*. London: Prentice Hall.
Wyllie DC (1992) *Foundations on Rock*. London: Spon.

GEYSERS AND HOT SPRINGS

G J H McCall, Cirencester, Gloucester, UK

Introduction

Geysers and hot springs that emerge at openings on Earth's surface are primarily found in regions of senescent or dormant volcanic activity. A second type of activity unrelated to volcanism also produces thermal waters; these 'meteoric' waters, the product of rain and snowfall, after descending deep into the ground through rock fissures and pores, have elevated temperatures at depth due to the global geothermal gradient. Both types of geothermal waters, convective and conductive, respectively, serve as geothermal energy resources and as resources for other purposes, but convective waters are more commonly exploited at present.

A unique ecology exists at geothermal sites. Under the sea, black smokers that emanate from geothermal vents relate closely to valuable sulphide mineral deposits found in ancient rocks; these sites also harbour extraordinary tube worm populations. On land, various minerals are deposited at sites of geothermal activity, micro-organisms thrive in the surrounding heated environment, and the warm waters have long been used as health spas. In addition to their popular use for health and medicinal purposes, geysers and hot springs serve as indicators of below-surface events. Increased geothermal activity provides a valuable warning of renewed volcanic activity. The entire spectrum of mild eruption as a volcanic process has an extension into the energy field, into the field of geomedicine, and into commercial enterprise. Geothermal energy is used in a number of countries worldwide, and some of the geothermal mineral deposits (for example, sulphur and borates) and associated rock formations (travertine and tufa) are of commercial value.

Geothermal Systems

The rocks of the uppermost 2–4 km of Earth's crust are generally porous and/or fissured and may be aquifers, filled with groundwater that has percolated down from rain and snow falling onto the surface. Those rocks that are not porous or fissured, i.e., are impermeable, form impermeable aquacludes, which separate or cap groundwater bodies. Heat within Earth is of two types: heat produced by gravitation on accretion and radiogenic heat from radioactive mineral decay. The transport of heat to Earth's surface is manifested in two ways. The first of these is a convective process, associated with the rise of hot magma, generated by partial melting in the mantle or crust. The molten rock cools in magma chambers, and this is the main source of geothermal heat as manifested in geysers and hot springs. The heat is passed from the cooling magma in the chamber to the surrounding country rocks, especially those above the chambers. Where these chambers contain groundwater bodies, the water expands and rises buoyantly, to be replaced by cold water flowing in from the sides, which is in turn heated and rises. This process establishes a geothermal circulatory system, which cools the

magma. Heat is initially brought up by the rising magma and then by the geothermal system as water or steam (the latter, if the water boils, which is a factor of temperature or pressure). Volcanoes, especially those with caldera development, are typically underlain by shallow magma chambers, which are fed from other larger reservoirs deeper in the crust or mantle. Thus geothermal systems are characteristic of regions with active, senescent, or dormant volcanoes. Most active volcanoes have been in action for less than a million years, so geothermal systems are geologically very young systems (*see* **Volcanoes**).

The second type of geothermal system is not volcanic related, but is related to the conductive transfer of heat outward through the crust. Because there is a temperature gradient in the crust, increasing with depth from the surface towards the mantle and core, groundwater taken in by recharge from rain and snowfall can be heated simply by being carried down to depths of a few hundred metres. These heated waters may then return to the surface as hot springs or geysers, or else remain as contained, heated groundwater or steam bodies in the subsurface. Heated water bodies below the surface are essentially unstable if they are overlain by cold water bodies, and will tend to move upward if not confined. Some hot water bodies may be in artesian conditions and flow out naturally in wells and bores. Natural outflows can be used for spa purposes. Hot water bodies in the subsurface that are confined by aquacludes can be tapped by boreholes and utilized for energy purposes and space heating.

By far, a majority of geothermal occurrences are related to volcanic activity and thus occur in volcanic provinces, but some (for example, in Bath, England) have no such obvious relationship. The entire Paris Basin, a structure analogous to the Cretaceous chalk-underlain London Basin, is underlain by heated groundwater bodies that can be exploited to a limited extent for residential heating and other purposes.

Mild Eruption (Volcanic-Related Geothermal Processes)

Hydrothermal

Hydrothermal relationships on Earth have been treated concisely by Alwyn Scarth in the book *Savage Earth*. On land, mild eruptions are limited to emissions of gas (mainly carbon dioxide), steam, and hot water and to formation of sulphurous fields, bubbling mud pools, hissing holes, and fissures. Such developments are associated with dormant or dying volcanoes that have been much more violent in the past and may become so in the future. Where such manifestations are present, it is impossible to say that the volcano is extinct. Lassen Peak (California, USA), Tiede volcano (Tenerife), and Taftan Peak (Southern Iran) are examples of dormant volcanoes. In the Gregory Rift Valley of Kenya, throughout several hundred kilometres, no volcano is presently active, but geothermal indications are ubiquitous. There is a geyser accompanied by hot pools at Lake Bogoria and a hot-water waterfall south of the lake, at Kapedo (Figure 1). In volcanic provinces, to augment seismic records, ground movement measurements, and other monitoring systems, geothermal hot springs and geysers should be monitored, because any increase in activity provides a valuable indicator of likely renewed volcanic eruptivity.

Figure 1 Hot-water waterfall, Kapedo, Kenya. There are a number of hot springs immediately above the fall, the headwater of the Suguta River, and the water is believed to travel underground from Lake Baringo, ~50 km to the south. Photograph by GJH McCall.

Geothermal waters can be analysed to determine their origin. Isotopic ratios can differentiate between magmatic water ('juvenile' water, or water that has been brought to Earth's surface by upward movement of rocks) and meteoric water (groundwater); geothermal waters are almost completely meteoric. Meteoric water is water from rain and melting snow that has percolated downward. In many volcanic areas, hot magma still lies close to the surface, long after lava and ash eruptions have ceased, and the rocks above such residual magma chambers remain hot. As groundwater percolates down, heated rocks increase the groundwater to temperatures of more than 200°C, without boiling, because of the pressure. In the simplest case, this water returns to the surface to form hot water springs and pools, hissing and bubbling up, or it is tapped in wells and boreholes. Such developments at Rotorua, New Zealand, are well known tourist and health spa attractions.

Solfataras, Fumaroles, and Mofettes

Vent openings in the ground allow escape of heated waters and vapours. Escaping steam may mix with sulphurous gases and will deposit sulphur when its emerges at the surface. Where sulphur dioxide and hydrogen sulphide are emitted, the hissing escape channels are ringed at the surface with yellow sulphur deposits. Vent openings where such emissions occur are termed 'solfataras' (after La Solfatara volcano, Pozzuoli, Italy) and 'fumaroles'. Vents that are sources of toxic gases are termed 'mofettes'. In addition to carbon dioxide and sulphurous gases, inert gases such as nitrogen and argon can also be emitted. These gases may be almost undetectable and are in such cases dangerous. The principal gas emitted from most vents, carbon dioxide, can be lethal in high concentrations. It was the cause of a disaster at Lake Nyos, Cameroon, in 1986. Lake Nyos, the mouth of an ancient volcano, released a lethal cloud of CO_2 that asphyxiated more than 1700 people in nearby villages.

Geysers

Where hot water cannot circulate freely below the surface, it will gush out intermittently as pressure builds up. The water may appear muddy if it is mixed with material derived from the buried rock mass in which it is contained. In the extreme case, the phenomenon of the geyser (named after Geysir, a locality in Iceland) occurs, and water and steam spurt from the surface intermittently in a high-reaching fountain. 'Old Faithful' at Yellowstone, Wyoming, USA (Figure 2) reaches a height of 20 m, erupts every 60 minutes, and the eruption lasts for 5 minutes. It erupts so regularly that timetables have been created

Figure 2 'Old Faithful' geyser, Yellowstone National Park, Wyoming, USA. Reproduced with permission from Green J and Short NM (1971) *Volcanic Landforms and Surface Features: A Photographic Atlas and Glossary*. New York, Heidelberg, Berlin: Springer-Verlag.

announcing the schedule of eruption. The Nordic version of Old Faithful is Strokkur, in Iceland (Figure 3). Geysers have individual schedules: a geyser at the south end of Lake Bogoria (formerly Hannington), Kenya, erupts about every 10 minutes, whereas Beehive geyser, in Yellowstone, erupts only once a year, and a geyser at Rotorua, New Zealand, erupts four times a week. Changes in rainfall patterns can affect geysers; the Lake Bogoria geyser can temporally cease activity if the lake level gets too high, and the Great Geyser in Iceland, though active for 8000 years, now has to be stimulated by an injection of soap powder.

Figure 3 Strokkur geyser, Iceland. Reproduced with permission from Green J and Short NM (1971) *Volcanic Landforms and Surface Features: A Photographic Atlas and Glossary*. New York, Heidelberg, Berlin: Springer-Verlag.

Sinter and Travertine Terraces

Hot waters dissolve various chemicals as they traverse underground rock channels. When hot waters exit at surface openings, the dissolved chemicals are precipitated and form cowls or cones around the exit fissure or aperture. If the waters stream downslope, they may form spectacular terraces. The pink terraces at Rotorua, New Zealand, were long famous but were destroyed by earthquake action, and the best example now extant is the magnificent organ-pipe terraces at Mammoth Springs, Yellowstone, Wyoming (Figure 4). Such terraces are commonly composed of calcium carbonate and are referred to as tufa, where they are carious, or as travertine, where they occur in massive layers. Where such deposits are siliceous, they are sometimes referred to incorrectly by these terms, and are best referred to as 'siliceous sinter' (calcareous deposits are also sometimes referred to as 'calcareous sinter').

Figure 4 Mammoth Hot Springs terraces, Yellowstone National Park, Wyoming, USA. Reproduced from Science Photo Library.

Bioherms and Stromatolites

The mounds around hot springs may accommodate living creatures, forming bioherms. Stromatolites are layered, built-up structures found near hot springs; they are composed of populations of unicellular cyanobacteria and algae, organisms that have been dated to as far back as 3500 million years ago (*see* **Biosediments and Biofilms**, **Minerals:** Carbonates). At Lake Bogoria, Kenya, there are intermittent developments of stromatolites in the highly saline shallows of the lake close to the cluster of hot springs, pools, and geyser. Such stromatolites and algal fossils are commonly preserved in travertine that is used for ornamental building stone or cladding.

Exploitation of Volcanic-Related Geothermal Heat

In Reykjavík, Iceland, the volcanic-sourced heat is used to meet energy needs for general urban purposes. This local utilization is similar to that applied to non-volcanic-related geothermal heat in the Paris Basin (see later). The derived energy is

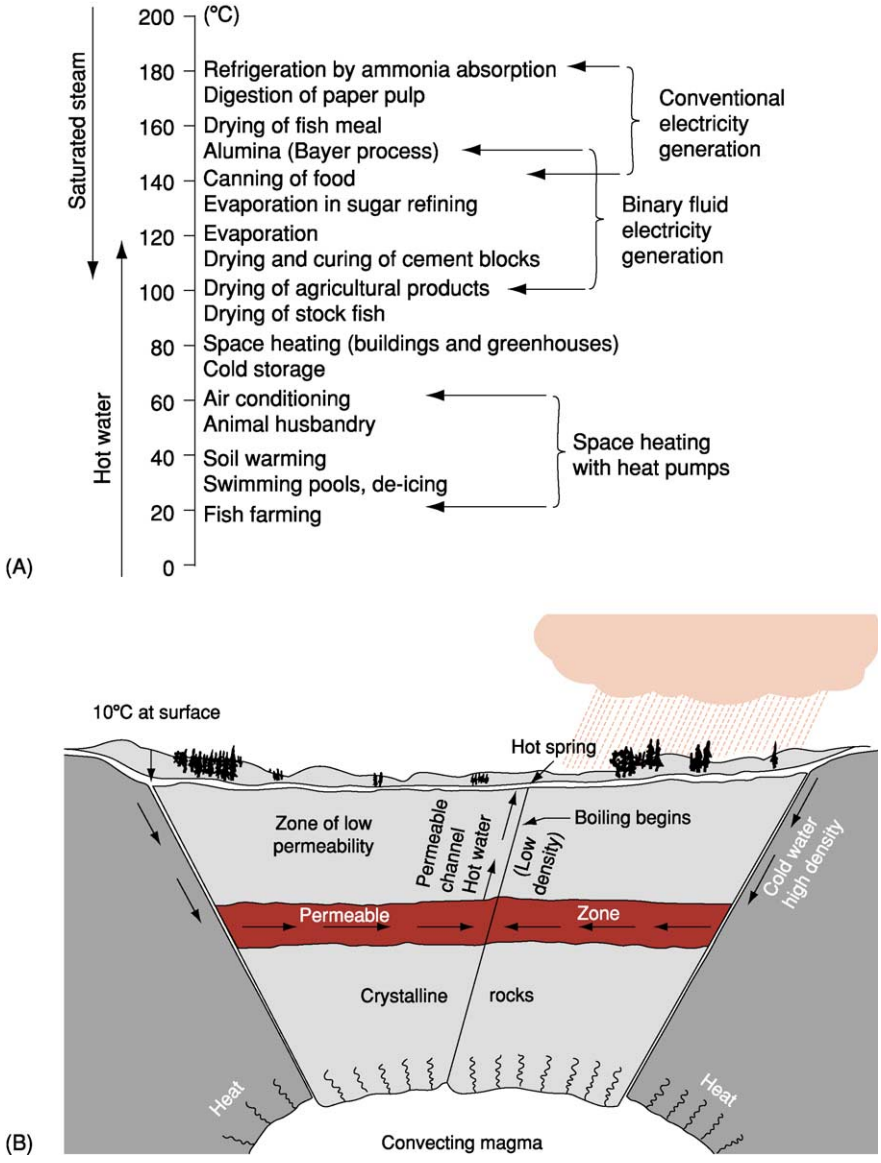


Figure 5 (A) Temperature requirements of geothermal fluids for various uses. Reproduced with permission from Arnorsson S (2000) Geothermal energy. In: Hancock PL and Skinner PJ (eds.) *Oxford Companion to the Earth*, pp. 437–440. New York: Oxford University Press. (B) Diagram showing a vapour-dominated geothermal field. Reproduced with permission from Cargo N and Mallory BF (1977) *Man and His Geologic Environment*. Reading, MA: Addison-Wesley.

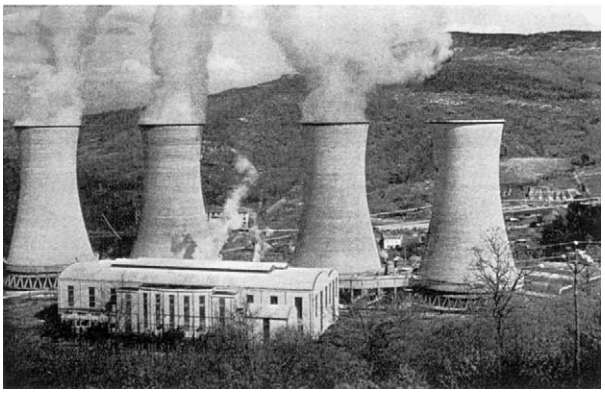


Figure 6 (A) 'Soffioni' blasting out at Larderello. (B) Power station installation at Larderello. Reproduced from Mazzoni A (1954) *The Steam Vents of Tuscany and the Larderello Plant*. Bologna: Arti Grafiche Calderini.

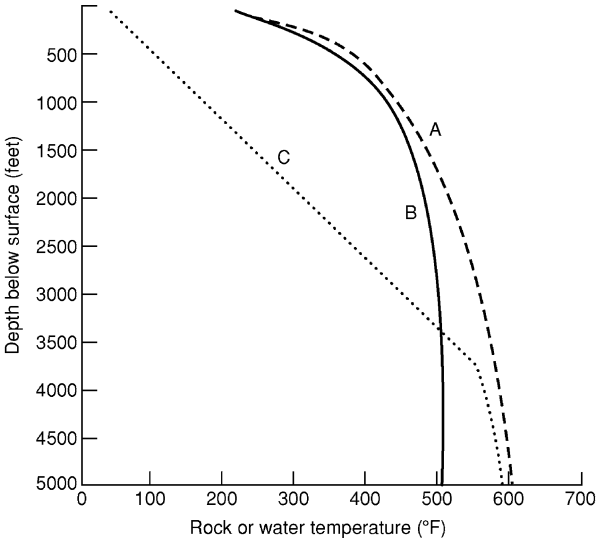


Figure 7 Water temperature variations with depth. Curve A: normal increase in boiling point. Curve B: increase in boiling point in hot spring region; magma or hot rocks near surface. Curve C: increase in boiling point in region where impermeable rocks prevent upward leakage; magma or hot rocks below impermeable rocks. Reproduced with permission from Cargo N and Mallory BF (1977) *Man and His Geologic Environment*. Reading, MA: Addison-Wesley.

suitable for space heating of homes and greenhouses and for use in local industry. The minimum temperatures of geothermal waters for various utilizations are given in Figure 5A.

Major exploitation of geothermal energy has been so far restricted to volcanic-related occurrences. For such utilization, a reservoir temperature of more than 200°C is desirable. The Larderello field, south of Pisa, Italy, initiated in 1906, is a 'dry' or 'vapour-dominated' steam field in that the deep resource is under high temperature but low pressure, not much more than atmospheric (Figure 5B). The water boils underground and the generated steam that is trapped in pores within the underground rock can be tapped by boreholes. A 'soffioni' (vent) at Larderello can emit a burst with immense power in a spectacular fashion; the steam reservoir there is capped by impermeable rock such as schist, and boring is hazardous in that equipment and personnel can be injured when the borehole reaches a steam pocket (Figure 6A). The steam is superheated and may be corrosive, especially related to boric acid, which was formerly extracted from Larderello for commercial use. The energy output of the several power stations at Larderello (Figure 6B) is very large, and most of the railways in northern Italy were at one time reported to utilize it. The commercial geothermal plant at the Geysers, north of San Francisco, California, is similar to the Larderello plant; it is the only steam-dominated geothermal field in the United States, and the output has now gone well beyond the original 12 500-kW capacity.

The 'wet', or liquid-dominated, geothermal plants tap reservoirs under high pressure, where the temperature may reach 350°–700°F without boiling (Figure 7). Under these pressures and temperatures, water flashes to steam up the borehole.

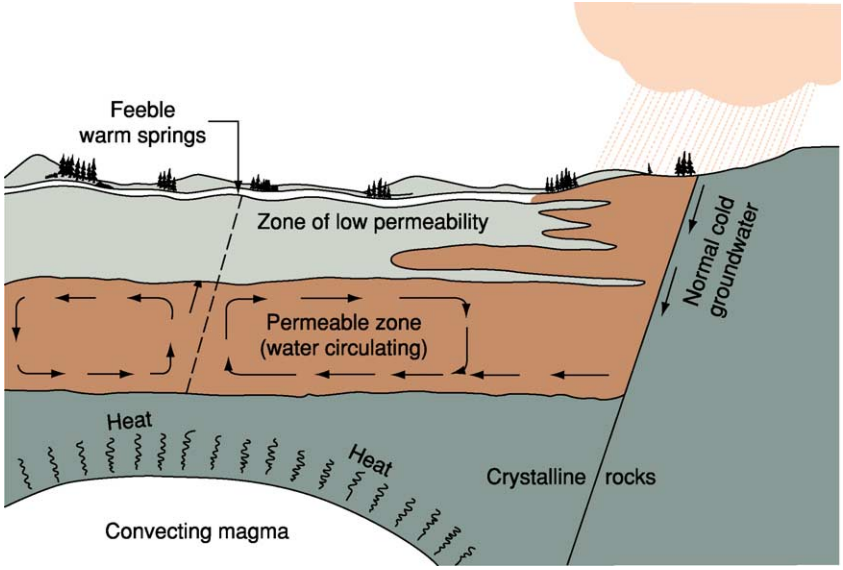


Figure 8 Diagram showing a liquid-dominated geothermal field. Reproduced with permission from Cargo N and Mallory BF (1977) *Man and His Geologic Environment.* Reading, MA: Addison-Wesley.

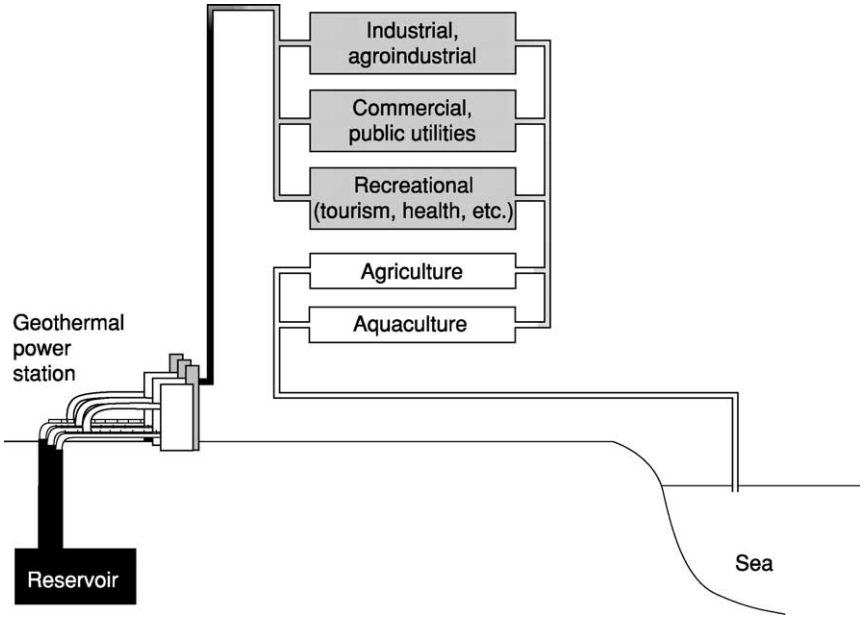


Figure 9 Diagram showing cascading relationships adopted at Reykjavík, Iceland. Reproduced with permission from Albu M, Banks D, and Nash H (1997) *Mineral and Thermal Groundwater Resources.* London: Chapman and Hall.

The geothermal installation at Wairakei, New Zealand is of this type (Figure 8). Other such installations are situated at Tongonan, Philippines, and Cerro Prieto, Mexico.

The most efficient use of volcanic-related thermal water is in Iceland, where 'cascading' applications have been developed, i.e., each step in the process uses a successively lower temperature (Figure 9). Chemicals and other raw materials may be extracted during the sequence of operations. It is sobering to realize how small a contribution geothermal energy makes towards meeting worldwide energy requirements. Table 1 lists

Table 1 Use of geothermal energy worldwide[a]

	Electric production		*Direct use*	
Country	*Power (MW)*	*Annual use (GWh)*	*Power (MW)*	*Annual use (GWh)*
Algeria	—	—	1	5
Argentina	0.7	3.5	—	—
Australia	0.4	0.8	—	—
Austria	—	—	21.1	84
Belgium	—	—	3.9	19
Bosnia–Herzegovina	—	—	33	230
Bulgaria	—	—	95	346
Canada	—	—	3	13
China	32	175	1 914	4 717
Costa Rica	120	447	—	—
Croatia	—	—	11	50
Czech Republic	—	—	2	15.4
Denmark	—	—	3.2	15
El Salvador	105	486	—	—
France	4	24	309	1 359
Georgia	—	—	245	2 145
Germany	—	—	307	806
Greece	—	—	22.6	37.3
Guatemala	5	—	—	—
Hungary	—	—	750	3 286
Iceland	140	375	1 443	5 878
Indonesia	590	4 385	—	—
Ireland	—	—	0.7	1
Israel	—	—	42	332
Italy	768	3762	314	1 026
Japan	530	3530	1 159	7 500
Kenya	45	390	—	—
Macedonia	—	—	75	151
Mexico	743	5 682	28	74
New Zealand	345	2900	264	1 837
Nicaragua	70	250	—	—
Philippines	1848	8000	—	—
Poland	—	—	44	144
Portugal	11	52	0.8	6.5
Romania	2	?	137	528
Russia	11	25	210	673
Serbia	—	—	86	670
Slovakia	—	—	75	375
Slovenia	—	—	34	217
Sweden	—	—	47	351
Switzerland	—	—	190	420
Thailand	0.3	2	2	8
Tunisia	—	—	70	350
Turkey	20	71	160	1 232
Ukraine	—	—	12	92
United States	2850	14 600	1 905	3 971
Europe	936	4309	4 368	20 505
America	3883	21 529	1 908	3 984
Asia	3031	16 092	3 075	12 225
Oceania	345	2901	264	1 837
Africa	45	390	71	355
Total	8240	45 220	9 686	38 906

[a]Reproduced from Arnorsson S (2000) Geothermal energy. In: Hancock PL and Skinner PJ (eds.) *Oxford Companion to the Earth*, pp. 437–440. New York: Oxford University Press. Data for megawatts (MW) and gigawatt-hours (GWh) are valid for 1998.

Figure 10 The Roman bath at Bath, England, fed by hot water from King's Spring. Reproduced from Kellaway GA (1991) *Hot Springs of Bath: Investigation of Thermal Waters of the Avon Valley*. Bath, UK: Bath City Council.

Table 2 Total heat output from hot springs at Bath, England, November 1990

Source	*Water output ($m^3\,day^{-1}$)*	*Temperature excess (above 10°C)*	*Heat output ($Mcal\,day^{-1}$)*
King's Spring	1180	35°	41 300
Hetling Spring	75	37°	2775
Cross Bath Spring	35	31°	1085
Total	1290		45 160[a]

[a]Adding 20% for losses gives a total heat output of 54 192 Mcal day^{-1}.

Table 3 Basic geochemistry of geothermal waters of Bath, England

Variable[a]	*King's Spring and Stall Street borehole*	*Hetling Spring*	*Cross Bath Spring*	*Kingsmead borehole*	*Sports Centre borehole*	*Hot Bath Street borehole*	*Weston borehole*	*Batheaston coal shaft*
Temperature	44°C	47°C	41°C	35°C	20°C	41°C	13°C	17°C
Na	187	195	183	300	550	200	420	476
Ca	390	358	380	360	296	367	110	261
SO_4	1010	1015	1050	1060	1000	1020	795	940
Cl	286	340	288	440	770	350	235	964
HCO_3	199	193	189	154	198	—	245	81
Mg	53	57	54	56	71	—	35	65
Si	21	21	—	17	7	13	—	10
Fe	1	0.5	0.2	1	—	2	0.2	1.5

[a]Mineral content is given in milligrams per liter.

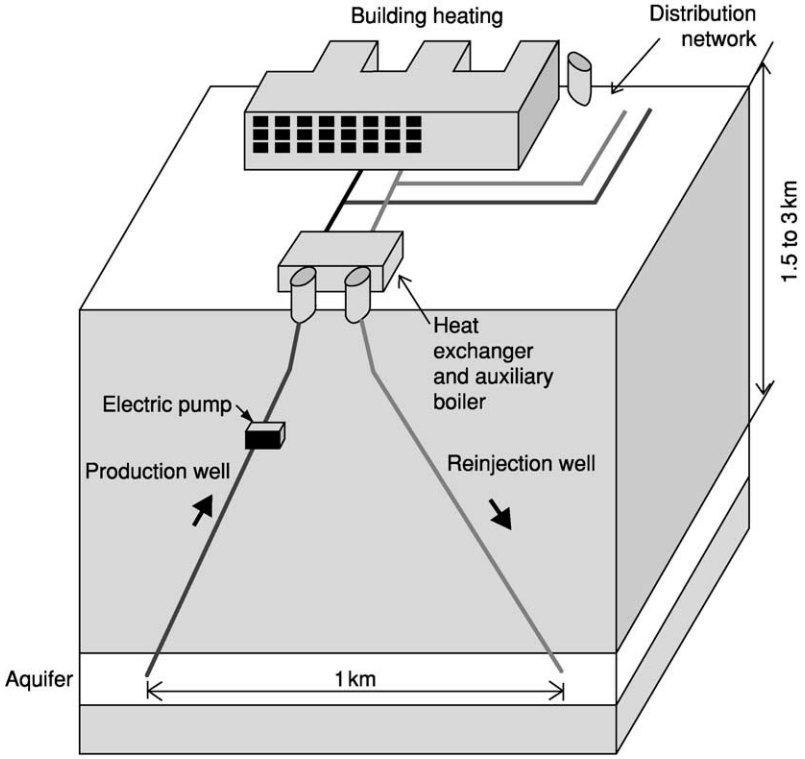


Figure 11 Doublet system with up and down circulation as used in the Paris Basin. Reproduced with permission from Albu M, Banks D, and Nash H (1997) *Mineral and Thermal Groundwater Resources*. London: Chapman and Hall.

the countries that utilize geothermal resources, and the sum total of electrical power produced is less than 10% of the electricity consumption of Great Britain.

Non-Volcanic-Related Geothermal Processes

Geothermal processes unrelated to volcanic activity are quite widespread. Because Earth's interior has a geothermal gradient, increasing in temperature with depth, when precipitation in the form of rainfall and snowfall descends deep enough through subsurface pores and passages, it is heated. Subsurface heated waters eventually rise to the surface under pressure, naturally as hot springs or artificially when tapped by boreholes. The first case is exemplified by hot springs in Bath and Bristol, England; the Paris Basin in France is an example of a region where 'heat mining' activities have been developed.

Bath and Bristol Hot Springs

The hot springs in western England (Figure 10) were utilized for spa purposes by the Romans, and possibly earlier, by the Celts; after the end of Roman rule, the baths fell into disrepair, but were rejuvenated during the Elizabethan Era and were an attraction in the Bath region from the seventeenth century onwards. The heated waters in the natural springs of Bristol and Bath have no known volcanic association and are widely believed to relate to rainfall recharge in the Mendip Hills, about 40 km away. In this process, rainwater is routed downward through fissures to the Carboniferous limestone at 600–900 m below sea-level, and beyond that, to the Lower Palaeozoic rocks. There is up to 4600 m of section above the base of the Ordovician, and calculations show that the water could be heated in depth to 121°C. The water returning to the surface under Bath and Bristol (Hotwells), through a series of fissures and aquifers, reaches the surface (artesian conditions are locally operative) still hot, though having lost much of its heat. The passage to Bath from the Mendip recharge area is estimated to take 3000 years. Flow measurement, temperature, and heat output and the basic chemistry of the Bath waters are given in Tables 2 and 3. Radiogenic elements and dissolved

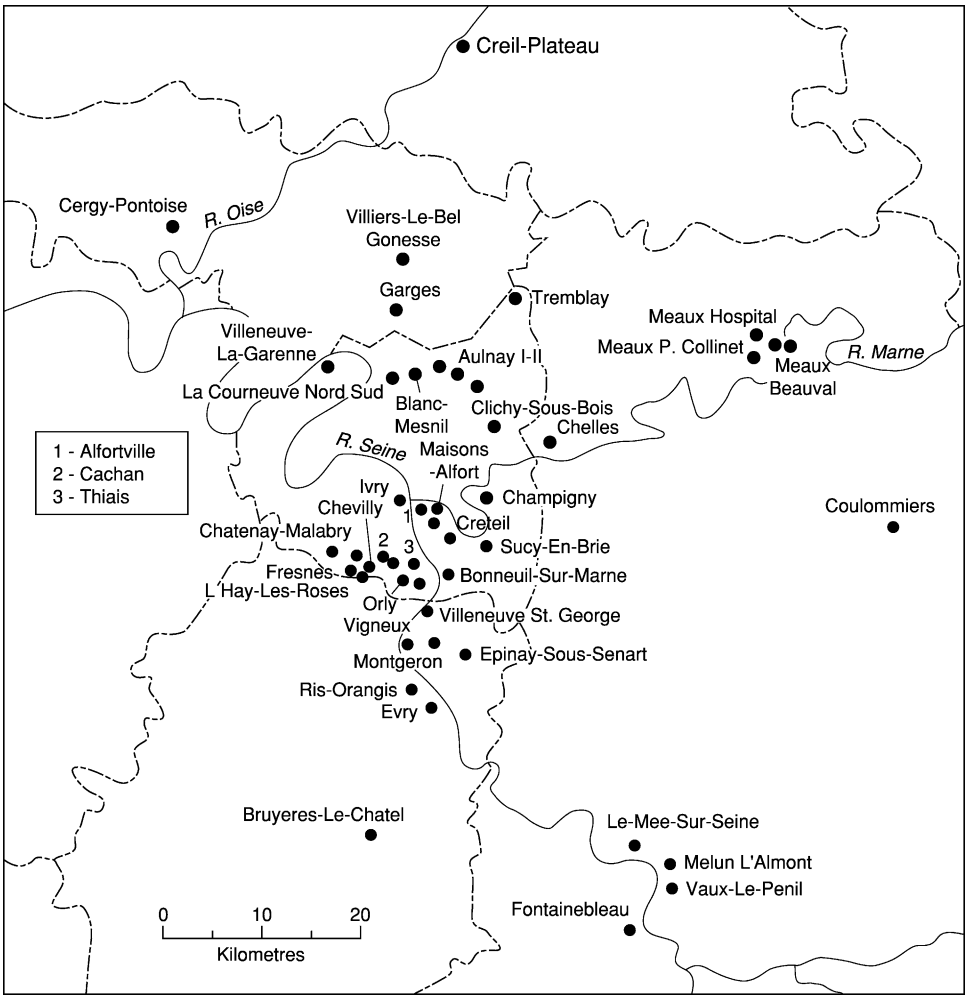


Figure 12 Geothermal utilization in the Paris Basin. Reproduced from McCall GJH, de Mulder EFJ, and Marker BR (1996) *Urban Geoscience*. Rotterdam and Brookfield, VT: Balkema.

noble gases in the waters are believed to come from Lower Palaeozoic rocks, being collected by the water on its passage through the substrata.

The Paris Basin

Rainwater in the urban outskirts of Paris, France, is carried down to deep aquifers in the same way as occurs at Bath and Bristol, but instead of using the heated waters for spa purposes, the practice in the Paris Basin is to use the heat mainly to service large estates of apartments. Borehole pairs form up and down doublets (Figure 11), and the extracted geothermal water is utilized as a direct heat source. The waters have temperatures of 60–70°C. As shown in Figure 12, the French have developed this type of low-level geothermal utilization widely. A cascading process is utilized here as in Iceland. Geothermal heat energy is a very local resource when used in this way. Thermal resources further south in the Auvergne are volcanic related, but the Paris Basin occurrences are not.

Seafloor Geothermal Activity

Hot springs emanate in volcanic parts of the ocean floor. Volcanism occurs in mid-ocean ridges, island arcs, and ocean islands (hotspots), but the mid-ocean ridges have become famous for their black smokers, regions where clouds of black mineralized smoke accompany thermal water emissions and strange, blind tube worms cluster feed on the emissions (*see* **Tectonics:** Hydrothermal Vents At Mid-Ocean Ridges). These occurrences are closely related to the formation of ancient volcano-related sulphide mineral deposits. In the Red Sea, such submarine

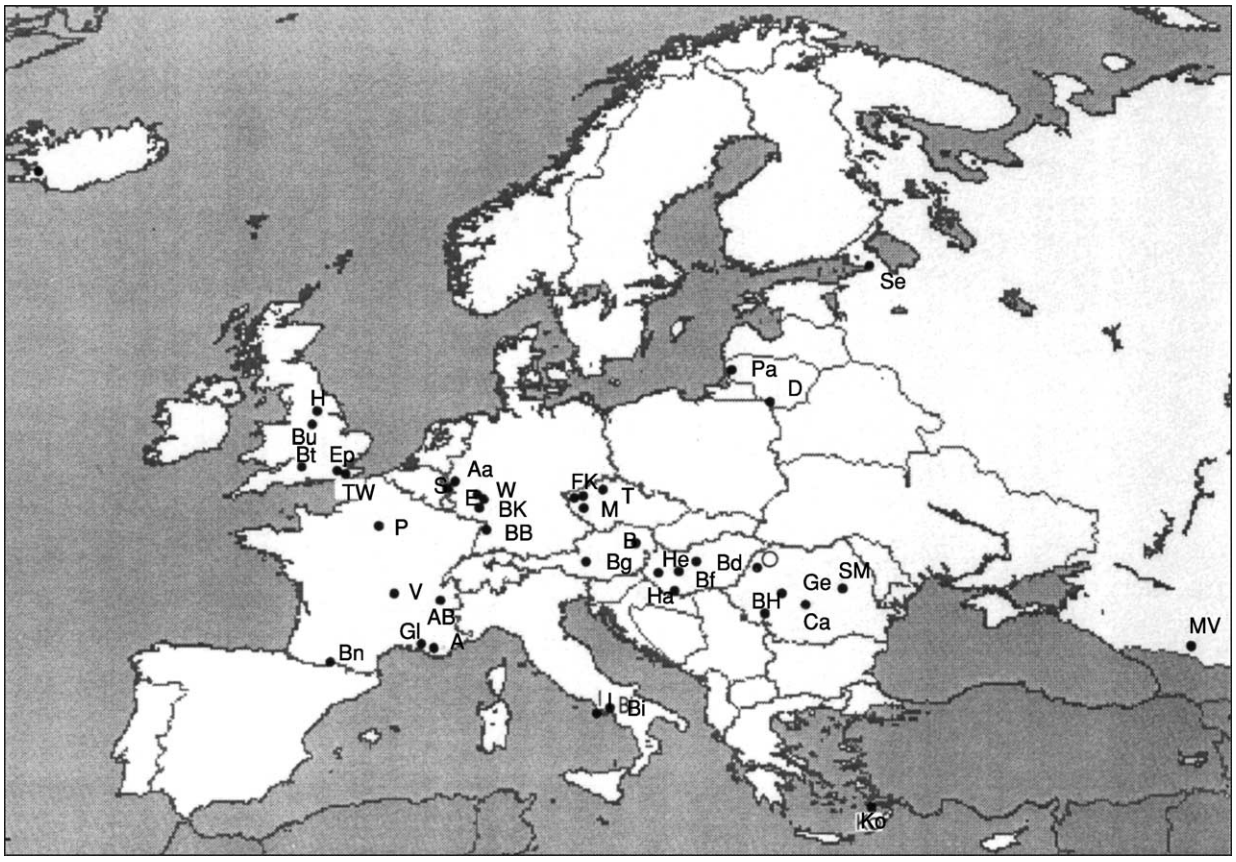


Figure 13 Sites where spas were once located throughout Europe. Reproduced with permission from Albu M, Banks D, and Nash H (1997) *Mineral and Thermal Groundwater Resources.* London: Chapman and Hall.

thermal springs are associated with zinc mineralization. There is probably little potential for economic recovery of minerals from such hot springs, though the hydrothermal deposits of gold in the Western Pacific might prove of economic significance.

Commercial Applications

There has been much research in Cornwall, England into utilizing the deeply buried hot rocks as an energy source, but this has so far not yielded any practical results. Health spas, however, are of economic interest. Spas were once numerous throughout Europe (**Figure 13**), and most utilized thermal waters. The Bath spa dates from Roman times, and the Roman thermal bath there (the King's Spring) is still in existence (**Figure 10**). Good accounts of it were written in the seventeenth century by Celia Fiennes, John Evelyn, and Samuel Pepys, and a recent account edited by GA Kellaway examines geophysical evidence for the origin of the hot springs. There were five baths originally, according to Celia Fiennes: Hot Bath, Lepours, Cross Bath, Kitchings, and King's Bath, but not all have survived. In 1978, after bathing in the waters at Bath, a child died from meningitic infection by *Naegleria fowleri*, a free-living amoeba found in hot spring water, and the spa was closed. This danger has been eliminated following a major investigation and restoration; the basis of success was correct assessment of the geological structure of the springs and understanding the ecological requirements of *Naegleria*. *Naegleria* requires oxidized water for sustained growth, thus the baths must be replenished by drilling for amoeba-free water, preventing oxidation of the water before it reaches the surface, and isolation of this water from oxidized water. The Bath spa is scheduled to reopen in 2004 or 2005.

Travertine and tufa, widely used as building stones and cladding, are another important economic contribution of geothermal features. Travertine is particularly widely used for cladding McDonald's restaurants. The cream-coloured stone comes from Tivoli, Italy; it often contains cavities made by

Figure 14 Travertine cladding on a McDonald's Restaurant, Hereford, England. Photograph by GJH McCall.

dissolution of reeds, and these have to be filled artificially, after which the stone makes a striking cladding, showing sinuous algal banding (Figure 14).

See Also

Biosediments and Biofilms. **Earth:** Mantle; Crust. **Minerals:** Carbonates. **Tectonics:** Hydrothermal Vents At Mid-Ocean Ridges. **Volcanoes**.

Further Reading

Albu M, Banks D, and Nash H (1997) *Mineral and Thermal Groundwater Resources*. London: Chapman and Hall.

Arnorsson S (2000) Geothermal energy. In: Hancock PL and Skinner PJ (eds.) *Oxford Companion to the Earth*, pp. 437–440. New York: Oxford University Press.

Cargo N and Mallory BF (1977) *Man and His Geologic Environment*. Reading, MA: Addison-Wesley.

Green J and Short NM (1971) *Volcanic Landforms and Surface Features: A Photographic Atlas and Glossary*. New York, Heidelberg, Berlin: Springer-Verlag.

Kellaway GA (1991) *Hot Springs of Bath: Investigation of Thermal Waters of the Avon Valley*. Bath: Bath City Council.

Mazzoni A (1954) *The Steam Vents of Tuscany and the Larderello Plant*. Bologna: Arti Grafiche Calderini.

McCall GJH (1967) *Geology of the Nakuru–Lake Hannington–Thomsons Falls Area. Geological Survey of Kenya, Report No. 78*. Nairobi: Ministry of the Interior, Geological Survey of Kenya.

McCall GJH, de Mulder EFJ, and Marker BR (1996) *Urban Geoscience*. Rotterdam and Brookfield, VT: Balkema.

Scarth A (1997) *Savage Earth*. London: Harper Collins.

GLACIERS

See **SEDIMENTARY PROCESSES: Glaciers**

GOLD

M A McKibben, University of California, CA, USA

Characteristics and Uses

Gold is a rare heavy metal that is soft, malleable, ductile, and bright sun yellow in colour when pure (Figure 1). The last property is reflected by its chemical symbol, Au, which comes from the Latin word *aurum* ('shining dawn'). Gold resists chemical attack and corrosion, has excellent electrical conductivity, and reflects infrared radiation.

Humans have made use of gold for more than 40 000 years. Because of its inertness and value, most of the gold that has ever been mined (about 130 000 metric tons) is still in use! Most gold is hoarded, in the form of bullion, coins, and jewellery. This usage stems from the metal's aesthetics, its role as a medium of exchange among banks and governments, and its perceived value by individuals and families as a hedge against economic uncertainty. Much like diamonds, gold has an emotional (and sometimes irrational) appeal that drives its free-market value to levels far above those that would otherwise be justified solely by its practical uses in technology.

Practical uses of gold take advantage of its chemical inertness, electrical conductivity, malleability, and ductility. Gold's high electrical conductivity and solderability make it excellent for creating reliable electrical contacts; it can be drawn into wires thinner than a human hair and pounded into sheets thin enough to pass light. It is therefore used in electronics (circuit boards, connectors, contacts, thermocouples, potentiometers), corrosion-resistant processing equipment (acid vats), infrared reflectors (windows of high-rise buildings, astronaut's helmet visors), and dental applications (fillings, crowns, etc.).

Pure gold is too soft, malleable, and ductile for uses that require physical endurance, particularly jewellery, so instead mixtures or alloys of gold with other metals (copper, nickel, silver, platinum, etc.) are used to enhance an object's durability. The gold content of such alloys can be defined in terms of carats, or parts of gold per 24 parts of total metal by weight. Durable jewellery is commonly made from 14 carat gold (containing 58.3% gold). Knowing the remaining metals used in gold alloys can be important to people who are allergic to specific metals in jewellery, such as nickel, particularly when the skin is pierced (earrings, nose rings, etc.).

The term 'gold-filled', which is used in jewellery making, means a layer of gold alloy that is placed over a less valuable core of base metal – in a few countries such layers are required by law to be at least 10 carats. 'Rolled gold plate' may be applied to layers that are less pure. Gold 'electroplate' jewellery must have at least seven millionths of an inch of gold overlaid, otherwise terms such as 'gold flashed' and 'gold washed' must be used.

Ironically, during the Spanish exploitation of the New World, comparatively 'worthless' platinum objects from South America were sometimes plated over with gold and passed off as pure. In modern times the relative values of these two metals are usually reversed, and the opposite strategy would be more lucrative. The consumer must be cautious at all times!

Another scale for expressing the purity of gold (mainly in bars, ingots, and coins) is fineness, based on parts of gold per 1000 parts of total metal. A gold bar that is '995 fine' thus contains 99.5% gold by weight.

The most common unit of weight for gold is the troy ounce, which is equal to 1.097 avoirdupois ounces or 31.10 g. A typical bar of gold (such as held by central banks and governments) weights about 400 troy ounces, or about 27.5 pounds (= ~12.4 kg). A troy pound contains 12 troy ounces, and a troy ounce contains 20 troy pennyweights, the latter being the unit of weight most often used in jewellery. A troy pennyweight (dwt) equals 1.555 g.

Mineralogy, Geochemistry, and Natural Concentration of Gold

Gold, silver, and copper are often associated in nature, because each has a lone outer S-orbital electron and belongs to the IB transition-metal subgroup of the periodic table. Other elements commonly associated with gold include arsenic, antimony, bismuth, iron, lead, and zinc.

In addition to the zero-valent elemental form, gold has two oxidation states: Au^{+} (aurous) and Au^{3+} (auric). Most gold minerals contain zero-valent or aurous gold (Table 1). About 40 natural gold minerals are known; native gold and electrum are the most common forms, tellurides are rare, and other forms are very rare. Because most gold ore occurs as

Figure 1 Polished slab of hydrothermal gold–quartz vein ore from the Hollinger Mine, Porcupine District, Ontario, Canada. This occurrence is typical of orogenic lode (greenstone belt) gold. The width of the image is approximately 1 cm.

Table 1 Examples of gold minerals

Native gold	Au
Electrum	(Au,Ag)
Cuproauride	(Au,Cu)
Rhodite	(Au,Rh)
Amalgam	Au_2Hg_3
Calaverite	$AuTe_2$
Krennerite	$(Au,Ag)Te_2$
Sylvanite	$(Au,Ag)Te_4$
Petzite	Ag_3AuTe_2
Nagyagite	$Pb_5Au(Te,Sb)_4S_{5–8}$
Uytenbogaardtite	Ag_3AuS_2
Aurostibite	$AuSb_2$
Fischesserite	Ag_3AuSe_2

the relatively pure native form, smelting or roasting of gold ore is sometimes not required.

Average gold concentrations in upper-mantle and crustal rocks vary by about an order of magnitude, typically between 1 ppb and 10 ppb. Mafic rocks generally contain higher concentrations than felsic rocks, and some minerals such as magnetite and pyrite may act as natural 'concentrators' to enrich gold in rocks. Continental crust averages about 4 ppb gold, whereas seawater contains only about 0.01 ppb gold.

Gold can be mobilized and concentrated into economic deposits by many geological processes, including igneous melting and crystallization, hydrothermal leaching and precipitation, and fluvial hydraulic sorting. In silicate melts, gold tends to follow other chalcophile elements and is often concentrated in metallic sulphide and oxide phases. When magmas ascend to shallow levels in the crust, gold may also partition into chlorine- and sulphur-rich aqueous fluids, which are expelled into the surrounding wall-rock as the magma crystallizes. Vapour-phase transport of gold chlorides has been observed in some volcanic emissions. Metamorphism and consequent dewatering of mafic rocks in orogenic zones can also lead to gold remobilization and precipitation in veins, along with quartz and carbonate minerals. In shallow crustal settings where hot aqueous fluids circulate and sometimes boil within permeable volcanic and sedimentary rocks (hot-spring or epithermal systems), gold seems to be preferentially transported and precipitated with silica (quartz) in veins and pores.

Most major gold deposits have resulted from hydrothermal processes (Figure 2) or from subsequent erosion and hydraulic sorting of rock grains derived from hydrothermal deposits (*see* **Tectonics:** Hydrothermal Activity; **Mining Geology:** Hydrothermal Ores). Thus, the simplistic terms lode (vein) and placer (sediment) for the origin of gold dominate the rich history and legal framework of gold exploration and mining. Lodes form when hydrothermal fluids pass through fractured permeable rock and deposit gold and other minerals as tabular veins and adjacent pore fillings. Placers form where surface waters (mainly fluvial) experience a change in hydraulic conditions (especially velocity), forcing them to drop suspended sediment grains of a certain density, size, or shape. Tributary intersections, meander loops, and canyon mouths are examples of locations where placer gold can accumulate in rivers.

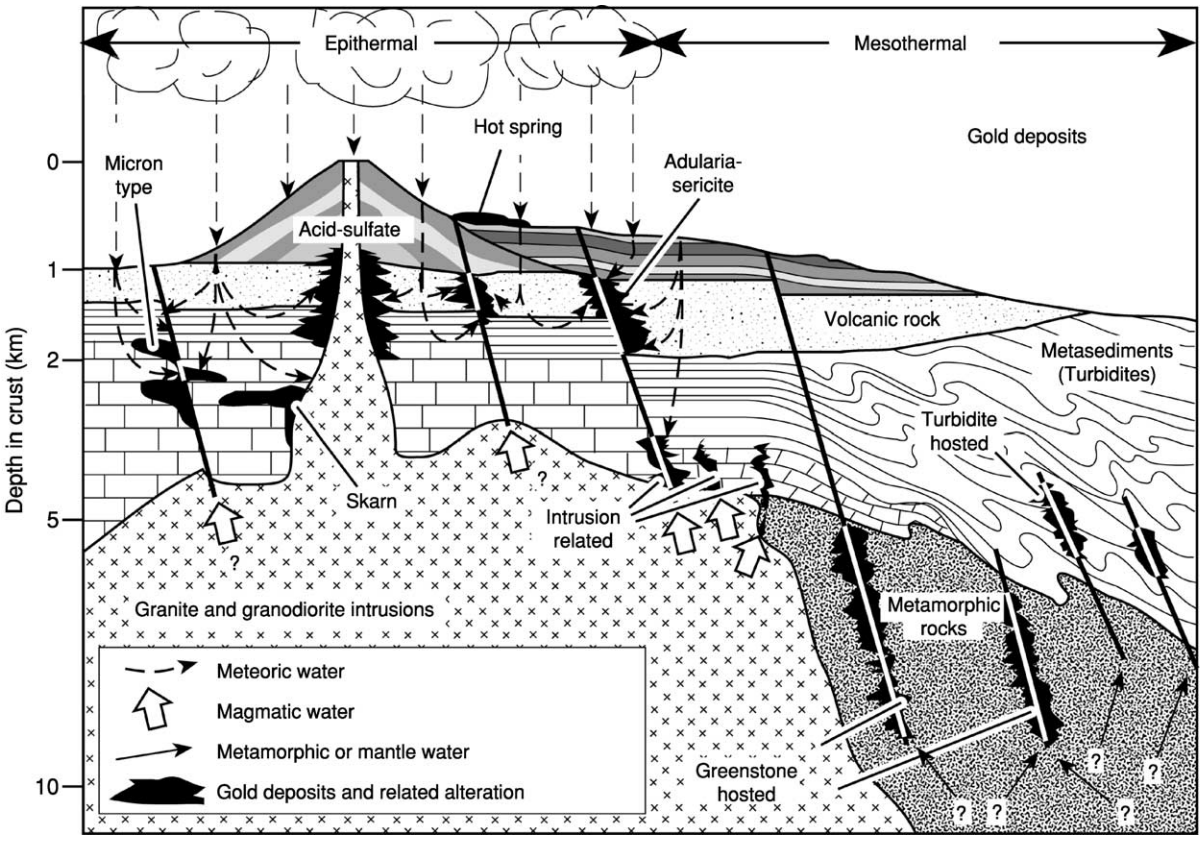


Figure 2 Schematic illustration of the wide variety of occurrences of hydrothermal gold deposits in continental crust. The terms epithermal and mesothermal refer to shallow-crustal and mid-crustal level deposits, respectively. Reproduced from Kesler SE (1994) *Mineral Resources, Economics and the Environment*. New York: MacMillan.

In most hydrothermal fluids, gold is transported in the form of aurous aqueous complexes with hydroxide (OH^-), chloride (Cl^-), and bisulphide (HS^-) anions. Therefore, gold may be precipitated from such fluids via geochemical processes that destabilize the aqueous complexes and reduce the aurous gold to the elemental state, for example:

$$Au(HS)_2^- + H^+ + 0.5H_{2(g)} = Au_{(S)} + 2H_2S.$$

Environmental changes that can destabilize such complexes include cooling, pH changes, and changes in fluid oxidation state that accompany boiling, mixing, and chemical reaction of migrating hydrothermal fluids with wall rock and other crustal waters. Many gold deposits are associated with minerals, such as silica (quartz), carbonate, sulphates, and clays, that form during such geochemical changes.

Common host minerals for gold include iron sulphides, oxides, and hydroxides, and such phases may thus play a role in the local reduction of aurous gold. For example, the oxidation of ferrous to ferric iron can facilitate the localized reduction of aurous gold from solution in the presence of ferrous mineral phases:

$$AuCl_2^- + Fe^{2+} = Au_{(s)} + Fe^{3+} + 2Cl^-$$

The occurrence of tiny grains of gold within larger mineral hosts (pyrite, oxides, hydroxides, etc.) and the presence of encapsulating and metallurgically recalcitrant phases such as silica can be significant factors influencing the degree to which mined gold ores must be ground, leached, and otherwise processed to release the gold. Failure to examine and characterize adequately the distribution and texture of fine gold occurrences within ores has led to some spectacular (and very expensive) mistakes in designing mineral processing facilities at some gold mines.

In the past few decades, the advent of inexpensive bulk-mining and cyanide heap leaching methods has made rocks containing as little as 1 ppm gold (0.3 troy ounces per ton) rich enough to be ores profitable for mining. Many of these rocks occur in the feebly mineralized (and previously ignored) areas surrounding now mined-out richer vein deposits.

Though likewise hydrothermal in origin, such bulk-disseminated low-grade gold ores do not fall conveniently into the traditional 'lode' category, leading to some interesting legal ambiguities in their status. Rigid governmental regulations for the dimensions of claim blocks, designed originally for idealized tabular 'lode' ore bodies, cannot be easily applied to more disseminated and irregular forms of mineralization.

Gold Ore Deposits

Being so valued a metal, gold has been actively sought out and mined in a wide variety of geological settings, wherever it has occurred. Only some of the major geological types of gold deposit will be briefly described here; some of the articles in the further reading section provide a more exhaustive catalogue of world gold-deposit types.

Archaean Gold–Quartz Conglomerates ('Palaeoplacers')

A majority of the world's gold has been mined from what were traditionally viewed as ancient placer deposits (Figure 3). While many people might envision a wizened prospector panning or sluicing for loose gold nuggets along a modern river bank, the most important of these ancient deposits are in hard lithified metamorphosed coarse sediments of Archaean age (2.3–2.8 Ga). The Witwatersrand ('Wits') gold fields of South Africa have produced more than one-third of the gold mined on Earth since 1886, and one single Wits mine (the Vaal Reefs mine) has produced almost as much gold as all of the gold mines in the USA and Australia combined. Similar ancient 'palaeoplacer' districts include the Elliot Lake–Blind River district in Canada. The gold in Wits ore is typically associated with detrital pyrite, uraninite, and carbon in quartz pebble conglomerates and quartzites that have been subjected to peak metamorphic temperatures of 300–400°C.

Placer and hydrothermal-replacement models for Wits gold have both been advanced, but most recent evidence favours a hydrothermal origin for the gold in these ancient sediments. The emplacement of the gold seems to have occurred at the time of peak metamorphism, via low-salinity fluids channelled along fault structures, unconformity surfaces, and bedding planes. Reaction with carbon- or iron-bearing (pyritic) rocks located just above the unconformity surfaces caused gold precipitation. The ultimate origin of the gold was probably underlying mafic rocks, rather than eroded lode quartz veins as traditionally implied by sediment clasts and the placer model.

In addition to their significance as major sources of gold, the Wits palaeoplacers provide strong evidence that the Earth's Archaean atmosphere was very different from our modern atmosphere. Detrital grains of pyrite and uraninite cannot survive significant fluvial transport in our oxygen-rich modern atmosphere; their abundance in the Wits placers indicates that the ancient atmosphere was less oxygenated and perhaps more

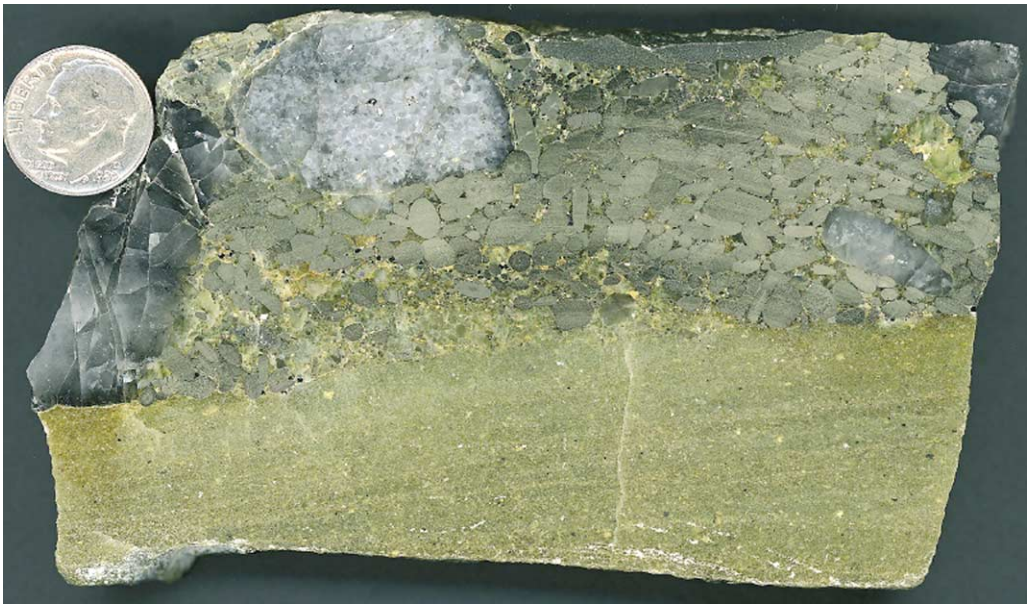


Figure 3 Polished slab of Archaean quartz-pebble conglomerate gold ore from the Vaal Reefs Mine, Witwatersrand District, South Africa. Rounded detrital pyrite is clearly visible. Curved marks across the pyrite grains are saw-blade artefacts. A US 10 cent coin is shown for scale.

enriched in carbon dioxide and/or methane than the modern atmosphere (*see* **Atmosphere Evolution**).

Orogenic Lode Gold

The Mother Lode District of California may be the most famous of a class of lode gold-ore deposits that are found in metamorphic terranes associated with accretionary plate tectonics. They have also been called greenstone-belt and mesothermal lode ores (Figure 1). The discovery in 1848 of placer gold at Sutter's Mill, downstream from this district, led to the California Gold Rush of 1849, one of the most significant New World social migration and settlement events in recent history. To this day many people still wear a popular miner's legacy from this gold-driven rush: Levi's blue jeans.

Orogenic lode gold deposits range in age from Archaean (such as the Barberton district in South Africa) to Mesozoic–Tertiary (such as the Mother Lode district). There remain a number of theories for their origins, but all seem to be associated with major accretionary orogenic events (plate collisions) and occur near major translithospheric structures (shear zones) bounding metamorphosed terranes. Typical conditions of formation are 250–400°C and 1–3 kbar.

The mineralization is typically of syn- to post-peak metamorphic age, associated largely with mid-crustal-level greenschist facies (gold plus quartz plus carbonate) mineral assemblages. The deposits are structurally controlled, often along splays associated with high-angle faults; the gold is syn-kinematic and localized near the brittle–ductile crustal transition. Fluid pressures range from superlithostatic to sublithostatic within the brittle–ductile shear zones. Within individual deposits, the quartz–carbonate vein systems are sometimes more than 2 km in vertical extent. The most common alteration minerals associated with the gold are quartz, carbonate, white mica, chlorite, and pyrite; both gold and silver are enriched in the deposits, while the abundance of base metals is relatively low. Fluid-inclusion data indicate that the aqueous fluids that formed the ores were of low salinity (salts) and rich in carbonic components (carbon dioxide, methane) (*see* **Fluid Inclusions**).

These hydrothermal or metasomatic systems thus seem to have developed during plate collisions as deep metamorphic fluids were expelled during shearing and deformation and transported into structurally favourable settings where they deposited the gold.

Epithermal Gold

As the name implies, epithermal gold deposits form in shallow (1–2 km depth) anomalously hot (100–300°C) crustal settings. Most develop around areas of volcanic activity, with distal to proximal magmatic heat flow driving the convective circulation of groundwaters in brittle fractured rock. Boiling hot springs and geysers near active volcanoes are modern examples of the surface manifestations of such systems. Some epithermal systems may grade with depth into magma-hydrothermal systems that are more characteristic of porphyry systems. The terms 'low-sulphidation state' and 'high-sulphidation state' are used to describe epithermal systems dominated by meteoric and magmatic fluids, respectively.

Epithermal deposits form most often in volcanic arc settings above subduction zones, where sustained high heat flow, active tectonism, and meteoric precipitation encourage fluid heating and circulation. The vein and disseminated gold ores in epithermal systems are typically associated with silica, alkali feldspar, and other secondary minerals (clays, micas, sulphates) formed by acid alteration of the host rock.

Examples of major epithermal gold deposits include Hishikari (Japan) (Figure 4), El Indio (Chile), Round Mountain (Nevada, USA), and Cripple Creek (Colorado, USA).

Carlin-Type Gold

The term 'Carlin-type' is applied to enigmatic mid-Tertiary gold deposits in western North America

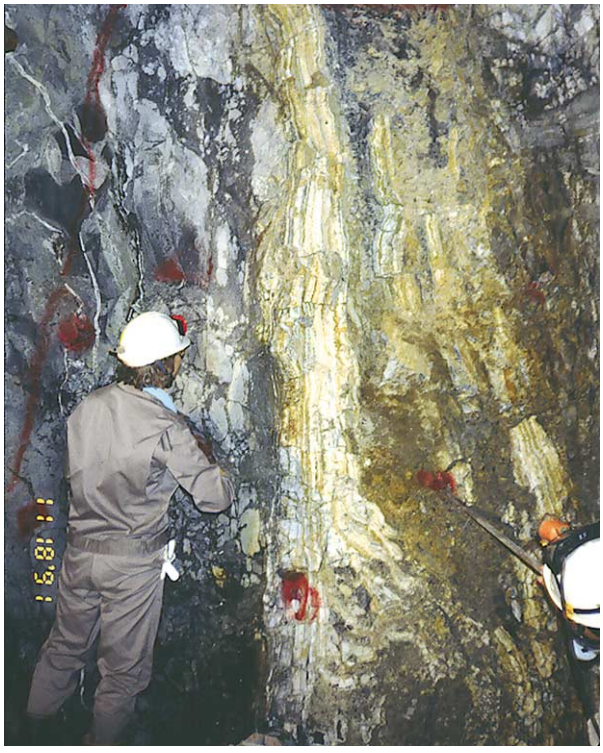


Figure 4 Banded crustiform epithermal veins of quartz in brecciated rock comprise the main ore type at the underground Hishikari Mine in Japan, one of the world's richest gold mines.

(epitomized by the Carlin district in Nevada) whose origin seems to fall somewhere between the epithermal and orogenic lode gold types. The epigenetic gold in these deposits is finely disseminated and hosted by Palaeozoic calcareous sedimentary rocks. Arsenical pyrite, marcasite, and arsenopyrite are the main host minerals. The host rocks exhibit evidence of carbonate dissolution, argillic alteration, silicification, and sulphidation. It has been proposed that reduced hydrogen sulphide-rich auriferous fluids migrated up pre-existing fault structures and spread laterally into permeable reactive carbonate rocks, depositing gold at temperatures of 150–250°C at depths of more than 2 km. The direct role (if any) of crustal magmatic processes in driving circulation and contributing components to the fluids is not clear. More recent models favour a metamorphic origin, driven by heat from a mantle plume underlying the region during crustal extension.

Though Carlin-type gold deposits are not as widely distributed in space and time as the other types of gold deposit described above, their more recent discovery was instrumental in driving the development of inexpensive cyanide heap leaching technology in the late twentieth century, making these and other lower-grade disseminated gold deposits economic and competitive with the traditionally dominant Witwatersrand gold deposits in South Africa.

Gold as a By-Product in Other Metal Deposits

Gold is frequently mined as a by-product from ore deposits dominated by base-metal production (Cu, Pb, Zn). Such deposits include porphyry, skarn, sediment-hosted, and volcanogenic massive sulphide types. In some cases, the addition of by-product gold makes the base-metal deposit economic. There are also deposits of these types that are mined only for gold.

Examples of major base-metal deposits with significant by-product gold include Grasberg (Indonesia) (porphyry and skarn copper), Bingham (Utah, USA) (porphyry copper), Red Dog (Alaska, USA) (sediment-hosted zinc), Rammelsberg (Germany) (sediment-hosted zinc and lead), and Mount Morgan (Queensland, Australia) (volcanogenic massive sulphide copper).

Gold Mining

Placer Mining

Mining of unconsolidated placer gold at the Earth's surface has taken place for centuries, ranging from the simple panning and sluicing techniques still used by weekend 'hobby' prospectors to the far more massive dredging and hydraulic techniques used by mining companies. All of these techniques involve the use of flowing water to wash away (sluice) less dense grains (mainly silicates) from the heavier gold grains contained in sediment. Dredges are barges with automated conveyor-belt mechanisms for scooping up sediment at the front end, processing it on board to extract the gold, and then dumping the worthless tailings out the back end. The ecologically destructive effects of dredges on river systems can be significant, not only locally, but also far downstream because of silting.

Following the 1849 California Gold Rush, unconstrained hydraulic mining of placers with large water cannons (monitors) and river dredges wreaked environmental havoc on scenic landscapes and waterways from the Sierran foothills to San Francisco Bay, prompting John Muir to form the Sierra Club and ultimately forcing government legislation banning most such practices in the USA. However, such practices still continue in less-developed countries, particularly in Latin America and the Pacific Rim. Offshore (marine) placers, where rivers have discharged their sediment loads into the oceans, are also mined for gold.

Relying solely on hydraulic techniques to separate gold from sediment is not 100% effective, especially for fine gold grains. Amalgamation, or dissolving gold grains in liquid mercury, is far more effective and was used extensively until the development of cyanidation in the 1890s. Copper-lined tubs were coated (amalgamated) with mercury and the placer concentrate was placed in the tub along with more liquid mercury. The tub was then agitated until the gold amalgamated and adhered to the copper amalgam. The amalgam was then scraped off and the gold was reconstituted to a solid state via evaporation (distillation) of the mercury. Unless it was captured by retorts that recondense the vapour, the mercury was released into the environment and entered the food chain (and humans) mainly via fishes. The historic and modern use of amalgamation to recover gold from placers continues to cause a serious health problem in some areas. Areas downstream of the Mother Lode district still suffer from mercury contamination, more than a century after mining has ceased. Amalgamation of placer gold is still used in remote areas of Brazil by native miners who have no other means of employment and cannot afford retorts or cyanidation; their communities suffer from mercury poisoning as a consequence.

Hard-Rock Mining

Upon rapid depletion of the most easily found New World placer deposits in the late nineteenth century, gold miners' attention shifted to hard-rock mining

with the discovery of South Africa's Witwatersrand deposits in 1886. These enormous deposits have since been the world's major gold source, yielding over 30 000 metric tons of the metal. Hard-rock mining involves drilling, blasting, digging, and scraping, which increase the costs far above those of placer mining. Open pits and underground workings need to be shored up and stabilized, and tailings disposed of properly. For underground operations, engine exhaust, dust, and rock gas must be vented and fresh air continuously injected for workers to survive. Those mines reaching below the water table require constant pumping.

Surface mining of hard-rock gold ore is feasible even at modest grades, but costs rise rapidly with depth, and so underground mining is possible only if the grades and tonnages of ore are high. Some mines start out as open-pit operations and then convert to underground workings as the shallower ore is mined out. Though still quite rich in ore, many of the Witwatersrand mines are now quite deep and hot (3–4 km and up to 65°C) adding greatly to costs and reducing labour productivity. Dramatic increases in gold price in the late twentieth century fuelled global exploration and led to the discovery of many new gold districts, some of which are amenable to inexpensive cyanide heap leaching techniques. These competitive trends caused gold production from South African mines to peak in the 1970s and steadily decline thereafter. An ounce of gold recovered from Wits mines now costs roughly twice as much as one recovered from typical surface mines in the western USA. Adding further to the woes of the South African mining industry is the problem of a workforce that will be significantly affected by the AIDS virus, both in terms of reduced labour productivity and long-term healthcare needs.

Gold Ore Processing

Once hard-rock gold ores have been blasted and taken from the ground, their processing can involve several steps. If the gold is 'locked up' in phases such as tellurides or sulphides, then efforts must be made to separate these minerals from the rest of the rock. Crushing and grinding are accomplished by brute force, using steel jaws or cone crushers followed by ball or rod mills (giant tumblers containing steel balls or rods). Next comes classification, or segregation of ground mineral grains into fractions of different sizes. This step is accomplished by using screens of different mesh sizes, much like sifting flour. Following that, the different minerals are separated from one another by taking advantage of their differing densities or other physical properties. For example, an air-driven cyclone separator can effectively segregate dense gold and sulphide from less dense silicate minerals. These steps together constitute beneficiation and result in an ore concentrate that is separated from the worthless minerals or tailings. These steps typically take place as close to the mine as possible.

Finally there is the finishing step, in which the metal is extracted from the concentrate, purified, and readied for market. The precise steps taken to extract the gold depend strongly on the nature of the gold in the ore.

Concentrates containing only native gold can be treated by amalgamation (as described above). Aqueous solutions containing both an acid and an oxidant can also be used to dissolve gold. For example, aqua regia ($HCl + HNO_3$) dissolves gold as auric chloride

$$Au_{(s)} + 4HCl + HNO_3 = HAuCl_4 + 2H_2O + NO.$$

Since the 1890s, however, the most important method used to dissolve gold from concentrate has been cyanidation with oxygen

$$4Au_{(s)} + 8NaCN + O_2 + 2H_2O = 4NaAu(CN)_2 + 4NaOH.$$

The concentrate is placed in vats of cyanide solution and agitated to increase the contact time for the reaction between the grains and the solution. The pregnant (gold-bearing) liquor is then filtered from the solids and further processed to recover the gold, sometimes by reacting it with powdered zinc metal in a process known as cementation

$$Zn_{(s)} + NaAu(CN)_2 + H_2O + 2NaCN = Au_{(s)} + Na_2Zn(CN)_4 + NaOH + 0.5H_2.$$

Another method of enhancing gold recovery uses activated carbon (such as found in aquarium filters) to adsorb gold cyanide complexes from the pregnant liquor. Once loaded up with adsorbed gold (up to 250 troy ounces of gold adsorbed per ton of carbon) the carbon is leached with acid solutions at high temperature and pressure, releasing the gold to be recovered by cementation or electrowinning. Electrowinning involves passing an electrical current through steel-wool electrodes that are immersed in the pregnant leachate, causing the gold to plate out in place of the iron via an electrochemical reaction. The high surface area of the inexpensive steel wool makes the process very efficient and cost-effective.

Cemented or electrowon gold is impure (typically 70–75% gold) because it contains residual iron or zinc along with non-aurous metals derived from the ore, such as silver, copper, and mercury. Such impure material is refined into nearly pure gold by heating it

in crucibles in a furnace to more than 1000°C in the presence of silica, oxidizers, and fluxes. The impurities oxidize and melt to form a separate silica-rich slag, which floats on top of the nearly pure molten gold. The slag is skimmed off and the gold (containing more than 95% gold at this stage) is called dore. Molten dore is further refined to in excess of 99% purity by injecting it with chlorine, which picks up the remaining metal impurities and removes them as volatile chlorides and solid chloride salts. A final sequence of redissolution and electrowinning (or precipitation) improves the purity to more than 99.9% gold, which is then remelted and cast into gold bars for market.

When gold ores yield less than 80% of their gold after normal grinding and cyanidation, they are termed refractory – such ores typically contain gold locked up in sulphides, tellurides, and arsenides. Such ores require more expensive (finer) grinding and sometimes roasting methods to free up the gold prior to cyanidation. Roasting such ores in a furnace (to break the host phases down and release the gold) has environmental impacts because of the release of sulphur dioxide (which contributes to acid rain) and toxic arsenic gases.

Besides complicating the freeing up of gold during beneficiation, host phases can interfere with the effectiveness of cyanidation by having low solubility in the cyanide solutions, consuming oxygen or cyanide, or competing with gold to form aqueous cyanide complexes. Other phases in the ore, such as coatings of carbon or silica, or the presence of clay minerals can inhibit the reactions and adsorb gold and other ionic constituents, causing gold recoveries to be less than ideal.

High-pressure and high-temperature oxidation in strong acids is sometimes used to process refractory ores, but this adds substantially to costs. Other more recent (and less costly) methods include microbiological approaches, in which bacteria are used to break down refractory gold-bearing phases prior to cyanidation. Species such as *Thiobacillus ferrooxidans* catalyze the oxidative destruction of pyrite and arsenopyrite (*see* **Minerals:** Sulphides), freeing up gold grains that are locked up in such minerals.

Cyanide Heap Leaching

In many modern low-grade gold mines, particularly those of the western USA, the gold is finely disseminated in fractured near-surface rock that can be easily mined by open-pit methods and piled into heaps after only modest blasting and beneficiation. The clever and inexpensive technique of cyanide heap leaching is used to extract gold very effectively from such ore heaps, avoiding the costs associated with grinding and smelting more traditional high-grade hard-rock ores.

The ore heaps are piled on top of thick plastic mats, and a sprinkler system is installed on top of the heaps (Figure 5). A cyanide solution is sprinkled on the

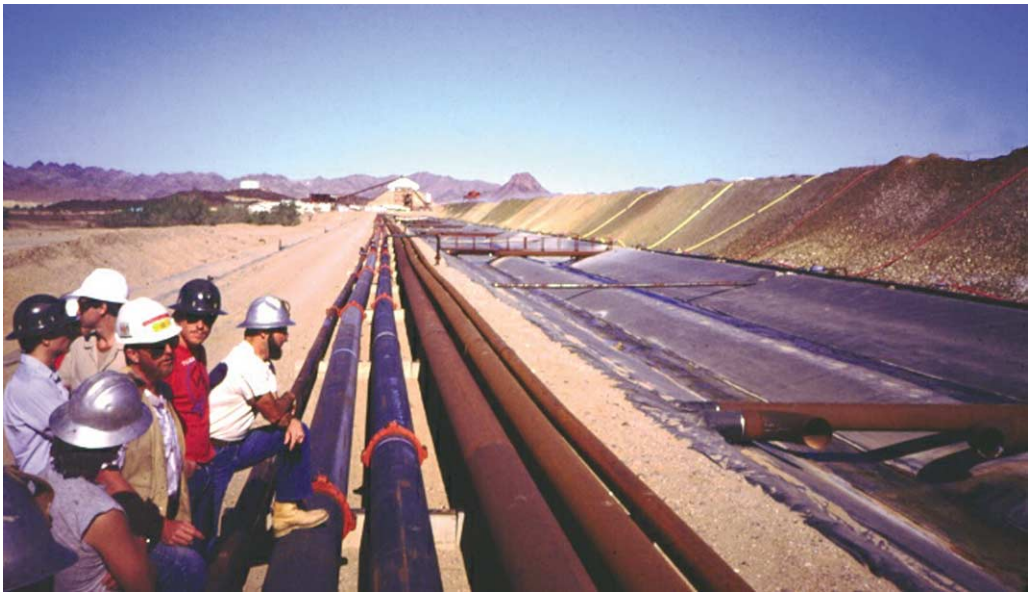


Figure 5 Heaps of low-grade epithermal gold ore are piled on black plastic mats and sprinkled with sodium cyanide solution at the open-pit Mesquite Mine, Imperial County, California, USA. One set of steel pipes delivers fresh solution to the sprinkler systems atop the heaps, while the other set carries away 'pregnant' gold-laden cyanide liquor that drains from the heaps and accumulates in the trenches.

heaps and trickles down through them, dissolving out the gold and accumulating on the mat surfaces. Drains and pipes carry the pregnant effluent from the mats to an activated-carbon circuit, where the gold is adsorbed and recovered; the stripped carbon and barren cyanide solution are then reused. Such technological innovations have made low-grade near-surface gold ores economically competitive to mine in recent decades, especially in the western USA. Some mines have tried combining cyanide heap leaching and microbiological approaches to reduce costs further. Gold recovery via cyanide heap leaching often exceeds 95%.

Though innovative and cost-effective, the use of open cyanide circuits in close proximity to wildlife and water resources has led to public concerns and calls for increased regulatory scrutiny of such mining operations. After some mis-steps in the 1970s and 1980s, the mining industry has responded to such concerns. Most operations now take pre-emptive measures to keep birds and other wildlife away from the heaps and cyanide ponds, such as using net covers and installing propane cannons that discharge at 30 s intervals. Although natural bacteria in the soil probably decompose most escaping cyanide before it can percolate into groundwater, the name 'cyanide' still carries a very negative connotation and therefore care must be taken to design a leaching circuit that is as leak-free as possible.

Gold Assaying

While visible nuggets of gold found near Sutter's Mill were indisputable evidence of rich placers and got the '49ers' of the California Gold Rush very excited, most modern gold ores are far less dramatic and typically contain gold that is very fine-grained and often invisible, even under some microscopes. Much of this gold is in solid solution in sulphide and arsenide minerals, and only an assay (chemical analysis) or the use of far more costly electron microscopy can prove its presence.

Gold can be qualitatively detected in samples by dissolving them in aqua regia and using a simple colorimetic test based on the addition of tin salts. When stannous chloride ($SnCl_2$) is added to the solution, stannic ions are formed and hydrolyze to a stannic hydroxide flocculant. As the tin is oxidized to the stannic state, gold is simultaneously reduced and adsorbed on the stannic hydroxide, turning the solution a brilliant purple colour called the 'Purple of Cassius'. This is a very sensitive test for gold, detecting as little as one part in a hundred million (10 ppb).

The most common and traditional assay technique for precisely quantifying gold content is the gravimetric lead fire assay. A powdered ore sample is mixed with carbon and lead oxide and fused at 900–1100°C for at least an hour. The carbon reduces the lead in the oxide to molten lead metal, and any metals present in the powder dissolve (amalgamate) into the molten lead. After cooling, the solid lead 'button' is placed in a special cup called a 'cupel' (made of bone, ash, or magnesia) and heated to 950–1000°C; the cupellation forces any remaining lead oxide and base metals in the remelted button to be absorbed by the cupel. After cooling there remains a metal 'bead' containing only silver, gold, and platinum-group elements. Treating the bead with nitric acid dissolves out the silver, leaving a bead composed only of gold and platinum-group elements to be accurately weighed.

Other methods used for gold determination include volumetic titration by iodine, spectrophotometric absorbance by gold bromide, and atomic emission and absorption methods.

Making judgments of ore grades based on limited assays of small samples can be risky because of the nugget effect, whereby the presence of one large gold grain in a small rock or drill-core sample can give the misleading impression that the entire mineralized area has high grades of gold. History is replete with cases of such high apparent assays (or assays of 'salted' barren rock) being used to sway and cheat naive investors. Assaying is most accurate when performed by trustworthy individuals on multiple large samples of mineralized rock.

Knowledge of the local geology can also help in recognizing fraudulent or misleading assay claims. In an undercover news investigation, the author once witnessed scam artists impressing naive potential investors with 'representative' vials of coarse gold nuggets – but the nuggets were far too large and rounded to have come from the local fine-grained dry lake beds that were being touted as the site of rich placer gold.

Gold Markets and Economics

Three events in human history have lead to dramatic increases in global gold production: the discovery of the New World in 1492; the California Gold Rush of 1849; and the end of government-regulated gold prices in 1968.

Governments and individuals have long valued gold as a measure of wealth and a hedge against economic uncertainty. Many governments back their paper currency with gold reserves held in central banks, and some issue gold coinage – the Canadian Maple Leaf and the South African Krugerrand are current examples. Gold (and other precious metal)

prices are quoted in dollars per troy ounce for 900 fine metal.

For many years the 'official' value of gold was set by the USA at $20.67 per troy ounce. The St Gaudens Double Eagle, a $20 US coin, was minted from 1907–1933 and weighed 34 g. In 1934 the value of gold in the USA was increased to $35.00 per troy ounce, to build up reserves and stabilize the dollar after the Great Depression. Although this had a stabilizing effect on the currency markets, it gradually devalued gold over the next 34 years as inflation caused other commodity prices to rise slowly. American citizens were also prohibited from owning gold bullion, whereas foreign parties could purchase gold from the US government at the fixed price.

In 1968, partly because of the growing gap between the 'real' value of gold and the fixed price, a two-tiered pricing system was briefly established whereby gold was still transferred among governments at the 'official' price of $35.00 per troy ounce while the price on the private market was allowed to fluctuate. This tiered price system soon failed (mainly because of South African government sales at the higher market price) and in 1971 the US government finally abandoned the gold monetary standard, allowing gold prices to float freely. In 1974 American citizens were once again allowed to own gold bullion.

With the global oil crises of the 1970s, oil and gold prices spiralled upwards dramatically, the latter peaking at about $850 per troy ounce in 1980. Americans were now free to invest in gold bullion and gold futures markets, which expanded dramatically. The increased prices and influx of investment capital stimulated a new wave of global exploration for gold deposits. It was during this period that South Africa's century-long domination of global gold production from the Wits deposits began to wane. Many gold districts in the USA, Australia, and other nations were discovered or expanded, their financial bottom lines aided by cheap cyanide heap leaching technology.

Increases in the price of gold in the 1980s prompted a wave of gold investment scams, with crooks setting up temporary field operations and fancy offices that were designed to lure investors into buying 'future' gold ore in lots of $5000–$15 000. Enticements of free food, alcohol, and female companionship were often used in such 'investment seminars'. Apparently some investors were hesitant to report losses from such scams because they were using money that they did not want tax officials or former spouses to know about.

Being highly influenced by emotion, greed, international conflict, and economic uncertainty (rather than a more rational basis of technological usefulness), trends in gold price have never lent themselves to accurate prediction. Recent years have seen downward pressure on prices as some western governments sell off some of their central banks' gold reserves, while continued fears of economic and political conflicts in Asia, Europe, and the Middle East have pushed prices upwards. The net result is, as always, erratic and unpredictable trends.

More than any other Earth resource, the concept of 'caveat emptor' applies to gold in all of its manifestations. The most successful investors in gold are probably those who buy stock in reputable gold-mining companies with adequate reserves and low overheads.

See Also

Atmosphere Evolution. **Economic Geology**. **Fluid Inclusions**. **Minerals:** Native Elements; Sulphides. **Mining Geology:** Hydrothermal Ores. **Tectonics:** Hydrothermal Activity.

Further Reading

Amey EB (2002) Gold. In: *US Geological Survey Minerals Year book 2002*, vol. 1, pp. 1–6. US Geological Survey, Washington, DC: Govt. Printing Office.

Boyle RW (1987) *Gold: History and Genesis of Deposits*. New York: Van Nostrand Reinhold.

Craig JR, Vaughan DJ, and Skinner BJ (2001) *Resources of the Earth: Origin, Use, and Environmental Impact*, 3rd edn. Prentice Hall, Upper Saddle River.

Foster RP (ed.) (1991) *Gold Metallogeny and Exploration*. Glasgow: Blackie.

Gasparrini C (1993) *Gold and Other Precious Metals: From Ore to Market*. New York: Springer-Verlag.

Hagemann SG and Brown PE (eds.) (2000) *Gold in 2000*. Reviews in Economic Geology 13. Littleton: Society of Economic Geologists.

Kesler SE (1994) *Mineral Resources, Economics and the Environment*. New York: MacMillan.

Kirkemo H, Newman WL, and Ashley RP (2000) *Gold*. Factsheet. US Geological Survey.

Yannopoulos JC (1991) *The Extractive Metallurgy of Gold*. New York: Van Nostrand Reinhold.

GONDWANALAND AND GONDWANA

J J Veevers, Macquarie University, Sydney, NSW, Australia

Introduction

Gondwanaland existed from the 600–500 Ma accretion of the African and South American terranes to Antarctica–Australia–India, through the 320 Ma merging with Laurussia to form Pangaea, until breakup between 180 Ma and 100 Ma (Figures 1 and 2).

The name Gondwanaland was introduced in 1885 by Eduard Suess for the regions with the *Glossopteris* flora, in particular the Gondwana System of peninsular India. In 1912 Alfred Wegener interpreted Gondwanaland as the supercontinent with Gondwanan floras. Later there was a view that Gondwanaland should be replaced by Gondwana, which, interpreted as 'Land of the Gonds', included one 'land' already. Semantic confusion of supercontinent and stratigraphical system was averted when 'wana' was found to stand for 'forest', so that Gondwanaland means the supercontinent and Gondwana means the Indian kingdom of forest dwellers. The distinction is valuable because the Gondwana facies started only after Gondwanaland merged with Laurussia to form Pangaea. However, not all workers use this terminology.

This account of Gondwanaland is told through a set of maps that stretch from assembly, through the merger with Laurussia to form Pangaea, to breakup.

Early–Middle Cambrian (530–500 Ma)

In the Early–Middle Cambrian, Gondwanaland was bounded to the north (at the modern coordinates of Africa) by terranes now in Laurentia, Europe, and Asia, and to the west and south by a trench (Figures 3 and 4). The interior was crossed by fold belts generated during the terminal Pan-Gondwanaland (600–500 Ma) deformation, which endowed Gondwanaland with a thick buoyant crust and lithosphere, and nonmarine siliciclastic deposits behind a peripheral shoreline.

Between 650 Ma and 570 Ma, stress 1 (Figure 3) was generated from the oblique collisions of, first, Avalonia–Cadomia with the West African Craton and, second, West Gondwanaland with East Gondwanaland during the closure of the Mozambique Ocean. The West African Craton, which was rotated counterclockwise, imparted clockwise rotation to the Amazonia Craton, which, in turn, rotated the Congo Craton counterclockwise. Between 550 Ma and 490 Ma, stress 2 was generated first by oblique subduction of the Palaeo-Pacific Plate beneath Antarctica and second by transcurrence beneath Avalonia–Cadomia and the West African and Amazonia Cratons. The India–East Antarctica–West Australia Craton was driven into counterclockwise rotation, which imparted clockwise rotation to the North Australia Craton, modelled by dextral shear along small circles about a pole in the Pacific. A fold belt was extruded between the West African and Congo Cratons, and another between the Congo and Amazonia Cratons. The cycle ended at about 500 Ma with final convergence along the Palaeo-Pacific margin and uplift and cooling in Gondwanaland. The heat emitted during convergence, added to that generated during the Pan-Gondwanaland cycle, built buoyancy into Gondwanaland by underplating the lower crust with mafic magma to promote isostatic uplift and concomitant downwearing.

The shoreline alongside Antarctica, through Australia and north-west India, continued along the northern margin to the Levant and north-east Africa, with marine sediment deposited in belts of terrigenous, mixed, and carbonate facies across West Africa and southwards (in modern coordinates) into South America (past the north pole) and then across the Damara fold belt, to link with carbonate in the Transantarctic Mountains, and over flood basalt on the Australian platform. In the south-east, a newly generated marginal basin started to close by north-eastward-directed subduction beneath a volcanic arc.

Above subducting slabs, the margin includes granite in the Suwannee terrane of Florida, granite in Argentina and beside a rift projected towards the Transantarctic Mountains, the string of Ross plutons, the Delamerian granites, and granites in north-eastern Australia. The Prydz-Leeuwin Belt and Mozambique Orogenic Belt were metamorphosed.

Early Ordovician (490–458 Ma)

In the Early Ordovician the north pole lay in the Sahara (Figure 5), but ice did not appear until 444 Ma. Australia was crossed by the (Larapintine) sea behind a magmatic arc generated by westwards-directed subduction of ocean floor that was flooded by fans of quartzose sediment from Antarctica.

Areas of Pan-Gondwanaland deformation had cooled. The shoreline lapped the Beardmore Shelf and Table Mountain Shelf. In the ancestral Paraná

Basin, north-east-trending rift basins filled with nonmarine detritus, including rhyolitic volcaniclastics. Beach sand was deposited in Florida, and carbonate was deposited around the ancestral Sierra Pampeanas, which were intruded by granite.

The Avalonian terranes drifted away from Africa at 470 Ma in the first of many transfers of material from Gondwanaland to the 'northern' continents.

Late Ordovician (458–443 Ma)

The end-Ordovician glaciation affected Saharan Africa and southern Africa–South America (Figure 6). Glacial advances and retreats climaxed with a big advance at the end of the Ordovician (444–443 Ma). In the Sahara, directional structures indicate uplands including Sudan and Arabia. Interglacial marine incursions swept over nonmarine glacigenic landforms and sediment. Distal glacimarine deposits extend over much of the northern margin of Africa–Arabia, including the marginal terranes. Other ice centres on uplands lay 3000 km away along the Pacific margin, suggesting that proximity of uplands to the sea was a factor. Nonmarine glacigenic deposits are found in Bolivia and Venezuela.

The interglacial sea received marine glacigenic sediment in the Volta and Bowé Basins, Iberia and Cadomia, North Africa, southern Turkey, and Arabia, which were all fringed by nonmarine basins and uplands and scattered anorogenic igneous complexes. The shoreline crossed northern India and continued alongside Western Australia and the North China Shelf. The intermittent Larapintine Seaway produced thick halite evaporites in the west. In the east, the sea opened onto a convergent margin and magmatic arc. The shoreline passed the Beardmore Shelf to the embayments of the Table Mountain Shelf and Don Braulio area.

Early Silurian (443 Ma)

At the beginning of the Silurian, the ice-sheets contracted to a strip in North Africa–Arabia and advanced over Brazil. In transgressing North Africa, the post-glacial sea accumulated glacigenic sediment, including 'hot shales' (which are sources of petroleum) with graptolites in the Fort Polignac Basin, shed from surrounding uplifts, including the Sudan upland, which was intruded by anorogenic complexes (Figure 7). Uplifts in South America likewise accumulated ice. Glacimarine diamictites and shales were deposited in the Paraná Basin, including the Iapó Formation, and in the Parnaíba Basin. Nonmarine diamictites were deposited in the eastern Paraná Basin and the Jatobá Basin. The marine diamictite of the Amazonas Basin crops out in a 1500 km long belt in the eastern Andes. Farther south, the uppermost Don Braulio Formation is not glacigenic. The Malvinokaffric (zoogeographical) Realm occupied South America, southern Africa, and Antarctica before it disappeared in the Middle Devonian.

In Antarctica, the nonmarine Crashsite Group in the Ellsworth Mountains lay behind a zone of deformation and metamorphism that extended through Marie Byrd Land and New Zealand into Eastern and Central Australia, including the Melbourne Terrane and Benambran Highlands (BH). Evaporites were deposited in the Bonaparte Basin.

Early Devonian (418–394 Ma)

In the Early Devonian the south pole lay off Patagonia but glacigenic deposits are unknown. Laurentia was about to make contact (Figure 8). Shelves in Arabia and North Africa widened during a marine transgression from the Lochkovian (415 Ma) to the Emsian (400 Ma) that formed a wide embayment with its head at Accra. Carbonate sediment indicates warm to moderate water, consistent with the tropical latitude. The Sudan upland, which was peppered with ring complexes, shed sand into nonmarine basins, including the Kufra Basin, which lay inland of the marine Ghazalat Basin. Past Arabia, the shoreline continued through the Chitral area of Pakistan and northern India to north-west Australia.

The Emsian shoreline passed the Lolén Formation and Bokkeveld Group. The rest of South America was crossed by the arms of a shallow sea between large islands that faced the Pacific margin on the west. Granite was emplaced in the Sierra Pampeanas.

The shoreline passed the Ellsworth Mountains and Ohio Range and was backed by nonmarine sandstone in the Pensacola Mountains, Beardmore Glacier area, and South Victoria Land. In New Zealand, marine sediment was deposited on terranes that amalgamated at 415 Ma. Amalgamation of north-east and west Tasmania followed at 400 Ma. In south-east Australia, rivers deposited sediment in the wide back-arc region of granite intrusion behind a volcanic arc and subduction complex. Beyond a group of nonmarine deposits in the centre, evaporites were deposited in the Bonaparte Basin, aeolian and playa deposits in the Canning Basin, and nonmarine redbeds in the Carnarvon Basin.

Late Devonian (382.5–362 Ma)

In the Late Devonian (Figure 9), the sea, initially (in the Frasnian) still in North Africa and South America and accumulating black shale, retreated in the Strunian

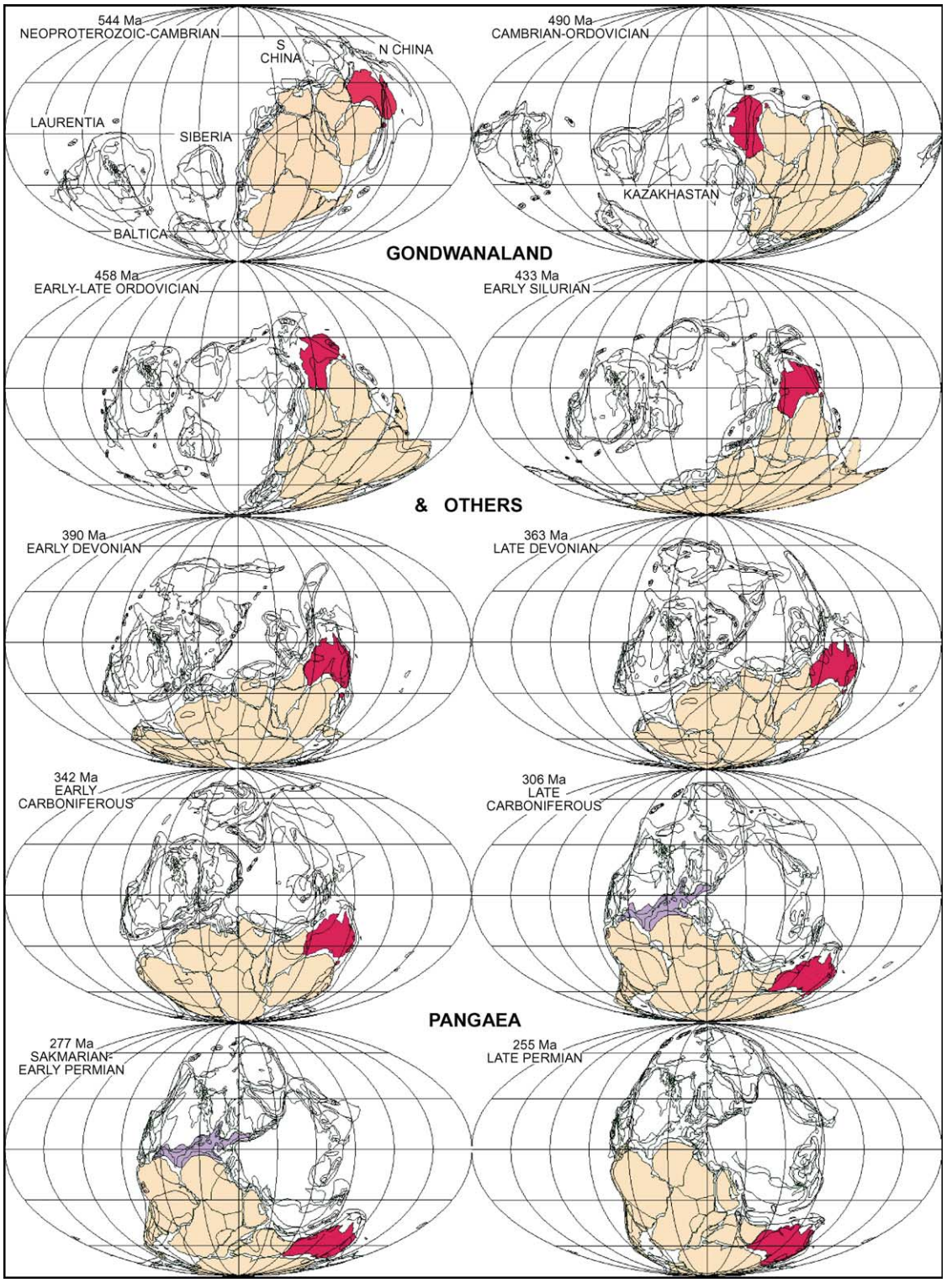


Figure 1 Australia (red), associates in Gondwanaland (brown), and other continents and terranes (Laurentia, Baltica, Siberia, Kazakhastan, Cimmeria, North and South China) through ten stages of the Palaeozoic. Reproduced with permission from Veevers JJ (2001) *Atlas of Billion-Year Earth History of Australia and Neighbours in Gondwanaland*. Sydney: GEMOC Press and C. R. Scotese.

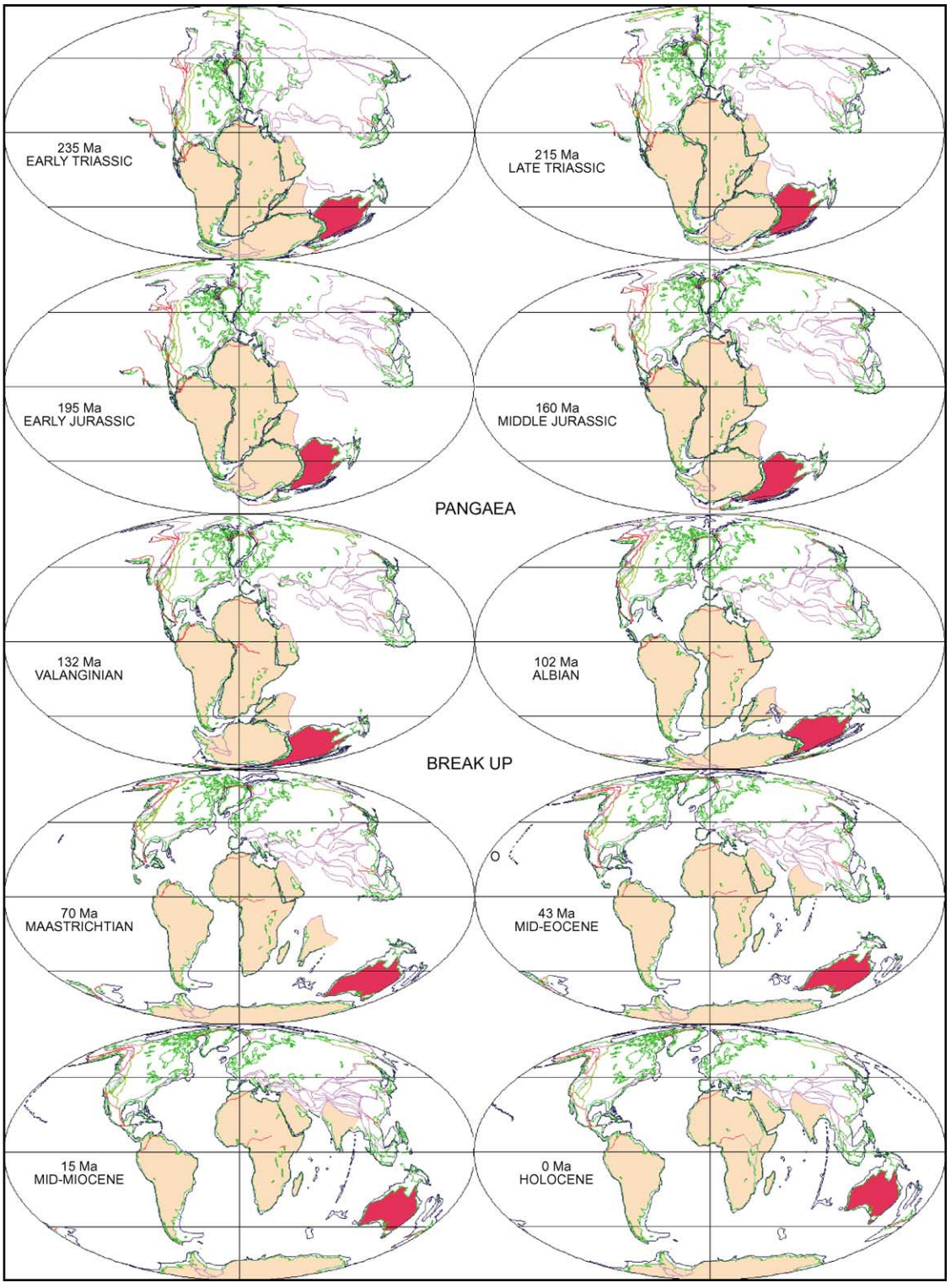


Figure 2 Australia (red), associates in Gondwanaland (brown), and other continents and terranes through ten stages in the Mesozoic and Cenozoic. Reproduced with permission from Veevers JJ (2001) *Atlas of Billion-Year Earth History of Australia and Neighbours in Gondwanaland*. Sydney: GEMOC Press and C. R. Scotese.

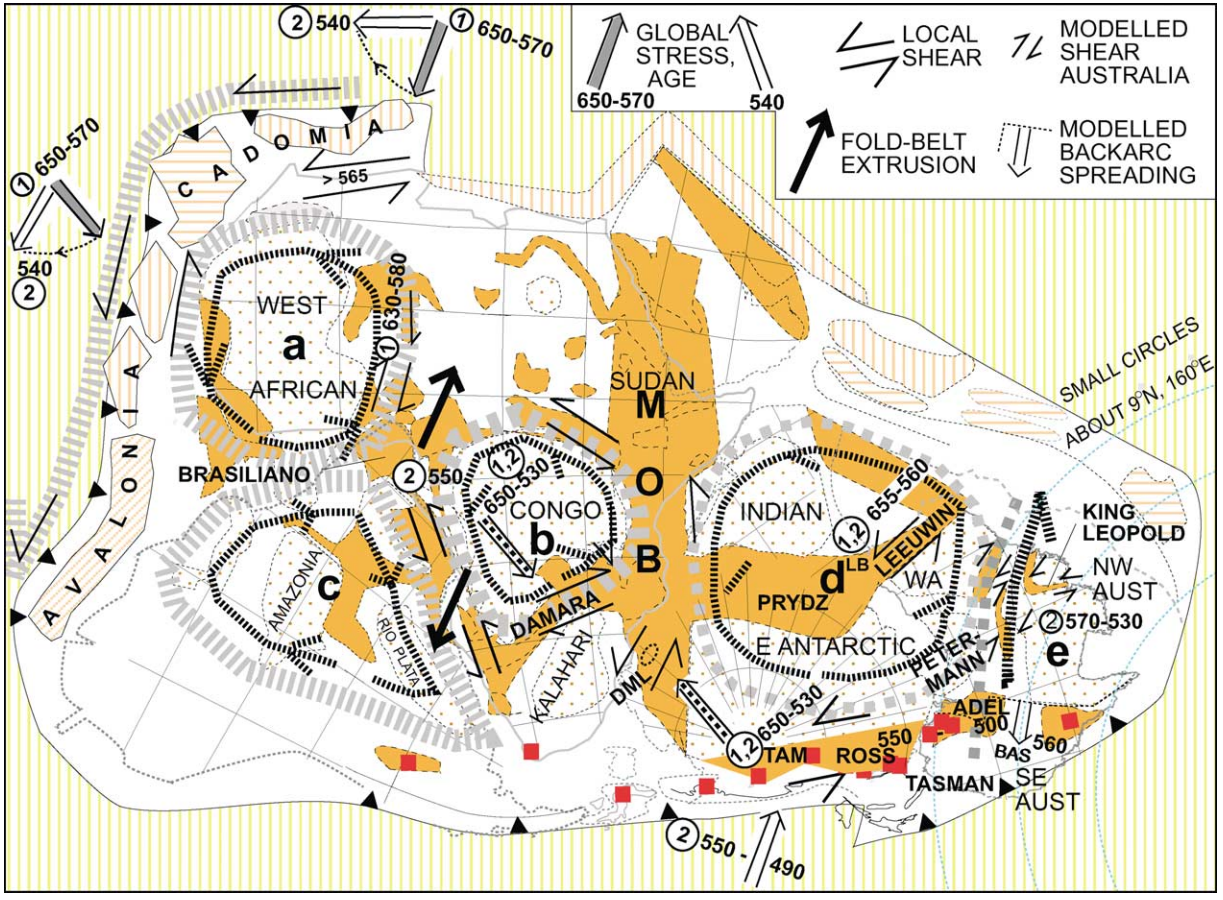


Figure 3 Early Cambrian–Middle Cambrian (530–500 Ma) tectonics of Gondwanaland, showing cratons and sense of rotation (narrow broken lines and arrows) grouped into composites (enclosed by wide broken line): a, West African Craton; b, Congo Craton; c, Amazonia–Rio de la Plata Craton; d, India–East Antarctica–West Australia Craton, bisected by the Prydz-Leeuwin belt; e, North Australia Craton. Pan-Gondwanaland (600–500 Ma) fold belts are orange. Ages are given in Ma. Circled numbers indicate stresses 1 and 2. DML, Dronning Maud Land; LB, Leeuwin block; MOB, Mozambique orogenic belt; TAM, Transantarctic Mountains; WA, Western Australian Craton; BAS, back-arc spreading. Reproduced with permission from Veevers JJ (2003) Pan-African is Pan-Gondwanaland: oblique convergence drives rotation during 650–500 Ma assembly. *Geology* 31: 501–504.

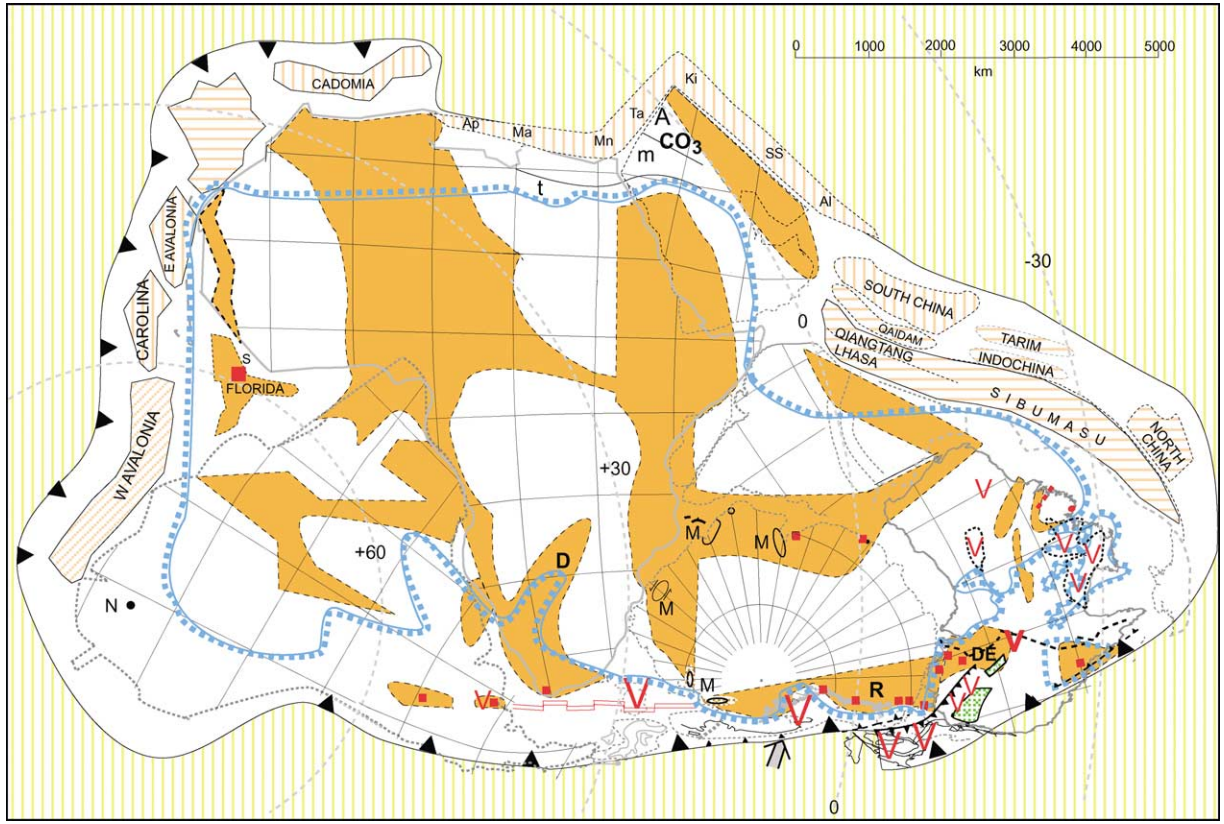


Figure 4 Early Cambrian–Middle Cambrian (530–500 Ma) palaeogeography, with shorelines (blue line–dots away from the land) skirting fold belts (orange). Prospective terranes run from West Avalonia to Cadomia and from Apulia to Sibumasu/North China. From west to east, the terranes on the north include: Ap, Apulia; Ma, Mani; Mn, Menderes; Ta, Taurus; Ki, Kirshehir; SS, Sanandaj-Sirjan; Al, Alborz. Facies indicated as: t, terrigenous; m, mixed; and CO_3, carbonate (Red Vs indicate volcanics, red squares granite). Other abbreviations: D, Damara fold belt; S, Suwannee terrane; DE, Delamerian granites; R, Ross plutons and M, metamorphism. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

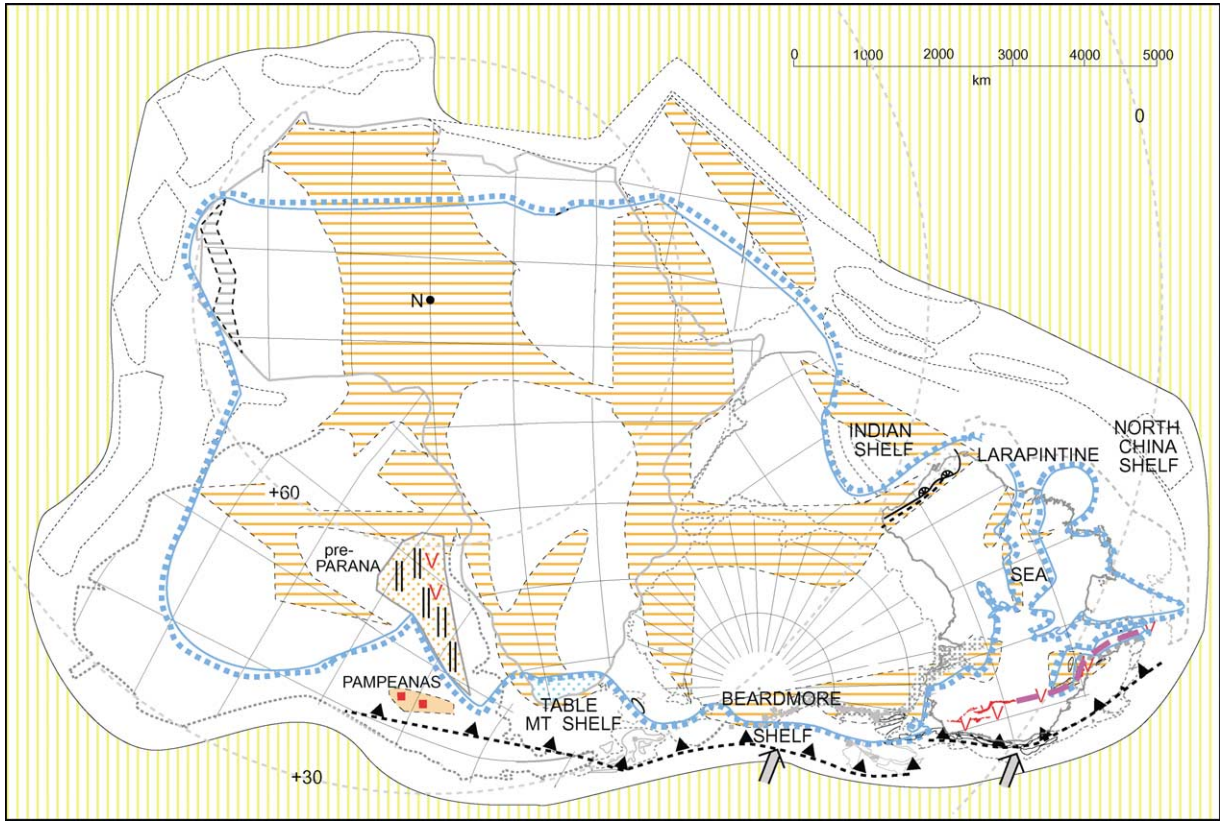


Figure 5 Early Ordovician (490–458 Ma) palaeogeography. N, position of north pole; orange horizontal shading, areas of Pan-Gondwanaland deformation. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

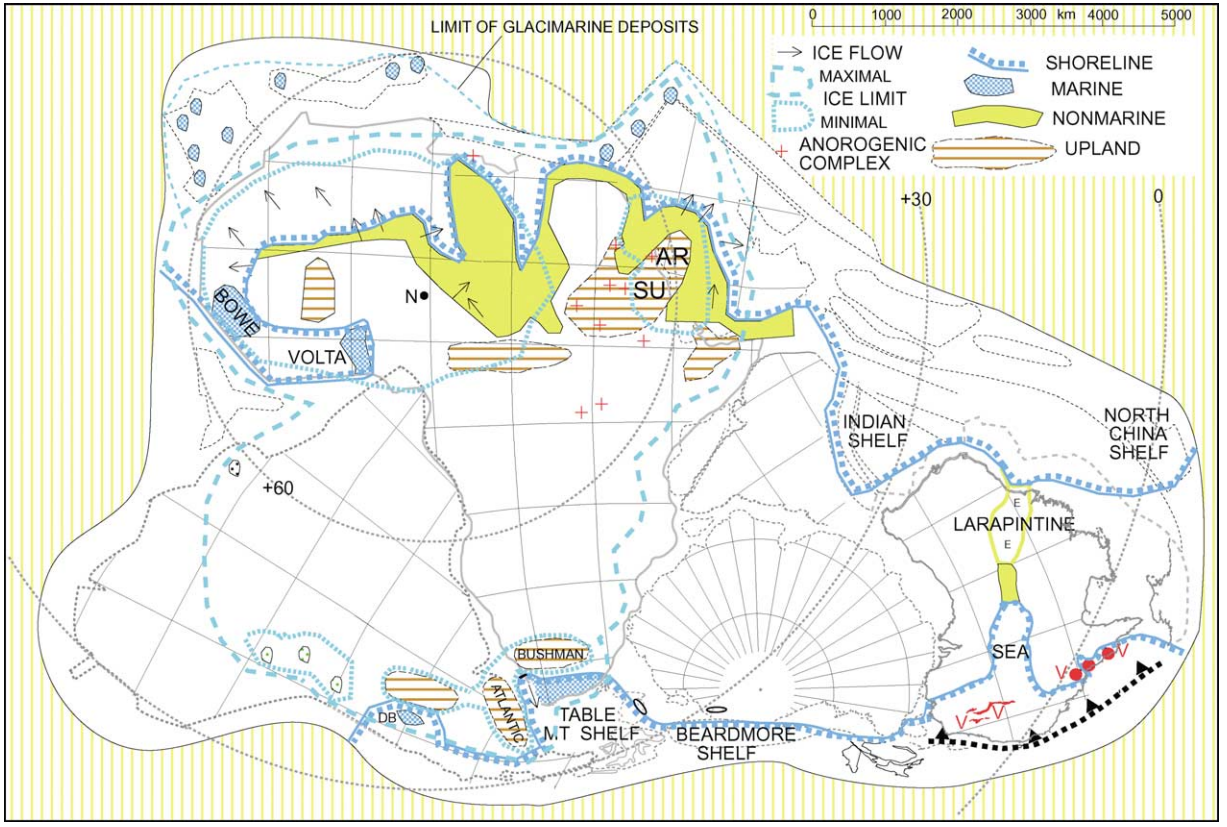


Figure 6 Late Ordovician (458–443 Ma) palaeogeography: DB, the glacimarine Don Braulio Formation; AR, Arabian upland; SU, Sudan upland; E, area of evaporite deposition; += anorogenic complex. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

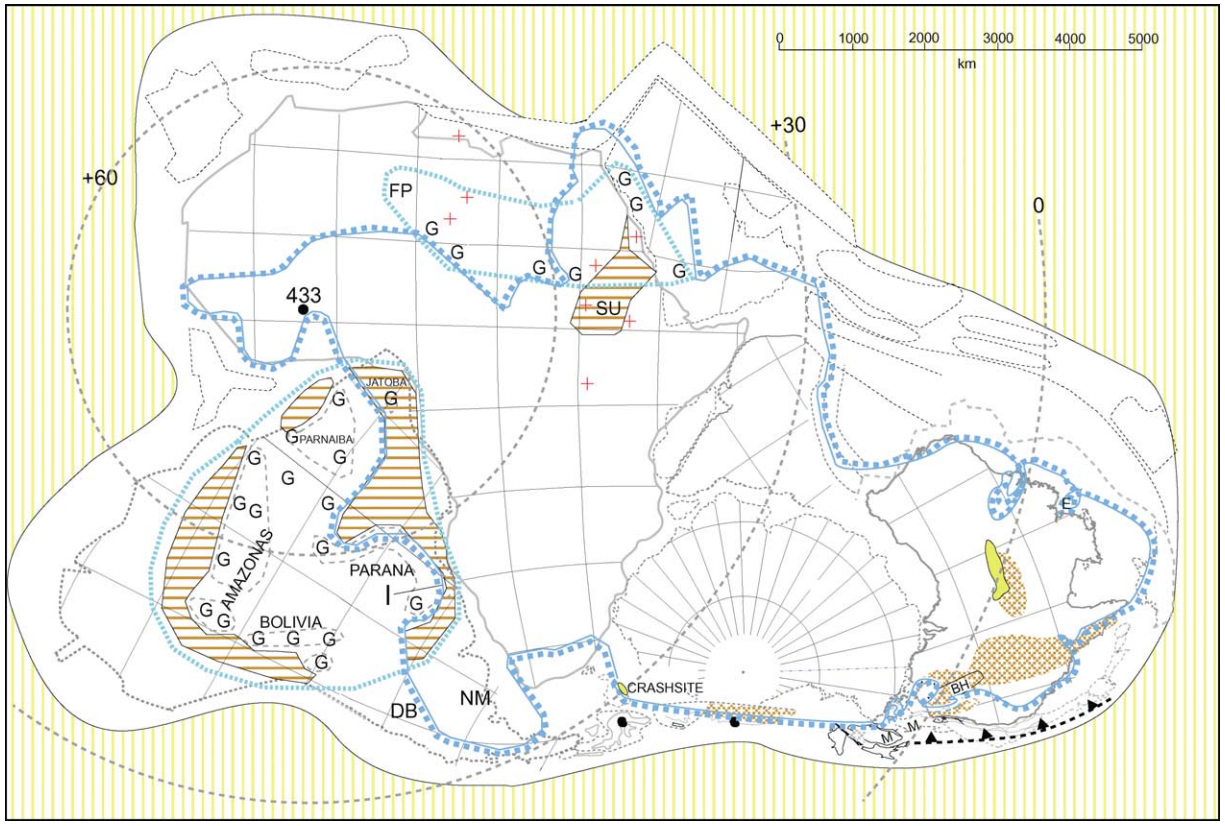


Figure 7 Early Silurian (443 Ma) palaeogeography. Shading as in **Figure 6**, with brown cross-hatching representing deformation and metamorphism. NM, nonmarine sediment in the Sierra de la Ventana of Argentina; G, glacigenic sediment; FP, Fort Polignac Basin; SU, Sudan upland; red crosses, anorogenic complexes; I, Iapó Formation; DB, Don Braulio Formation; M, Melbourne Terrane; BH, Benambran Highlands; E, evaporites. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

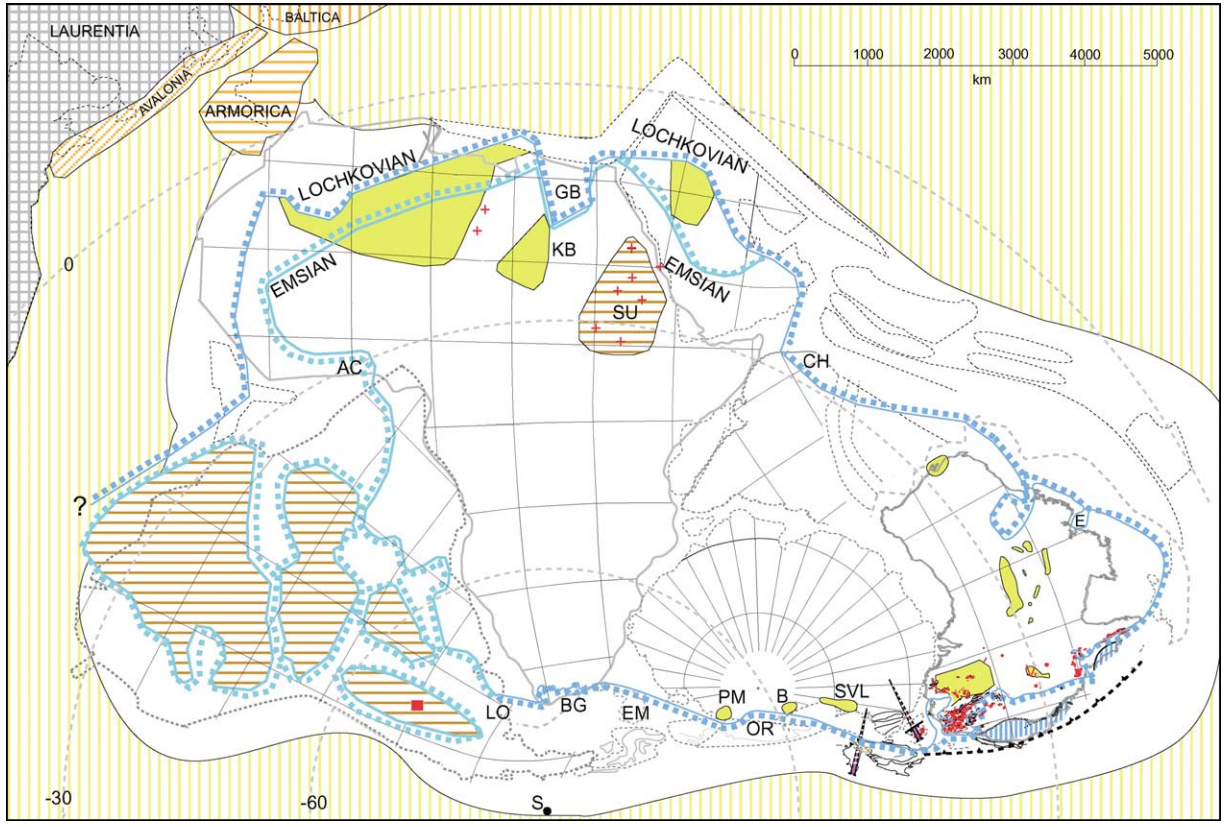


Figure 8 Early Devonian (418–394 Ma) palaeogeography: AC, Accra; SU, Sudan upland; KB, Kufra Basin; GB, Ghazalat Basin; CH, Chitral; LO, Lolén Formation; BG, Bokkeveld Group; EM, Ellsworth Mountains; OR, Ohio Range; PM, Pensacola Mountains; B, Beardmore Glacier; SVL, South Victoria Land; E, evaporites; horizontal orange shading, large islands separated by shallow sea; screw symbol, amalgamation of terranes. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

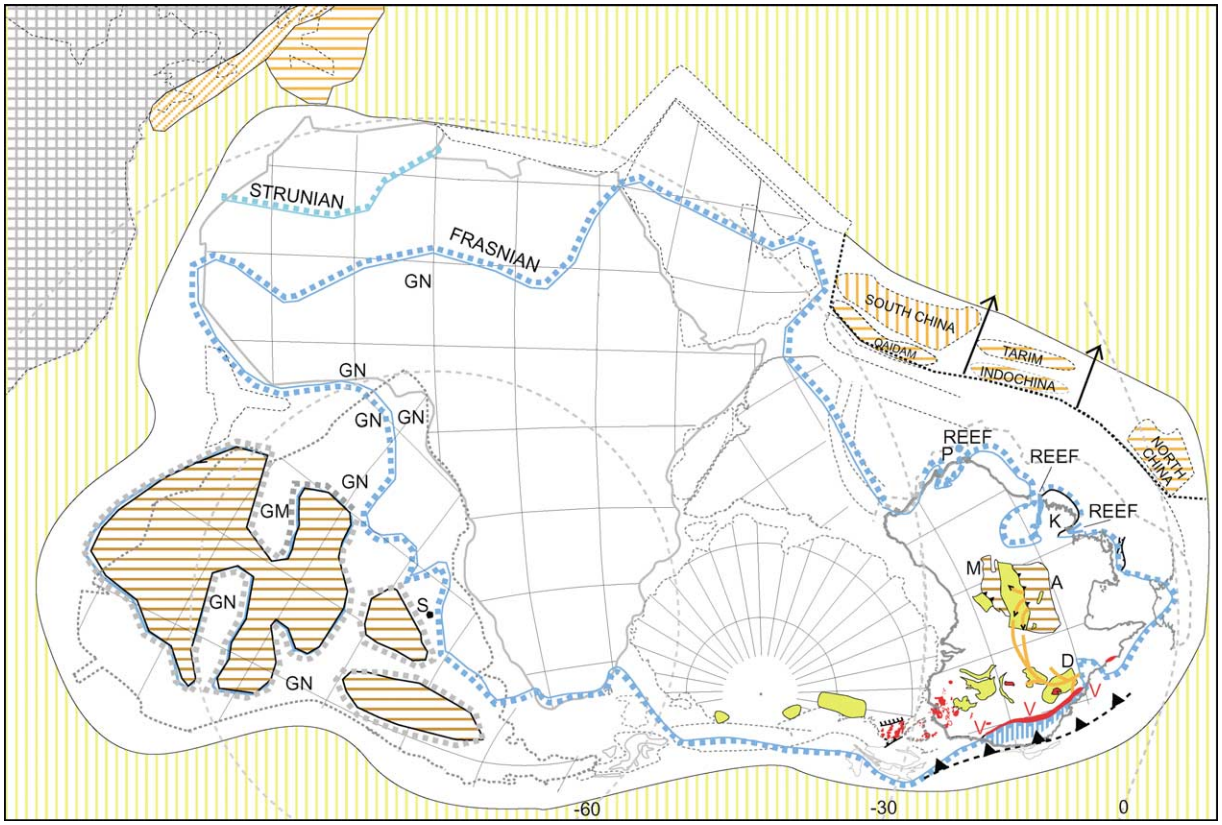


Figure 9 Late Devonian (382.5–362 Ma) palaeogeography: GN, glacigenic nonmarine sediment; GM, glacigenic marine sediment; A, Arunta block; M, Busgrave block; D, Drummond Basin; K, Kimberley block; P, reef facies. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

(latest Devonian). Terranes started to break off in the north-east leading to the generation of Palaeo-Tethys. The Centralian Superbasin, which had been subsiding since 840 Ma, was initially dismembered during crustal shortening, and sediment flowed into the convergent eastern Australia. Laurussia touched Gondwanaland.

In South America, black shale was deposited during the Frasnian, and glacigenic nonmarine sediment (glacimarine sediment in the Amazon Basin) was deposited during the end-Famennian regression. Glacigenic sediment extended into adjacent Africa. In North Africa–Arabia, the depositional facies changed from detrital in the south to carbonate in the north. In southern Africa and Antarctica, the shoreline and nonmarine depocentres remained in the same place as before. During continuing westward-directed subduction, granite was intruded in North Victoria Land and northwards, where volcanic rifts opened after the contractional Tabberabberan Orogeny. Farther north a magmatic arc and subduction complex developed behind the trench.

In central Australia, the overthrusting Arunta block disrupted the wider Amadeus Basin and culminated in 5 km of uplift reflected in extensive alluvial fans. Similar sediments were shed southwards from the overthrusting Musgrave block. In the Amadeus Basin, eastward flow joined the northward flow in the Drummond Basin to debouch on the margin. In the Bonaparte Basin, alluvial fans were succeeded by a reef complex. Likewise in the Canning Basin, gravel fans from the Kimberley block mingled with a reef complex on a platform in front of a deep axis; in the south, paralic sediment was deposited in an arm of the sea. In the Carnarvon Basin, shelf limestone gave way to reef facies.

Early Carboniferous, Visean (335 Ma)

In the Early Carboniferous, the Moroccan salient of Gondwanaland collided with the Armorican salient of Laurussia. Palaeo-Tethys further separated the Chinese blocks (Figure 10).

In North Africa–Arabia, the shoreline, bounded by nonmarine sediment, penetrated almost to Nigeria. In the east, uplifts include the Sudan Arch, the Central Arabian Arch, and the Summan Platform; igneous centres include rhyolite.

In South America, an Early Carboniferous contraction with arc magmatism (not shown) led to deformation and uplift that drove out the sea. Floodplain deposits with coal are found in the Lake Titicaca area. Glacigenic nonmarine sediment was deposited in the Pimenta Bueno, Jaurú, Solimões, central Amazonas, and Parnaíba–Itacaja areas. The pole was in north-east Africa, so the tillitic beds in North Africa were within 30° of the pole, but the glacigenic localities of South America, between 25° and 60° from the pole, all nonmarine, reflect alpine glaciation.

South-eastern Australia was subjected to the Kanimblan east–west contraction, followed by widening of the volcanic arc and finally north–south contraction in megakinks. Similar events took place in northern Queensland. In between, nonmarine deposition in the Drummond Basin continued until 330 Ma, when it was terminated by gentle folding, part of the Alice Springs terminal folding and thrusting that dismembered the Centralian Superbasin, a distant effect of the collision in north-west Africa. The youngest preserved sediment in the foreland basins of Central Australia is the 330 Ma Mount Eclipse Sandstone.

Mid-Carboniferous, Namurian (327–311.5 Ma)

In the Namurian, Laurussia and Gondwanaland merged to form Pangaea by definitive right-lateral contact along the Variscan suture, and the collisional stress and subsequent uplift was felt as far afield as Australia (Figure 11). The south pole had moved 60° since 335 Ma to a location in Marie Byrd Land, and ice sheets developed on the tectonic uplands south of 25° S in South America and south of 45° S elsewhere. The collisional uplands of equatorial north-west Africa and adjacent Europe were subjected to intense rainfall and backwearing.

Rapid uplift and concomitant downwearing must have produced copious sediment, yet the depositional record over the Australian platform is blank. Where could the sediment have gone? The paradox can be resolved by postulating that uplift combined with rapid polar movement triggered a continent-wide glaciation so that sediment shed from the nunataks of the central uplifts and from the eastern cordillera was carried away in the ice-sheet. Only in the east was nonmarine glacial sediment deposited by glaciers that broke through the eastern cordillera. During its retreat in the earliest Permian the ice released its load of sediment. Ice-sheets continued across Antarctica through the Ross and Gamburtsev areas, and into adjacent India and southern Africa, through uplands called Windhoek, Cargonian, and proto-fold belt. A separate ice-sheet covered uplands in South America, which shed glacimarine sediment in front of a volcanic arc.

Now part of Pangaea, Gondwanaland underwent a (Pangaean) cycle of tectonic and climatic events.

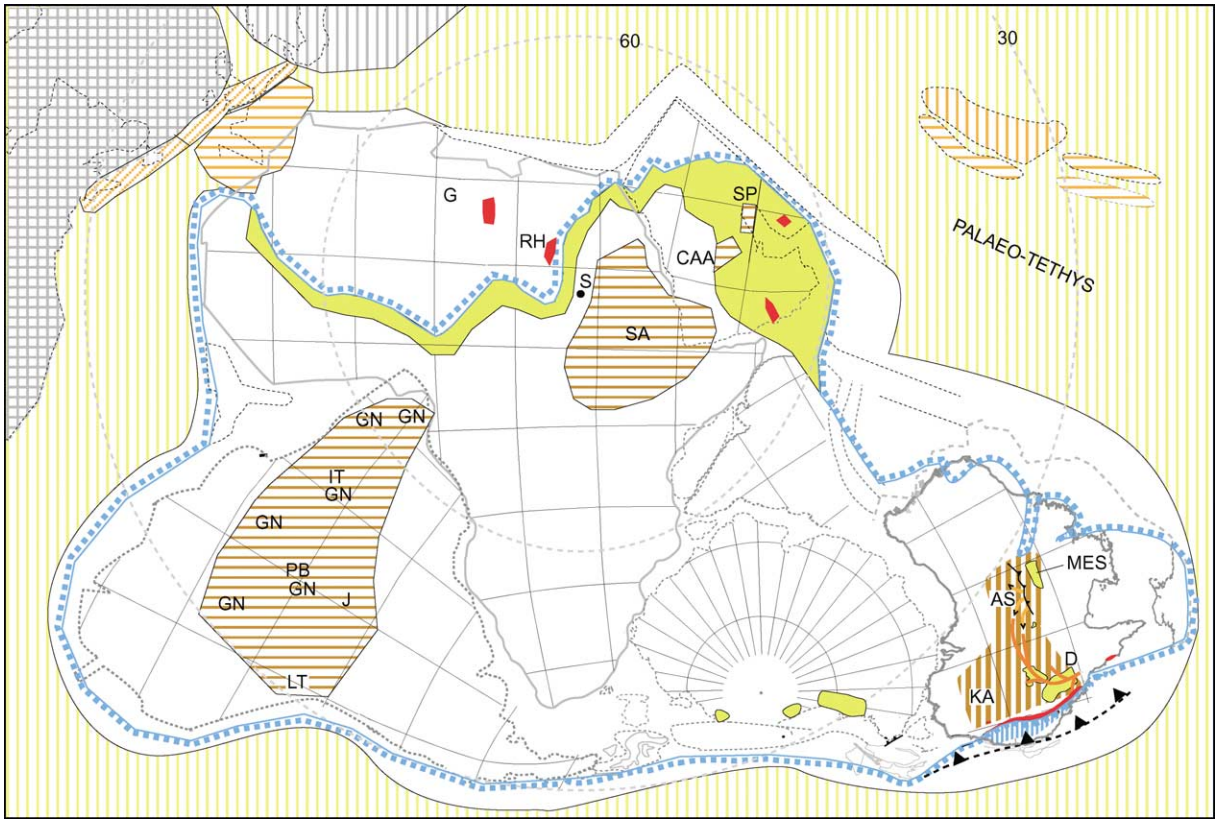


Figure 10 Early Carboniferous (335 Ma) palaeogeography: SA, Sudan Arch; CAA, Central Arabian Arch; SP, Summan Platform; RH, rhyolite; LT, Lake Titicaca; GN, glacigenic nonmarine sediment; PB, Pimenta Bueno; J, Jaurú; IT, Itacaja; G, tillitic beds; KA, Kanimblan contraction; D, Drummond Basin; AS, Alice Springs; MES, Mount Eclipse Sandstone; red shading, igneous centres; orange = upland; lime-green = nonmarine fringe behind shoreline. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

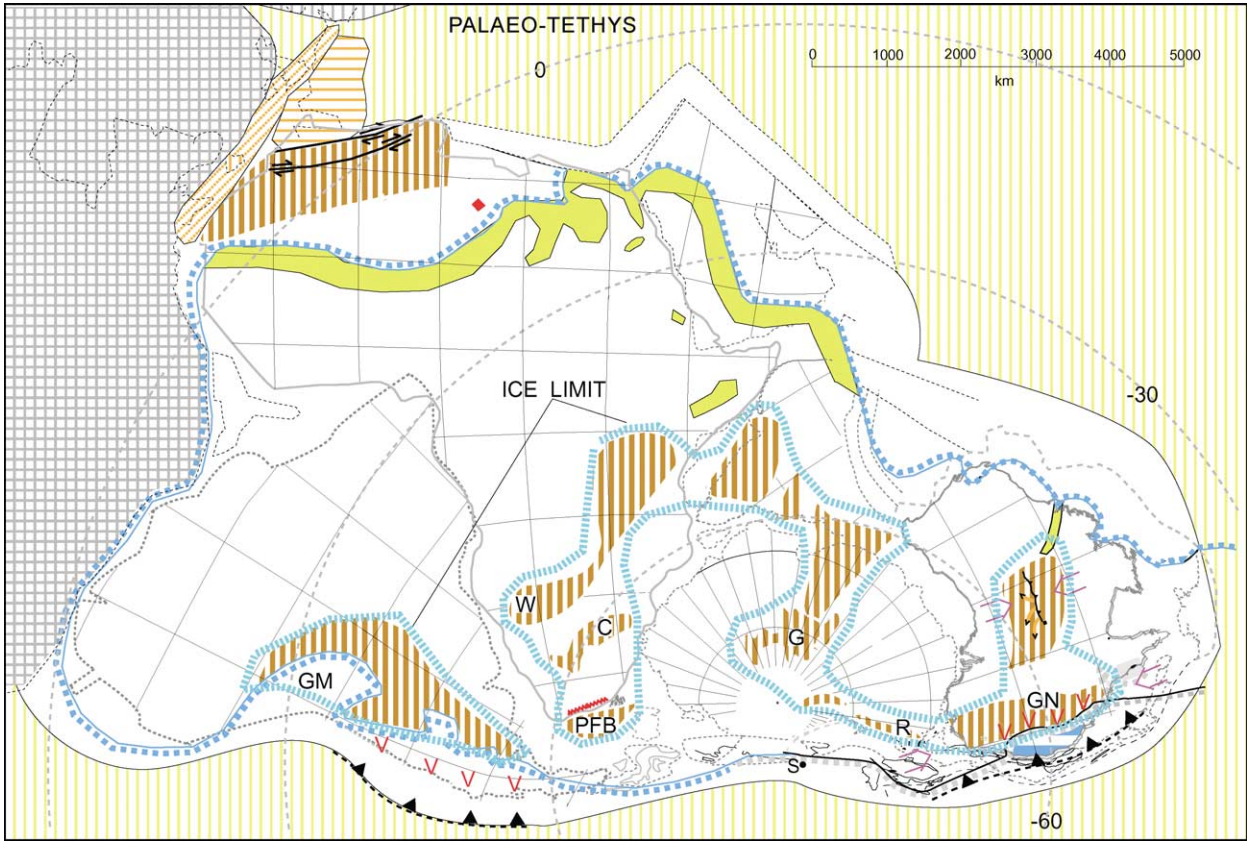


Figure 11 Mid-Carboniferous, Namurian (327–311.5 Ma), palaeogeography. Pink arrows in central and eastern Australia indicate farfield stress. GN, glacigenic nonmarine sediment; R, Ross area; G, Gamburtsev area; W, Windhoek upland; C, Cargonian upland; PFB, proto-fold belt; GM; glacimarine sediment; vertical orange shading, collisional stress and uplift. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

Tectonic–Climatic Model of the Pangaea-to-Dispersed-Continents Cycle

The heightened radioactivity of continental rocks led to the idea that a system of ascending convection ('monsoonal') currents beneath an insulating supercontinent would spread out at the top and eventually break the supercontinent into individual continents. Changes in the pattern and vigour of mantle convection would increase atmospheric carbon dioxide, creating a greenhouse effect. Heat was released in five stages (Figure 12).

- Stage 1 (platform lacuna). The Earth comprises the single continent (Pangaea) and ocean (Panthalassa). The amounts of spreading and subduction are minimal. Minimal turnover of mantle material leads to minimal venting of carbon dioxide and an icehouse climatic state. Maximal continental freeboard is due to, first, the short and narrow mid-ocean ridge displacing less water, second, the continental crust having a maximal mean thickness because its ocean frontage of thin (rifted) crust is minimal, and, third, the self-induced Pangaean heat store and accelerated mantle plumes generating a geoid high. The Earth is dominated by dry land.
- Stage 2. The heat impounded beneath Pangaea soon leads to localized thinning of the Pangaean crust and lithosphere, initially by sagging of cratonic basement and rifting of orogenic basement (extension I).
- Stage 3. Continued crustal thinning leads to rifting between the incipient continents (extension II).
- Stage 4. Pangaea breaks up by spreading of intra-Pangaean rift oceans to form dispersed continents and oceans. The mid-ocean-ridge spreading and subduction are maximal, leading to maximal carbon dioxide venting and greenhouse conditions. Low continental freeboard arises from, first, the mid-ocean ridges displacing more water, second, the crust of the continents having a minimal average thickness because its ocean frontage of thin crust is maximal, and, third, the rapidly depleting Pangaean heat store and decelerated mantle plumes supporting a lower geoid. The Earth is dominated by ocean.
- Stage 5. The depleted heat store leads to slower spreading and subduction and preferential closing of the rift oceans, so that eventually the continents reform Pangaea and the oceans reform Panthalassa in a return to Stage 1.

Earliest Permian (302–280 Ma)

Following a lacuna (Pangaean stage 1), the heat beneath Pangaea drove differential subsidence of the Gondwanaland platform (stage 2, extension I) to trap sediment released from the ice (Figure 13). Terranes left the northern margin, and rift zones penetrated East Africa and between India and Australia. Granite moved into transtensional rifts, driven by the right-lateral shear between Laurussia and Gondwanaland. Coal with the *Glossopteris* flora succeeds glacigenic sediment in all parts of Gondwanaland except tropical South America and Africa. A magmatic and orogenic zone along the Panthalassan margin (Alexander Du Toit's Samfrau Geosyncline) developed in South America.

The Sakmarian (288 Ma) postglacial shoreline ran along the margin in North Africa and made broad indentations across Arabia and India (the India–Australia Rift Zone) and narrow indentations across north-west Australia. It lapped the magmatic zone of eastern Australia and enclosed a gulf between Australia and Antarctica and another between Antarctica and South America.

The Gamburtsev upland shed glacial sediment into southern Africa, India, and south-west Australia, and the Beardmore-Ross upland continued to shed sediment into south-east Australia and New Zealand. Ice extended across the Congo Basin into North Africa and Arabia, with outwash material along the margin.

The inception of Gondwanan glaciation is linked to the lowered input of carbon dioxide in Pangaean stage 1, the removal of atmospheric carbon dioxide during accelerated erosion and weathering of the uplands, and the blocked oceanic circulation at the equator.

Glossopteris of the Gondwana palaeobotanical province made voluminous coals during the entire Permian (Figure 14). The Eurameria province extended into northern South America and Africa, and all three provinces – Eurameria, Gondwana, and Cathaysia – were juxtaposed in Arabia. The boundary between the Eurameria and Gondwana provinces approximates the ice limit (Figure 13). The Variscides also contain coal but without *Glossopteris*.

Late Permian (255–250 Ma)

In the Late Permian the sea retreated to the margins except in the north-east (Figure 15). The magmatic orogen (Gondwanides I) had propagated past northern Queensland. It was backed by the Karoo foreland basin, which was occupied by a vast lake with endemic nonmarine bivalves (preserved in the Waterford Formation, which extends into the Estrada Nova Formation of South America). Smaller lakes crossed the rest of southern and central Africa.

In eastern Australia, volcanogenic sediment flowed across a foreland basin and forebulge into

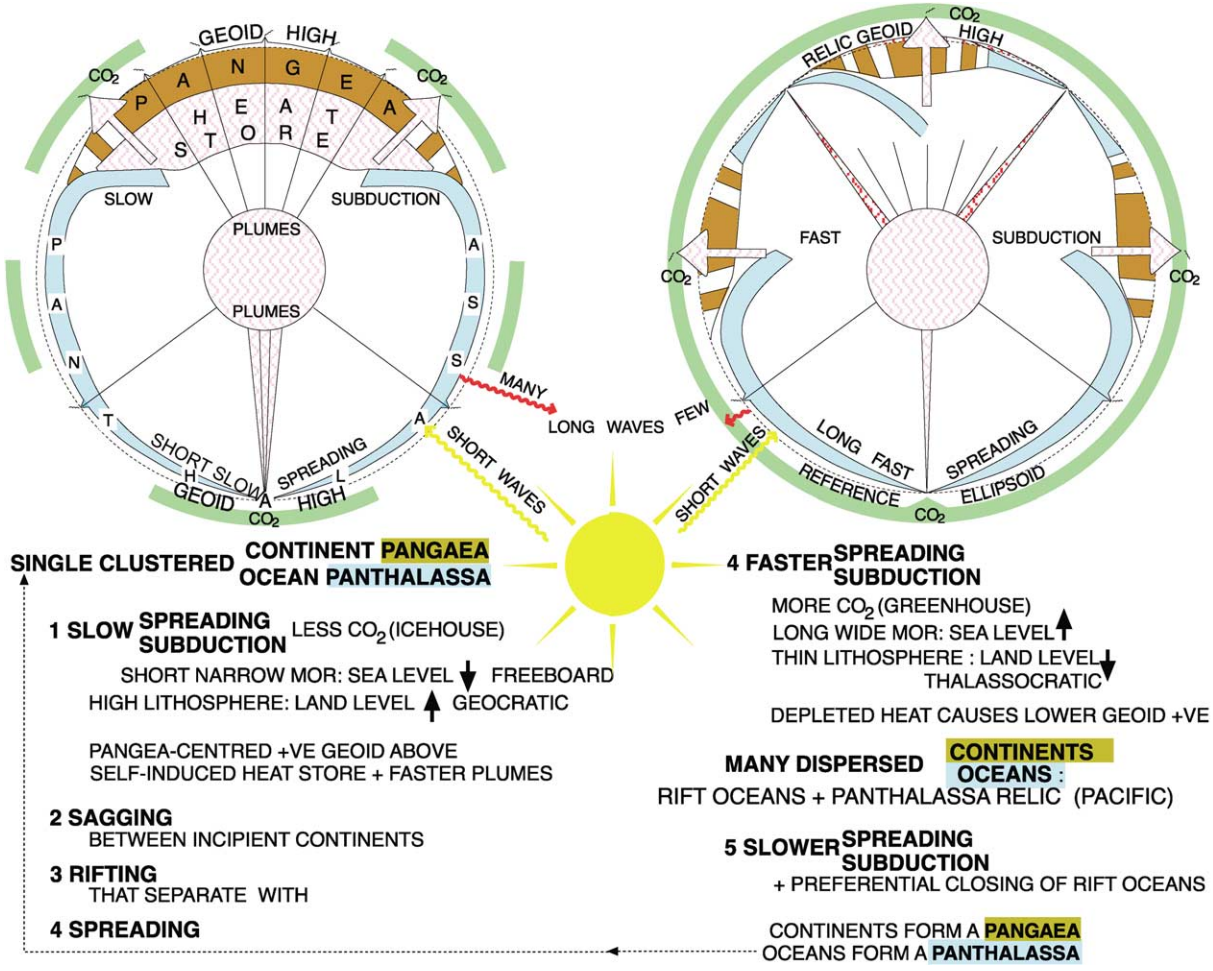


Figure 12 Diagrammatic model of the single continent (Pangaea) and ocean (Panthalassa) alternating with dispersed continents and oceans through five stages. Reproduced with permission from Veevers JJ (2001) *Atlas of Billion-Year Earth History of Australia and Neighbours in Gondwanaland*. Sydney: GEMOC Press.

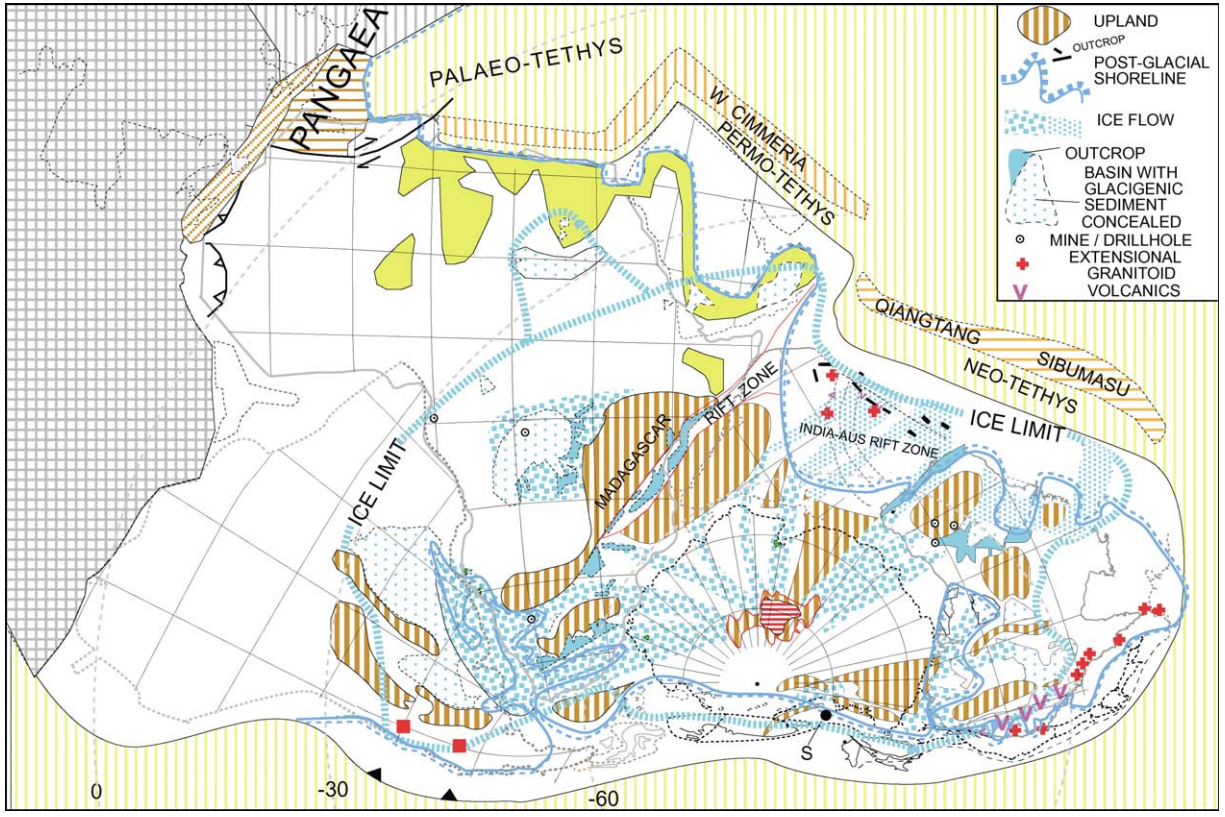


Figure 13 Earliest Permian (302–280 Ma) palaeogeography. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

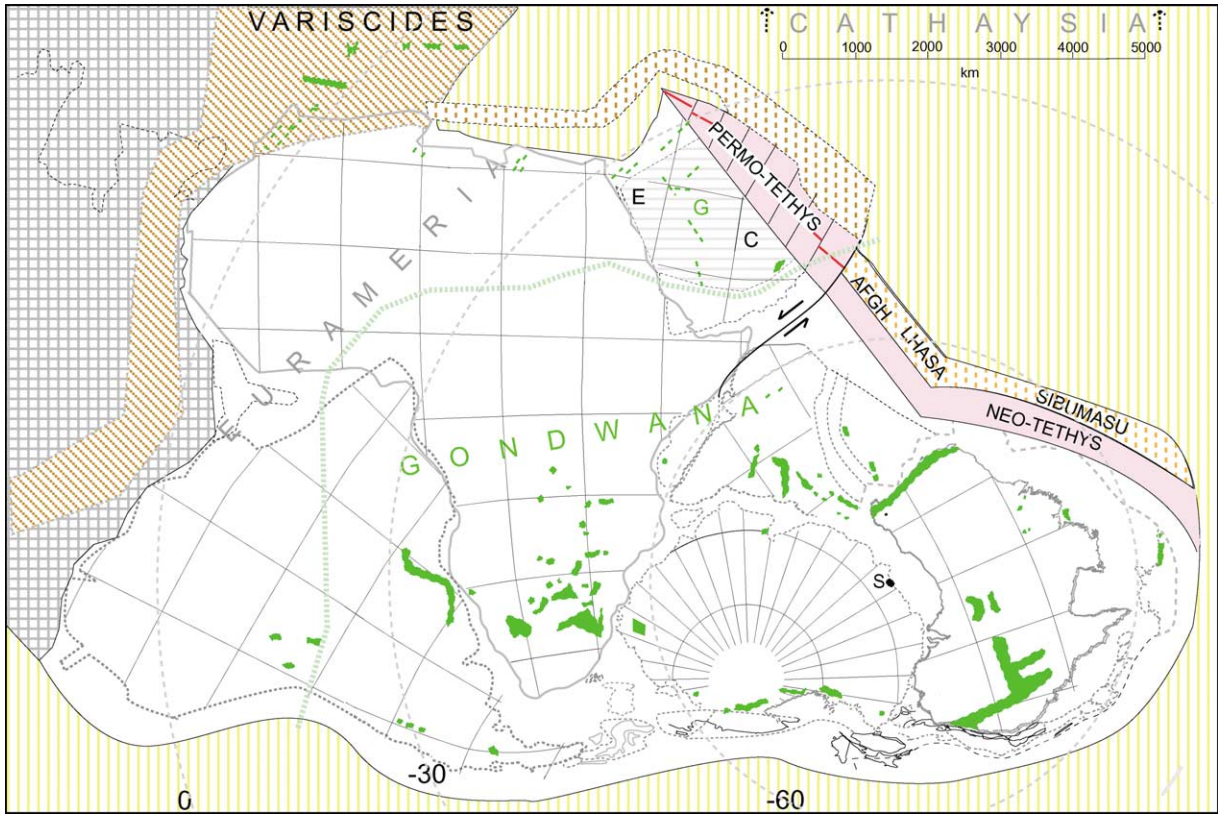


Figure 14 Distribution of Permian coal. The Permo-Tethys configuration pertains to the mid-Permian. E, Euramerica; G, Gondwana; C, Cathaysia. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

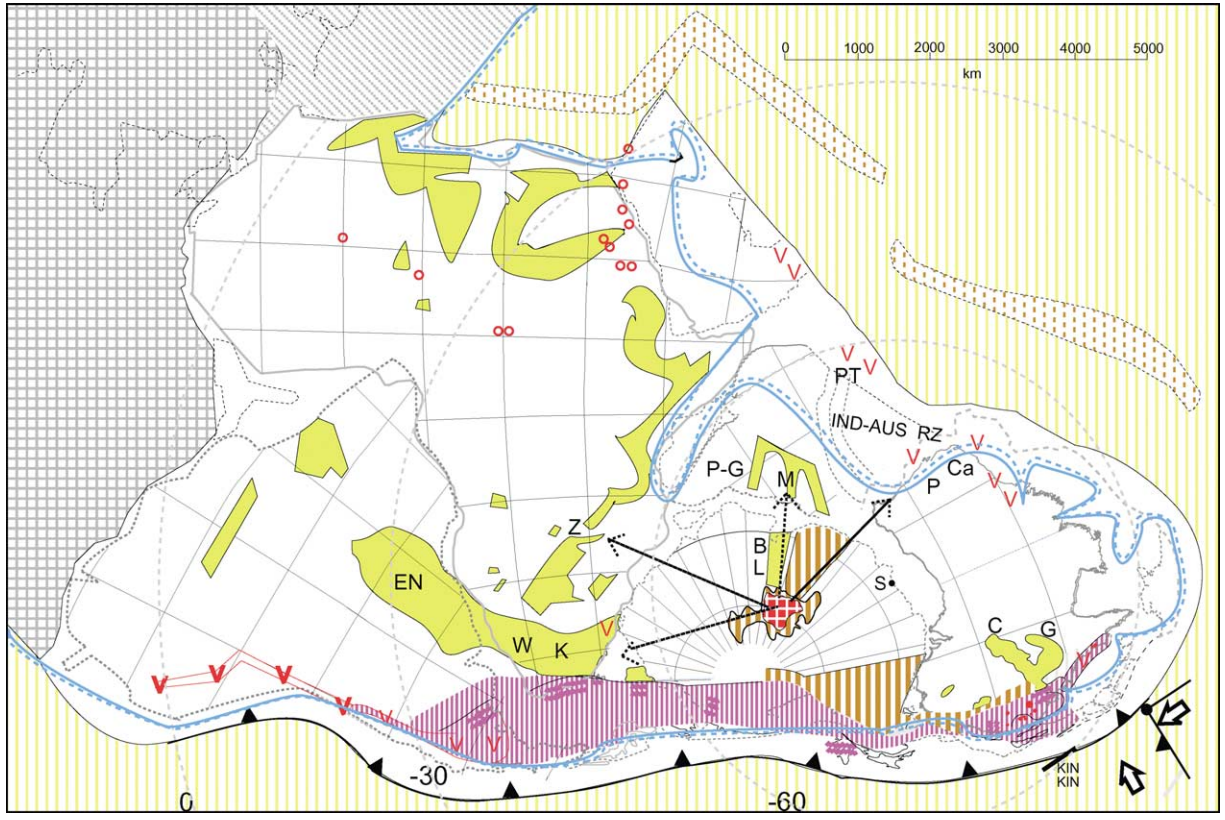


Figure 15 Late Permian (255–250 Ma) palaeogeography: The pink vertical shading indicates the magmatic orogen of Gondwanides I, the red open circles anorogenic igneous complexes, the red square pattern the central Gamburtsev upland. K, Karoo foreland basin; W, Waterford Formation; EN, Estrada Nova Formation; G, Galilee Basin; C, Cooper Basin; P, Perth Basin; Ca, Carnarvon Basin; BL, Beaver Lake Basin; M, Mahanadi Basin; P-G, Pranhita-Godavari Basin; Z, Zambezi Basin; PT, Panjal Traps, IND-AUS RZ, India–Australia Rift Zone. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

the epicratonic Galilee and Cooper Basins, and the Kin Kin Terrane docked. The Gamburtsev upland continued to send sediment into the Perth and Carnarvon Basins, through Beaver Lake to the Mahanadi Basin, and into the Pranhita-Godavari, Zambezi, and Karoo Basins. Broad fluviolacustrine basins covered North Africa. Volcanics, including the Panjal Traps, erupted in the India–Australia Rift Zone and in Arabia. Anorogenic complexes erupted in North Africa.

Early Triassic (250–235 Ma)

The end of the Permian saw a much lowered sea-level, and *Glossopteris* vanished as part of the end-Permian extinction event (*see* **Palaeozoic:** End Permian Extinctions). The Triassic saw global warming and the deposition of redbeds. The tectonic situation was unchanged. The Triassic sea returned to its former level, and nonmarine deposition resumed in most of the previous areas (Figure 16).

Mid-Triassic (234–227 Ma)

In the mid-Triassic the Panthalassan margin was terminally deformed (Gondwanides II). Deposition continued in South America and in North Africa-Arabia, which was intruded by anorogenic magmas (Figure 17).

Late Triassic (227–206 Ma)

During the late Triassic, the platform relaxed in Pangaean stage 3 (Extension II), and, following the Early and Middle Triassic coal gap, plant diversity and peat thickness recovered. Coal measures found accommodation space in Gondwanides II as the Molteno Coal Measures, the Lashly Formation, the Topfer Coal Measures, the New Town Coal Measures of Tasmania, and the Ipswich Coal Measures of Queensland (Figure 18). On the craton, the Leigh Creek Coal Measures were deposited on Neoproterozoic basement, and the carbonaceous sediment of the Peera Peera Formation filled the initial Eromanga Basin. Carbonaceous material is found in the McKelvey Member and the Dubrajpur and Colorados formations. Northern South America was crossed by volcanics, and North Africa was dotted with anorogenic magmas.

An arm of the sea crossed North America–Africa to produce evaporites; along its edge, grabens (Extension II) from Texas to Nova Scotia and Morocco filled with sediment that contained coal, as in the Productive Coal Measures of Maryland and Virginia.

Early Jurassic (200–184 Ma)

The vast 200 Ma Central Atlantic magmatic province of tholeiitic flows, dykes, and sills preceded the 190–180 Ma breakup of Pangaea (stage 4; spreading of intra-Pangaean rift oceans) by seafloor spreading (Figure 19). Another vast province of tholeiitic flows and sills was erupted between 184 Ma and 179 Ma in the back-arc region between southern Africa and south-eastern Australia, and southernmost South America was covered by felsic volcanics that have been dated at 187 Ma and younger.

End-Jurassic (145 Ma)

By the end of the Jurassic, the Central Atlantic had reached a width of 1000 km and was continuous through the Straits of Gibraltar with Neo-Tethys and its complex of marginal basins (Figure 20). The north-eastern Indian Ocean had opened at 156 Ma by seafloor spreading, wedging Argo Land off Australia, and the western Indian Ocean had begun to open at 150 Ma by seafloor spreading in the Natal and Somali basins.

The Chon Aike volcanics and a granite in Chile provide evidence of continuing subduction. The Antarctic margin was rifted in the Explora Wedge and Byrd Subglacial Mountains. Scattered magmatism continued in North Africa–Arabia. Nonmarine sediment was deposited in large areas of South America and Africa, on the north-western margin of India, in the rifts at the triple junction between India, Antarctica, and Australia, and in north-west and eastern Australia.

Mid-Cretaceous, Albian–Cenomanian (*ca.* 100 Ma)

Mid-Cretaceous shorelines, nonmarine basins, and igneous rocks are shown here on a pre-breakup base to maintain the same scale as previous figures; the space occupied by the oceans is denoted by a red broken line except between Antarctica and Australia, which were about to break up (Figure 21).

By the mid-Cretaceous, Gondwanaland had split into four pieces, with one of these about to split into two. Earth had entered the state of dispersed continents. The 94 Ma end-Cenomanian shoreline made its maximum penetration into North Africa and northern South America and its minimum into Australia. Earlier, in the Aptian (115 Ma), two-fifths of Australia had been covered by an epeiric sea; in the Cenomanian, the shoreline retreated to the present coast and beyond because uplift outpaced the eustatically rising sea; other continents sank passively beneath the rising sea. Australia's behaviour was caused by a

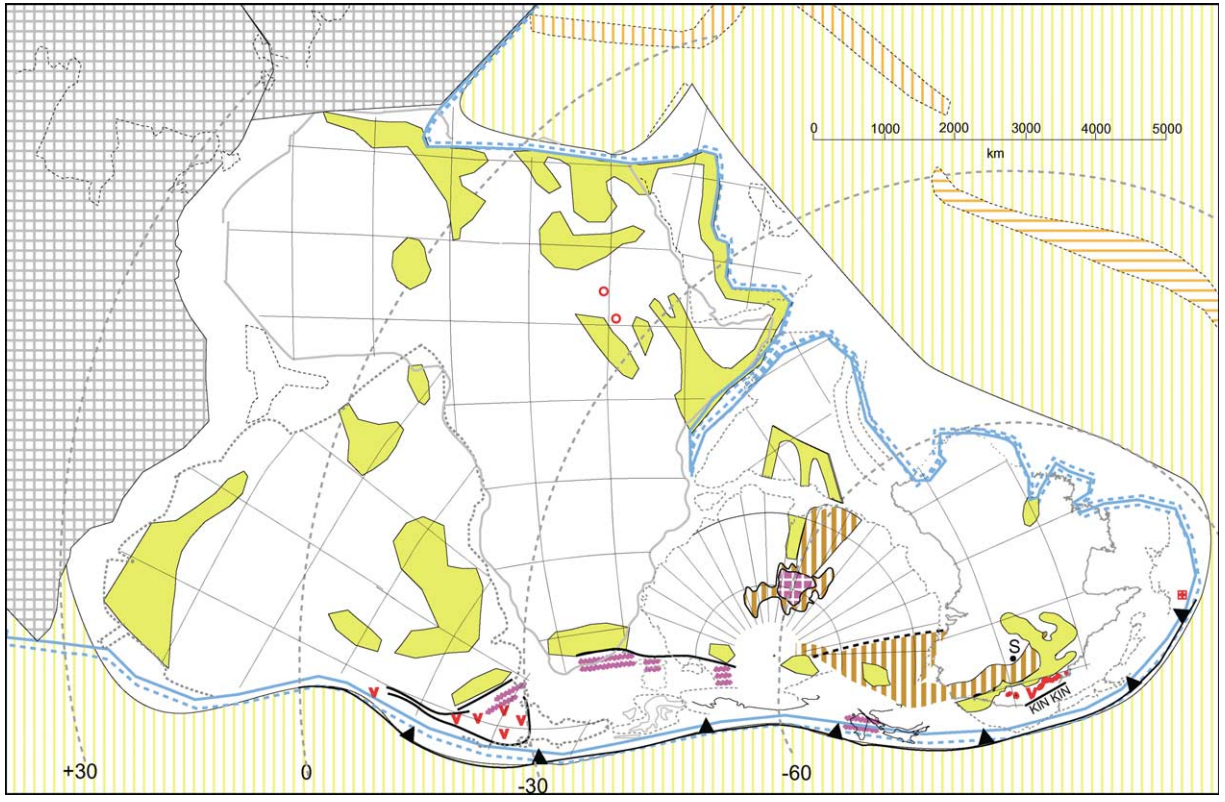


Figure 16 Early–Middle Triassic (250–234 Ma) palaeogeography. The orange vertical shading indicates uplands, the purplish wiggly lines zones of deformation, and red dots granitic plutons. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

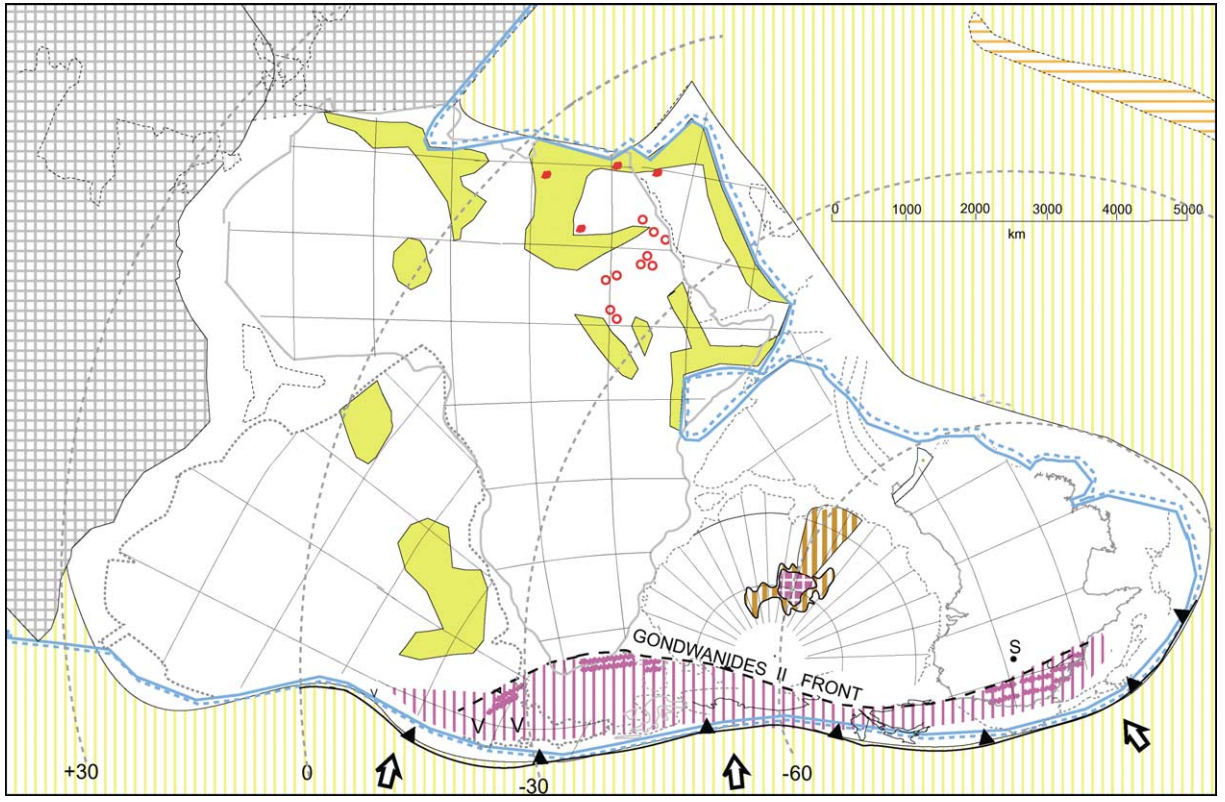


Figure 17 Mid-Triassic (234–227 Ma) palaeogeography. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

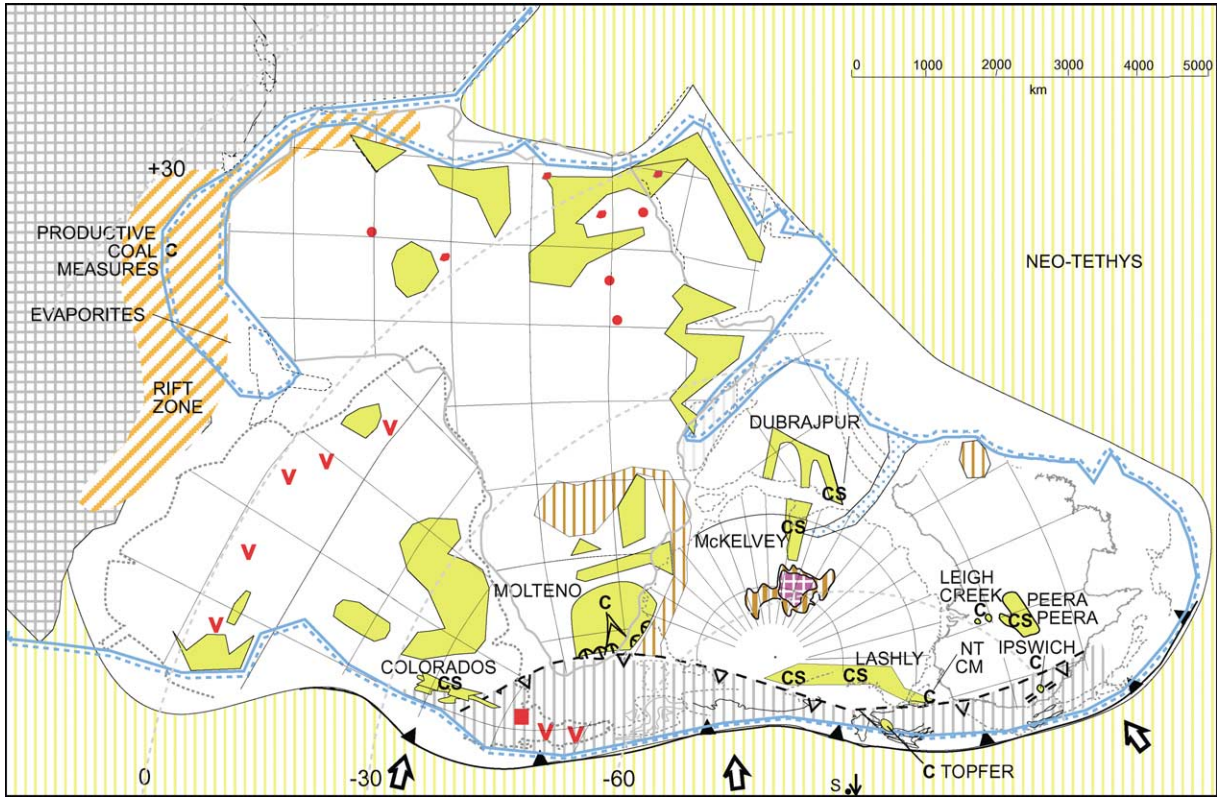


Figure 18 Late Triassic (215 Ma) palaeogeography: C, coal; CS, carbonaceous sediment; NTCM, New Town Coal Measures. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

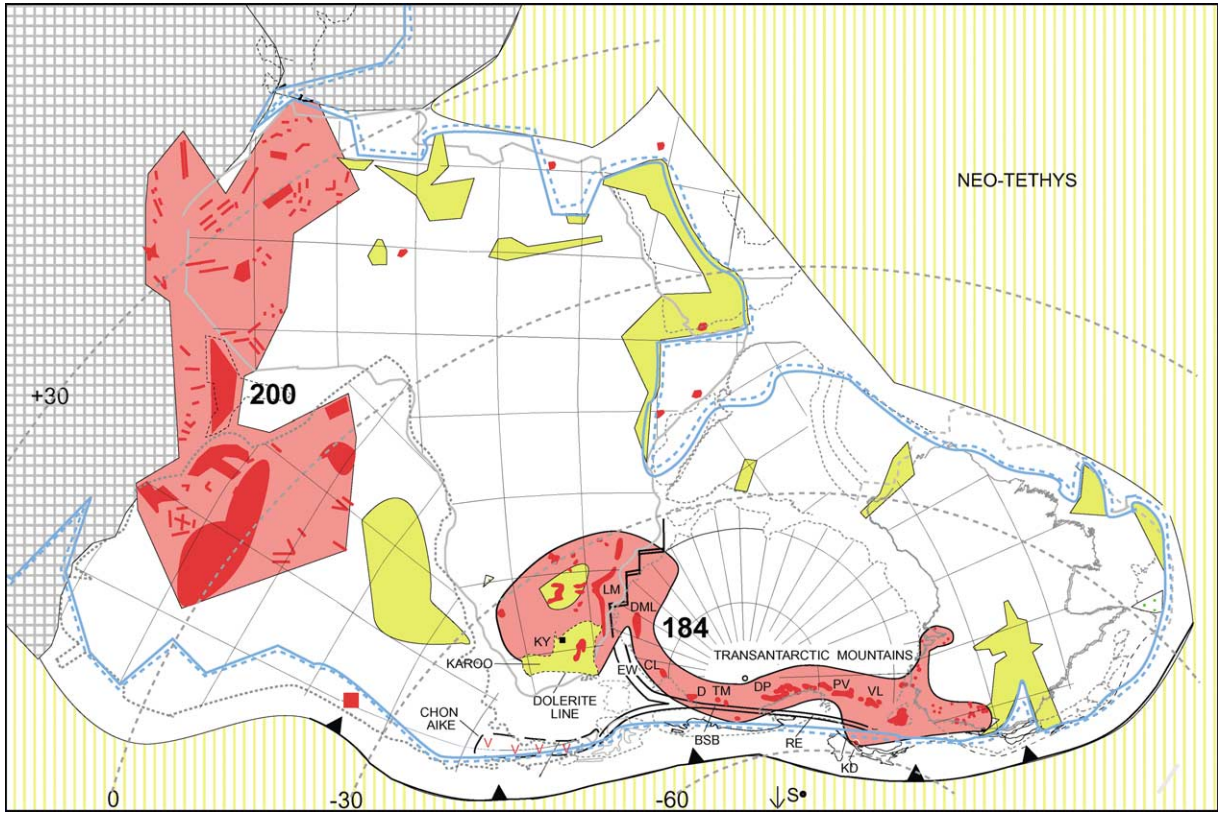


Figure 19 Early Jurassic (200–184 Ma) palaeogeography (ages are given in Ma). Areas covered by tholeiitic flows and sills (shown in pink, with outcrops in red) include the 200 Ma Central Atlantic magmatic province and another behind the Panthalassan margin that stretches from the Karoo through the Transantarctic Mountains, Tasmania, and New Zealand, to south-east Australia; individual localities, from left to right, are KY, Kimberley; LM, Lebombo monocline; DML, Dronning Maud Land; EW, Explora Wedge; CL, Coats Land; D, Dufek intrusion; BSB, Byrd Subglacial Mountains; TM, Thiel Mountains; DP, Dawson Peak; RE, Ross Embayment; PV, Pearse Valley; VL, Victoria Land; KD, Kirwans Dolerite; orange = upland; lime-green = nonmarine fringe behind shoreline. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

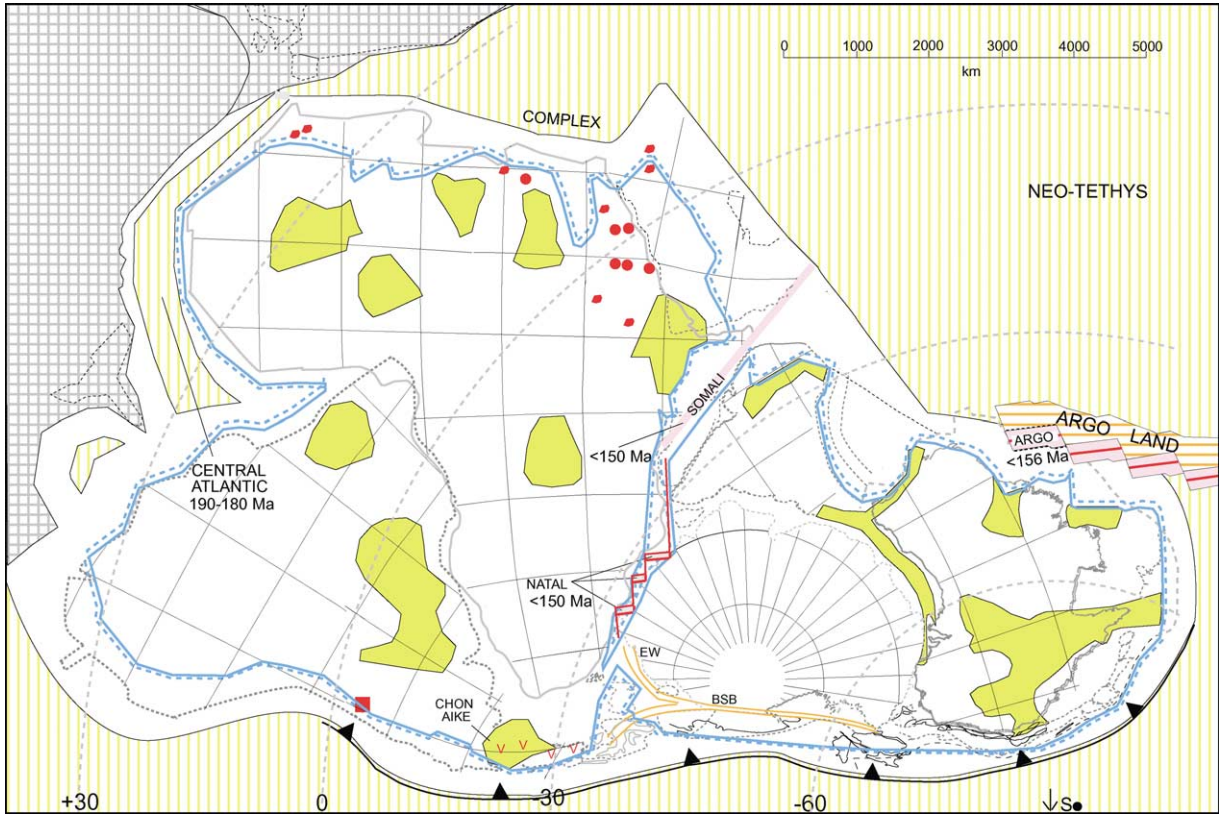


Figure 20 Late Jurassic (150 Ma) palaeogeography: EW, Explora Wedge; BSB, Byrd Subglacial Mountains. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

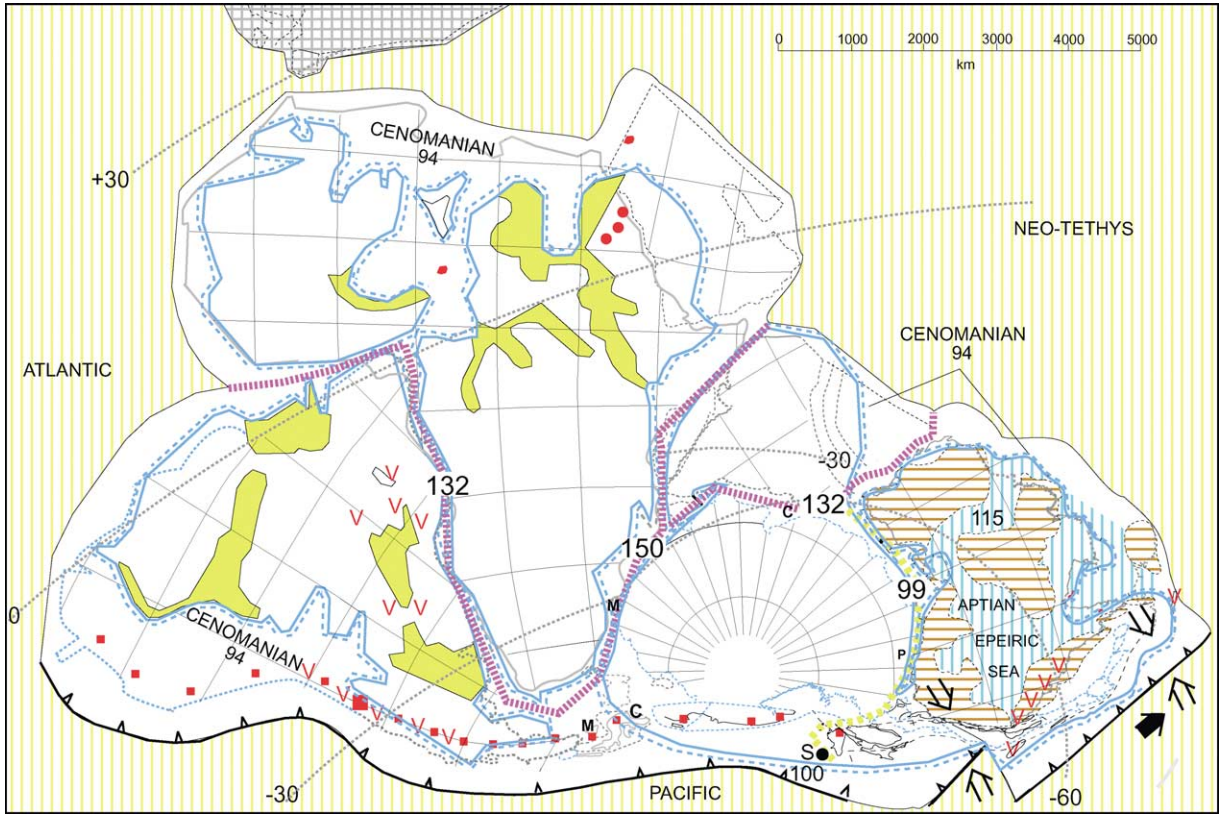


Figure 21 Mid-Cretaceous (*ca.* 100 Ma) palaeogeography of the individual continents, which are now dispersed, as shown in Figure 2, but which are shown together for cartographic convenience (the space occupied by the oceans is denoted by a red broken line). Ages of breakup are given in Ma. Also shown is the Aptian epeiric sea on Australia between land (blue pattern). Palaeolatitudes pertain to the conjoined Australia–Antarctica, and to the separate South America, Africa, and India. C, coal; P, nonmarine sediment with Aptian palynomorphs; M, marine deposits. Reproduced with permission from Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangaea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews*

change in the vector of the Pacific Plate with respect to Australia, from head-on collision in the Aptian to transcurrence or side-swipe in the Cenomanian, with concomitant cessation of subduction-related volcanism and uplift.

In Antarctica, coal is found on Alexander Island and in Prydz Bay; nonmarine sediment with Aptian palynomorphs is found in offshore Antarctica, with marine deposits on the opposite side of the continent. The Pacific arc glowed. Magmatic activity elsewhere was confined to volcanism in Brazil and anorogenic intrusion in North Africa–Arabia.

Cenozoic Aftermath (65–0 Ma)

During their dispersal (Figure 2), the Gondwanaland continents made new connections: South America with North America, India and Africa–Arabia with Eurasia, and Australia (almost) with south-east Asia. Antarctica alone remained unattached. Microcontinents (terranes) from the northern margin of Gondwanaland crossed successive generations of Tethys to lodge in Eurasia; New Zealand and New Caledonia remained isolated in the southern Pacific Ocean.

See Also

Africa: Pan-African Orogeny; North African Phanerozoic. **Antarctic**. **Argentina**. **Australia:** Phanerozoic; Tasman Orogenic Belt. **Brazil**. **New Zealand**. **Palaeozoic:** End Permian Extinctions. **Pangaea**.

Further Reading

Crowell JC (1999) *Pre-Mesozoic Ice Ages: Their Bearing on Understanding the Climate System*. Geological Society of America Memoir 192. Boulder: Geological Society of America.

Franca AB, Milani EJ, Schneider RL, *et al.* (1995) Phanerozoic correlation in southern South America. In: Tankard AJ, Suarez-Soruco R, and Welsink HJ (eds.) *Petroleum Basins of South America*, pp. 129–161. American Association of Petroleum Geologists Memoir 62. Tulsa: American Association of Petroleum Geologists.

Gaetani M (ed.) (2003) Peri-Tethys Programme. *Palaeogeography, Palaeoclimatology, Palaeoecology* 196: 1–263.

Nance RD and Thompson MD (eds.) (1996) *Avalonian and Related Peri-Gondwanan Terranes of the Circum-North Atlantic*. Special Paper 304. Boulder: Geological Society of America.

Schandelmeier H and Reynolds P-O (eds.) (1997) *Palaeogeographic–Palaeotectonic Atlas of North-Eastern Africa, Arabia, and Adjacent Areas*. Rotterdam: Balkema.

Selley RC (1997) The sedimentary basins of northwest Africa: stratigraphy and sedimentation. In: Selley RC (ed.) *African Basins*, pp. 3–16. Amsterdam: Elsevier.

Stampfli G, Borel G, Cavazza W, Mosar J, and Ziegler PA (eds.) (2001) *The Paleotectonic Atlas of the PeriTethyan Domain*. CD-ROM. Katlenburg-Lindau: European Geophysical Society.

Veevers JJ (ed.) (2000) *Billion-Year Earth History of Australia and Neighbours in Gondwanaland*. Sydney: GEMOC Press.

Veevers JJ and Powell CMcA (eds.) (1994) *Permian-Triassic basins and foldbelts along the Panthalassan margin of Gondwanaland*. Geological Society of America Memoir 187. Boulder: Geological Society of America.

Veevers JJ and Tewari RC (1995) *Gondwana Master Basin of Peninsular India between Tethys and the interior of the Gondwanaland Province of Pangea*. Geological Society of America Memoir 187. Boulder: Geological Society of America.

Veevers JJ (2001) *Atlas of Billion-Year Earth History of Australia and Neighbours in Gondwanaland*. Sydney: GEMOC Press.

Veevers JJ (2004) Gondwanaland from 650–500 Ma assembly through 320 Ma merger in Pangea to 185–100 Ma breakup: supercontinental tectonics via stratigraphy and radiometric dating. *Earth Science Reviews* [submitted November 2003].

GRANITE

See **IGNEOUS ROCKS: Granite**

GRENVILLIAN OROGENY

R P Tollo, George Washington University, Washington, DC, USA

Introduction

Orogenies are defined by extended periods of mountain building, usually resulting from convergence of tectonic plates. Such episodes in Earth's history typically involve a series of geological environments that reflect changes in the tectonic setting as convergence proceeds. The Grenvillian Orogeny is named after the village of Grenville in Québec, and the term is widely used to refer to a range of Mesoproterozoic tectonic events that occurred between 1.3 and ~1.0 Ga, resulting in development of a series of orogens that may have stretched across the globe for nearly 10 000 km. Within the Grenville Province of south-eastern Canada, which is the most thoroughly studied portion of this composite orogen, this period of orogenesis included (1) an early accretionary stage at 1.3–1.2 Ga, (2) an interval of widespread magmatism at 1.18–1.08 Ga, (3) and a period of continent–continent collision at 1.08–0.98 Ga that was rapidly followed by uplift and exhumation of the orogenic core. Development of the widespread Grenville orogen was the last major tectonic event to affect the Precambrian core of Laurentia, and marked the final stage in assembly of the Mesoproterozoic supercontinent of Rodinia. In North America, this widespread tectonism is recorded by a broad swath of igneous and metamorphic rocks extending 2000 km from the Atlantic coast of southern Labrador to Lake Huron in Canada and the Adirondacks in the United States (Figure 1). The belt of affected rocks continues south-westward for another 1500 km, mostly in the subsurface, to the Mississippi embayment in the United States, reappearing to the west in Texas and Mexico. Igneous and metamorphic rocks of similar age and tectonic affinity also occur in a series of internal and external massifs associated with the Appalachian orogen in the United States (Figure 2). Grenvillian rocks also constitute the Sveconorwegian Province of southern Norway and Sweden and are recognized as inliers within the Caledonides of Northern Ireland, Scotland, and Norway. Recent palaeogeographic reconstructions suggest that fragments of the dismembered orogen are also present in Antarctica, South America, and Australia.

Major geological events such as the Grenvillian Orogeny that result in assembly of supercontinents are relatively rare events in Earth's history, and the causal mechanisms are not yet well understood. Such events may be driven by global-scale geodynamic mechanisms, such as mantle downwelling, or may involve periodic random amalgamation of cratons resulting from subduction of the oceanic lithosphere. Geological studies of orogens and of the orogenies that produce them are typically spurred by economic factors, because orogenic belts contain much of the world's metallic mineral resources. Because orogens and orogenies are manifestations of large-scale Earth processes, enhanced understanding of the geological factors involved in their genesis also provides important evidence bearing on the mechanisms of plate tectonics and the physical evolution of the planet. The Grenvillian Orogeny represents an episode of unusually widespread tectonism that profoundly affected Earth's palaeogeography; like other major orogenies, such as the Permo-Carboniferous Appalachian Orogeny, the Grenvillian Orogeny marked the end of a major era in the geological time-scale of Earth's history.

Definition of the Grenvillian Orogeny

Geologists working in south-eastern Ontario and western Québec recognized in the early part of the twentieth century that the Grenville Province was structurally distinct from the rest of the Canadian Shield. Advances resulting from detailed field mapping, structural studies, and application of isotopic dating techniques in the 1960s and 1970s led to a more comprehensive understanding of the internal geology of the province and the timing of the Grenvillian Orogeny. During this time, researchers recognized that many areas of the orogen preserved tracts of older recycled crust that had not been completely overprinted by the effects of Grenvillian orogenesis. By 1980, geologists had determined that deposition of the Flinton Group, a Precambrian succession of metamorphosed clastic and carbonate rocks, was constrained to the interval 1080–1050 Ma, and that this deposition occurred during a tectonic hiatus separating two major episodes of orogenesis: (1) a pre-1080-Ma period of arc-related magmatism, uplift, and erosion and (2) a post-1050-Ma period of widespread regional metamorphism (see column A in Figure 3). The earlier episode of orogenesis was referred to as the Elzevirian Orogeny; the later episode is known as the Ottawan Orogeny. Both periods of orogenesis were considered part of the Grenville

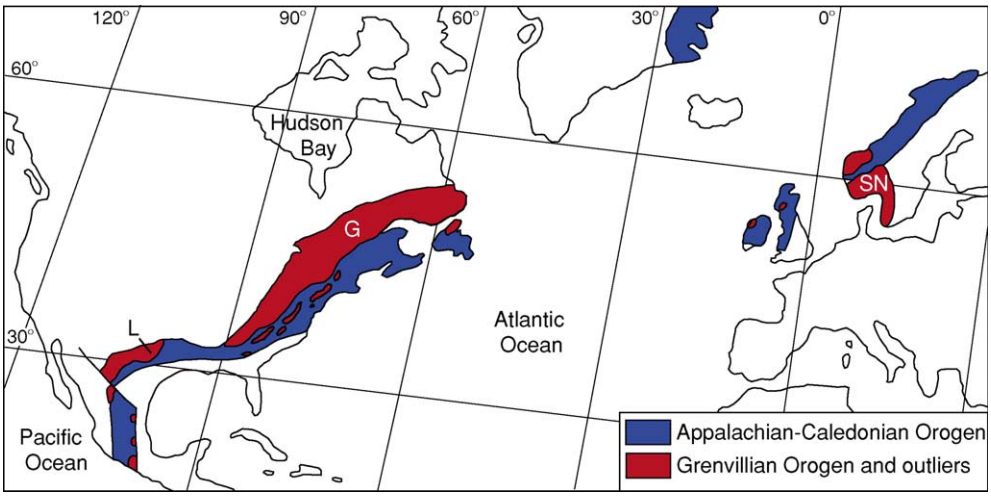


Figure 1 Map showing the location of the Mesoproterozoic Grenvillian Orogen (red) and outliers and the Palaeozoic Appalachian–Caledonian Orogen (blue). Abbreviations: G, Grenville Province of Canada; SN, Sveconorwegian Province of southern Norway and Sweden; L, Llano uplift. Modified with permission from Davidson A (1998) An overview of Grenville Province geology, Canadian Shield. In: Lucas SB and St-Onge MR (eds.) (coordinators) *Geology of the Precambrian Superior and Grenville Provinces and Precambrian Fossils in North America: Geological Survey of Canada, Geology of Canada, No. 7*, chapt. 3, pp. 205–270. Dartmouth, NS: Geological Survey of Canada.

Orogenic Cycle, which is widely defined as encompassing the interval 1.3–1.0 Ga. Definition of the Grenvillian Orogeny has evolved as new information has become available and as new perspectives have been presented (Figure 3). For example, in the Adirondack massif, where the age brackets on orogenesis closely resemble limits originally defined in the Canadian Province, the time interval between Elzevirian and Ottawan orogenesis did not involve widespread deposition of sediments; instead, this time interval was dominated by intrusion of large volumes of igneous rocks that collectively define the anorthosite–mangerite–charnockite–granite (AMCG) suite (see column B in Figure 3). Such rocks of similar age also occur in the Canadian Province. More recently, researchers recognized that evidence of Elzevirian orogenesis is limited to only a relatively small part of the Canadian Grenville Province and is indeed absent from much of the remainder of the Grenvillian orogenic belt. As a result, and because post-1080-Ma orogenesis ('Ottawan Orogeny' of the original definitions) is now understood to have been primarily responsible for the present geological configuration of the Grenville Province, this later period is referred to by many geologists as the 'Grenvillian Orogeny', although the time-frame for this activity and the existence of pulses of crustal shortening remain subjects of debate (compare columns C and D in Figure 3).

The evolution in the availability and interpretation of data has led to the term 'Grenvillian Orogeny', conveying different meanings to different geologists. To some, it refers to the *ca*. 1.3- to 1.0-Ga period of multistage orogenesis that ultimately resulted in amalgamation of the Grenville Province and creation of Rodinia; to others, it refers only to the culminating (typically post-1080 Ma) events of this large-scale process. In the following discussion of the tectonic events associated with creation of the Grenville Orogen, the 1.3- to 1.0-Ga interpretation is used because it is necessary to describe the sequence of convergent tectonic events that took place during closure of the ocean that bordered this part of Rodinia and to document more fully the nature of rocks constituting the orogenic belt. In this usage, the Grenvillian Orogeny (or Grenville Orogenic Cycle) includes two or more major periods of orogenic activity that have been given status as both 'pulses' and 'orogenies' by various authors (Figure 3). The terms 'Grenvillian Orogeny' (*sensu lato*) and 'Grenville Orogenic Cycle' are widely used in the geological literature to encompass the 1.3- to 1.0-Ga time-span of multiphase orogenic events. However, use of the latter term should be regarded with caution because orogenies are now understood to be lineally progressive phenomena in which crust evolves geologically from one state to another. In this way, orogenies differ from the more cyclic evolutionary history of ocean basins, which may involve successive periods of opening, closure, and reopening. Although unidirectional in modifying crust, orogenic tectonic activity is nevertheless typically diachronous along the length of an orogen, as illustrated by geological evidence from both

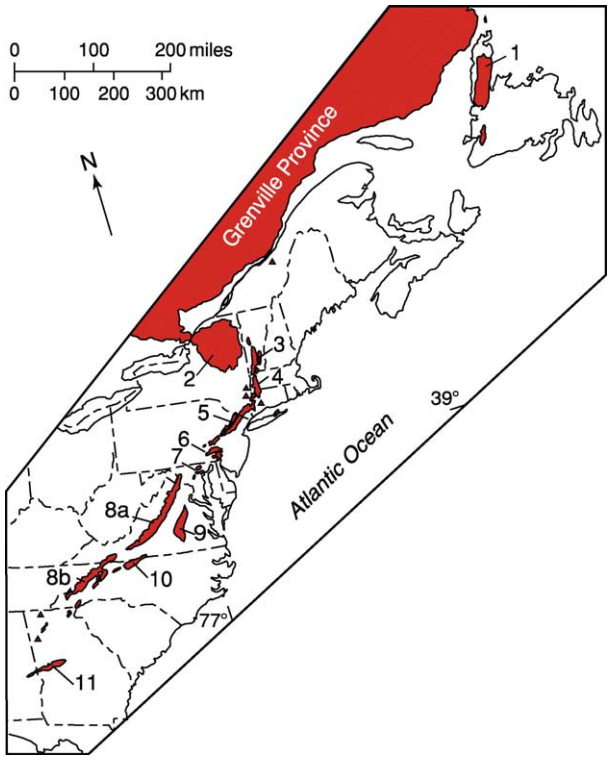


Figure 2 Map showing locations of major occurrences of the Grenville-age basement (shaded areas) in eastern North America, including inliers within the Appalachian orogen. 1, Long Range massif; 2, Adirondack massif; 3, Green Mountains massif; 4, Berkshire massif; 5, Reading Prong–New Jersey Highlands–Hudson Highlands; 6, Honey Brook upland; 7, Baltimore Gneiss antiforms; 8, Blue Ridge province, including Shenandoah (8a) and French Broad (8b) massifs; 9, Goochland terrane; 10, Sauratown Mountains anticlinorium; 11, Pine Mountain belt. Modified with permission from Rankin DW, Drake AA Jr., and Ratcliffe NM (1990) Geologic map of the U.S. Appalachians showing the Laurentian margin and Taconic orogen. In: Hatcher RD Jr., Thomas WA, and Viele GW (eds.) *The Appalachian–Ouachita Orogen in the United States*: *The Geology of North America*, vol. F-2, plate 2. Boulder, CO: Geological Society of America.

the Palaeozoic Appalachian and Mesoproterozoic Grenville belts.

Tectonic Evolution of the Grenvillian Orogeny

The most complete geological record of the Grenvillian Orogeny is preserved in the Grenville Province of Canada and the adjacent Adirondack massif in New York (Figure 2). Tectonic events in this region occurring within the 1.3- to 1.0-Ga timeframe record subduction of oceanic crust, closing of an ocean basin, accretion of magmatic arcs (Elzevirian Orogeny), sedimentation (Flinton Group and related metasedimentary successions) and AMCG magmatism, Himalayan-style continent–continent collision (Ottawan Orogeny), and uplift of the orogenic belt.

Elzevirian Orogeny

The culminating sequence of tectonic events that resulted in closure of an ocean bordering proto-North America and, ultimately, in creation of the Grenville orogen in its present form began between 1250 and 1230 Ma with a period of orogenesis termed the 'Elzevirian Orogeny'. (In this article, 'proto-North America' refers to the cratonic core of the eventual supercontinent Rodinia, prior to accretion of the Grenville orogen. 'Laurentia' came into existence following addition of the Grenville orogen.) Evidence for this orogeny is limited in Canada to the south-western portion of the province, and is unequivocally present in the United States only in the vicinity of the Llano Uplift in Texas (Figure 1). The Elzevirian Orogeny was preceded by subduction of oceanic crust and development of associated, possibly extensive, magmatic arcs that constitute the Composite Arc Belt in Canada and parts of the Adirondack Highlands (Figure 4) and Green Mountains in the United States (Figure 2). The Composite Arc Belt (Figure 4), where the most diverse assemblage of metamorphosed arc-related lithologies is recognized, consists mostly of <1300-Ma volcanic, plutonic, volcaniclastic, carbonate, and siliciclastic rocks derived from arcs and marginal basins. The inferred origin of many of these metamorphic lithologies as arc-related igneous rocks is based on the unique assemblage of lithologic types and chemical similarities between these rocks and calc-alkaline andesitic to tonalitic suites from modern arc terranes such as the western Americas, Aleutians, and Japan.

Prior to amalgamation with proto-North America, a composite Elzevirian volcanic-arc assemblage was separated from the main proto-North American craton by a back-arc basin and bordered by a subduction zone that dipped beneath the arc terrane and was responsible for production of the characteristic calc-alkaline magmatism (Figure 5A). This tectonic scenario bears similarities to the present Taiwan region in the western Pacific. The cessation of arc and back-arc magmatism at *ca.* 1230 Ma is interpreted to mark the termination of local subduction of oceanic lithosphere, an event that was followed by the Elzevirian Orogeny, which, by most recent estimates, occurred at 1230–1180 Ma, and is associated in the Canadian region with closure of the back-arc basin and accretion of the outboard Frontenac–Adirondack–Green Mountains crustal block (Figure 5B). Possible slices of the oceanic crust that previously separated the arc terrane from the continent are present as ophiolites within the accreted arc block. An important

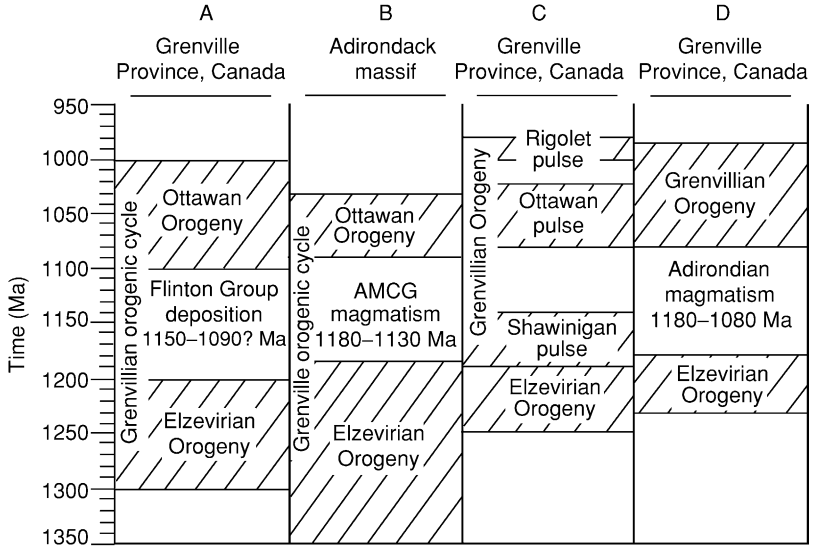


Figure 3 Comparative nomenclature for Late Mesoproterozoic orogenesis in the Grenville Province of Canada and the Adirondack massif, as published in the geological literature from 1980 to 2002. (Nomenclature was derived from the 1980, 1996, 1997, and 2002 work by Moore and Thompson, McLelland and co-workers, Rivers, and Gower and Krogh, respectively.) Diagonal pattern indicates time intervals of orogenesis. The various definitions of orogenic periods reflect differences in study areas, types of geological data compiled, and the type and quality of geochronological information available. AMCG, Anorthosite–mangerite–charnockite–granite.

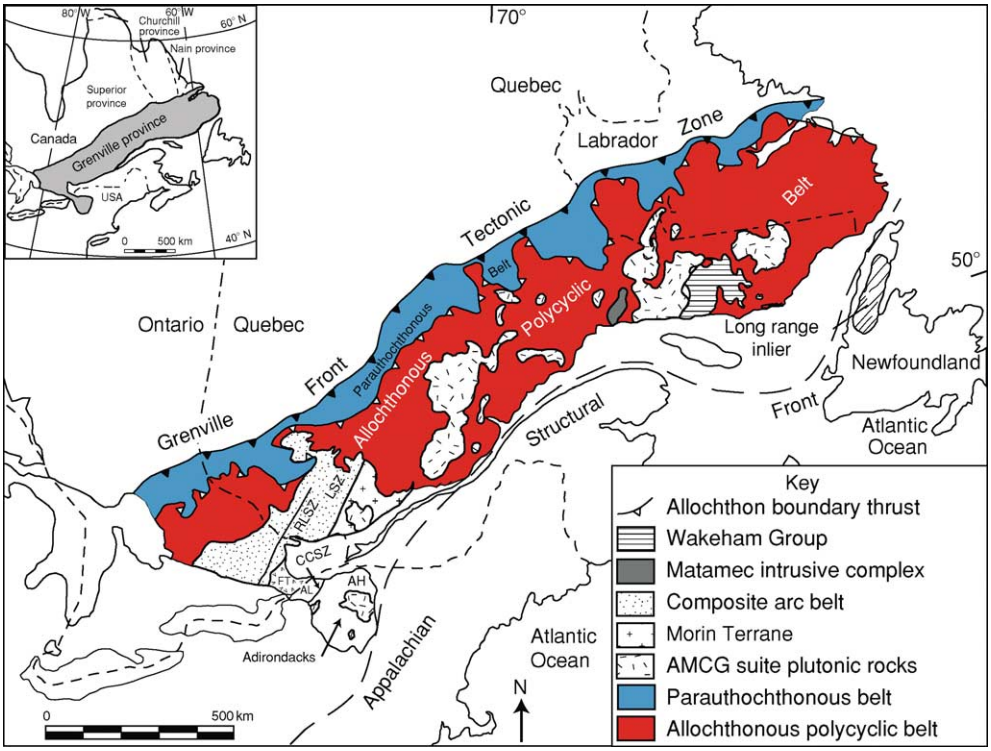


Figure 4 Map of the Grenville Province in Canada. Abbreviations: RLSZ, Robertson Lake shear zone; LSZ, Labelle shear zone; CCSZ, Carthage–Colton shear zone; FT, Frontenac Terrane; AL, Adirondack Lowlands; AH, Adirondack Highlands; AMCG, anorthosite–mangerite–charnockite–granite. Nomenclature for principal lithotectonic belts of the Canadian Grenville Province adopted from the work of the Carr group and Rivers and co-workers. Modified with permission from Davidson A (1998) An overview of Grenville Province geology, Canadian Shield. In: Lucas SB and St-Onge MR (coordinators) *Geology of the Precambrian Superior and Grenville Provinces and Precambrian Fossils in North America: Geological Survey of Canada*, *Geology of Canada, No. 7*, chapt. 3, pp. 205–270. Dartmouth, NS: Geological Survey of Canada.

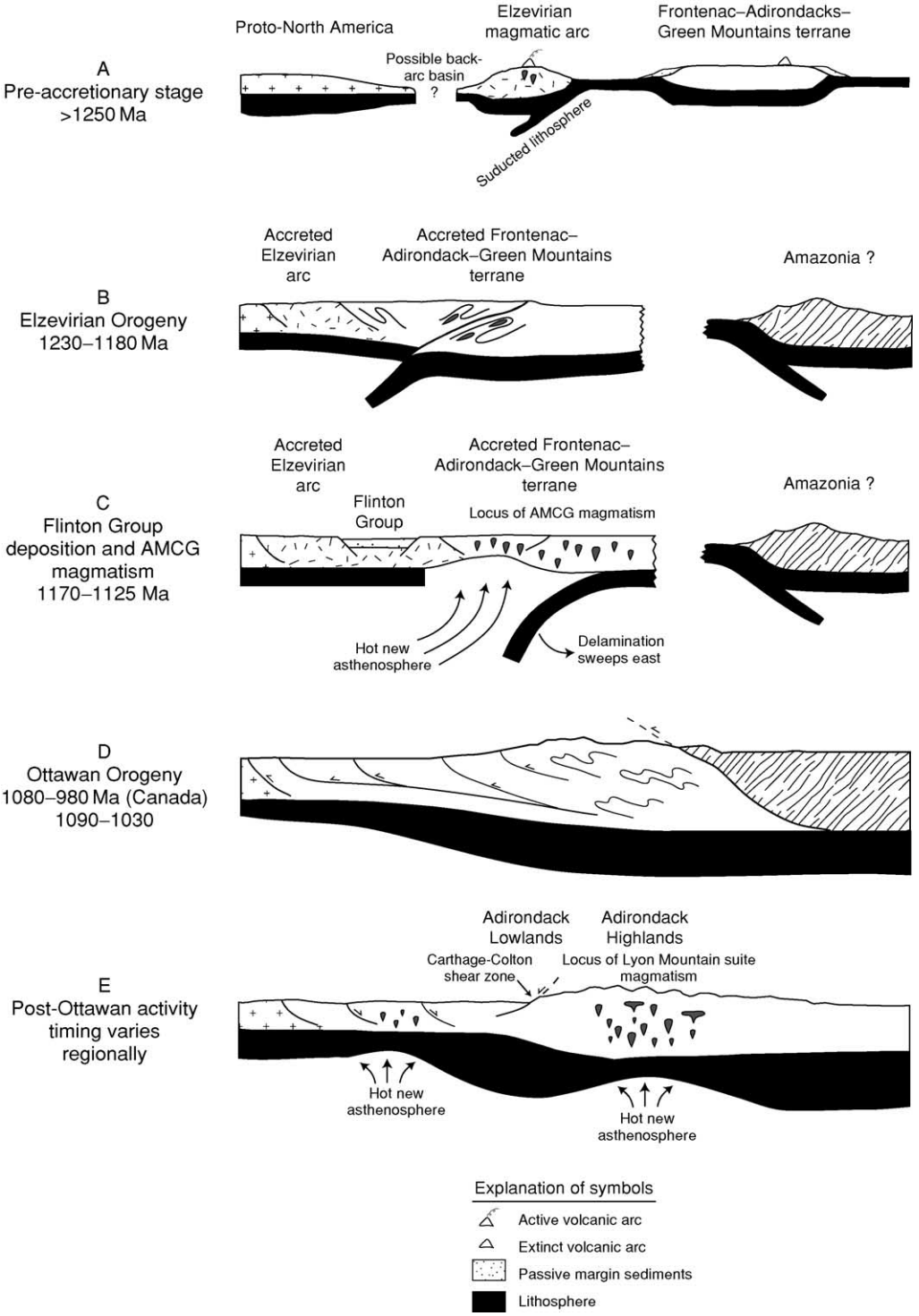


Figure 5 Summary of major events involved in Grenvillian orogenesis during the interval from 1.3 to 1.0 Ga in the southern Grenville Province of Canada and the Adirondack massif. Modified with permission from McLelland J, Daly S, and McLelland JM (1996) The Grenville Orogenic Cycle (*ca.* 1350–1000 Ma): an Adirondack perspective. *Tectonophysics* 265: 1–28.

consequence of closure of the back-arc basin was the shifting of the zone of active subduction outboard of the newly accreted crust and the establishment of a dominantly compressive tectonic setting; these events resulted in crustal thickening through thrust imbrication and the likely formation of a fold–thrust belt along the eastern (present coordinates) portion of the Composite Arc Belt, where rocks were locally

metamorphosed at amphibolite-facies conditions during Elzevirian orogenesis. A similar sequence of arc accretion accompanied by fold-and-thrust tectonism and metamorphism occurred during the Taconian orogenesis that affected much of the Appalachian region during the Ordovician. Such accretion of subduction-related arcs commonly occurs during the early stages in the tectonic evolution of collisional orogens involving closure of oceans. The regional scope of the Elzevirian Orogeny is a matter of debate because geological evidence of similar terrane amalgamation is generally lacking in the eastern Canadian Grenville Province. A possible arc-related assemblage of rocks is recognized in the Mesoproterozoic New Jersey Highlands (Figure 2), but is undated by modern isotopic techniques. The existence of a volcanic arc similar in age to the Elzevir Arc is documented in Texas, where it is believed to have docked with proto-North America at about 1150 Ma, producing polyphase deformation and high-pressure metamorphism within the Llano uplift.

Post-Elzevirian Activity

The period following Elzevirian orogenesis from 1180 to 1080 Ma is marked only by penecontemporaneous local sedimentation and widespread AMCG magmatism. In Canada, three small and isolated sedimentary successions of fluvial to shallow marine origin preserve evidence of crustal extension occurring across a broad area of the south-western Grenville Province. These rocks include the Flinton Group in eastern Ontario (Figure 5C), which is recognized as originally deposited unconformably on older basement rocks of Elzevirian affinity and subsequently isoclinally folded and metamorphosed to varying grade with the basement. Studies of detrital zircons indicate that the Flinton Group rocks, consisting of quartz-pebble conglomerate, pelite, quartzite, siltstone, and carbonate units, were deposited after 1150 Ma in a local fault-bounded basin. Studies of similar rocks exposed north-west of the Flinton Group indicate a likely comparable age of deposition.

The association of crust thickened by orogenic processes, deposition of sediments within encratonic basins, extensional faulting, and emplacement of mantle-derived AMCG igneous rocks led researchers to propose that subcontinental lithosphere beneath the Grenville orogen was probably replaced by asthenosphere during crustal shortening. Such juxtaposition of hot asthenosphere against thinned continental lithosphere would have profound effects throughout the overlying crust, and could have resulted from either lithospheric delamination or convective thinning of the lithosphere. These models produce similar consequences, including (1) establishment of steep temperature gradients in the continental lithosphere, (2) increases in surface elevations, and (3) creation of a thermal anomaly in the extended crust. In either case, magmatic activity such as represented by the Grenvillian AMCG suites is a likely result. In the southern Grenville Province and adjacent Adirondacks, where AMCG rocks are abundant, trends in crystallization ages of plutons indicate that the delamination model is more likely to account for this important period of igneous activity. The combination of overthickened crust and lithosphere resulting from Elzevirian orogenesis, and subsequent relaxation of contractural strain throughout much of the Grenvillian region, established conditions conducive to local delamination of upper mantle lithosphere (Figure 5C). Such delamination, which appears to have been concentrated in the southern Grenville Province–Adirondacks–Green Mountain region, resulted in crustal rebound and local structural collapse, forming isolated, fault-bounded basins in which the Flinton Group and related sedimentary successions were deposited. As delamination proceeded towards the south-east, the direction indicated by crystallization ages of related plutonism, high-temperature asthenosphere rose into the delaminated zone, bringing relatively hot, mafic melt towards the base of the crust. Tholeiitic gabbroic magmas produced by depressurization melting of the rising asthenosphere ponded at the base of the crust and underwent fractional crystallization within a relatively quiescent, nonorogenic environment, resulting in the production of plagioclase-rich crystal mushes through flotation and concomitant sinking of olivine–pyroxene-rich cumulates (Figure 6). Heat from the gabbroic magmas resulted in partial melting of the lower crust, producing relatively anhydrous felsic magmas that intruded to shallow levels to form plutons of syenitic, monzonitic, and granitic composition, which, on crystallization of orthorhombic pyroxene, locally yielded mangerites and charnockites. Ultimately, plagioclase-rich magmas originally produced through the aforementioned crustal melting used crustal fractures and other weaknesses to ascend to the level of the felsic melts, producing anorthosites and the related AMCG suites that are characteristic of the Grenville Province. Some of the anorthositic complexes form very large, petrologically diverse plutonic bodies, such as the Lac St. Jean AMCG suite in the central Grenville Province and the Marcy Massif in the Adirondacks. Further fractional crystallization of the plagioclase melts may have resulted in the formation of mafic-rich residual liquids represented by ferrodiorites (also referred to as ferrogabbros and jotunite), which, through immiscibility, may have in

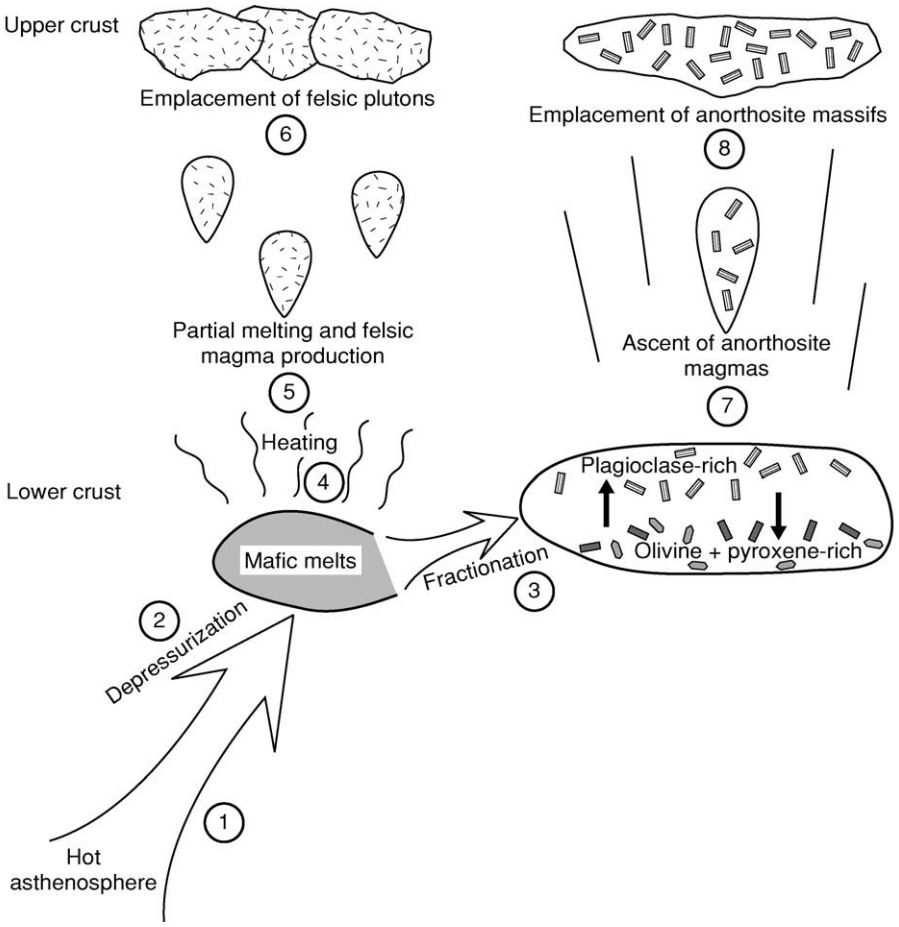


Figure 6 Schematic summary of a possible model for generation of Grenvillian anorthosite–mangerite–charnockite–granite suites. Following either lithospheric delamination or convective thinning of the lithosphere resulting from crustal thickening related to Elzevirian orogenesis, the hot asthenosphere rises towards the base of the intact crust (1), where depressurization reactions result in production of mafic magmas (2). These mafic magmas undergo crystallization under relatively quiescent, nonorogenic conditions, then fractionate to produce plagioclase-rich melts, through flotation, and olivine + pyroxene-rich melts, through gravitational settling (3). The mantle-derived mafic melts also provide significant heat transfer to the lower crust (4); this results in partial melting and production of felsic magmas (5) that rise buoyantly towards emplacement in the upper crust as plutons composed of granite, charnockite (syenogranite containing orthorhombic pyroxene), and mangerite (monzonite containing orthorhombic pyroxene) (6). The plagioclase-rich differentiates of the mantle-derived gabbroic magma use fractures and other crustal weaknesses to ascend within the crust (7) towards ultimate emplacement as petrologically complex massifs containing anorthosite and related anorthosite–mangerite–charnockite–granite rocks.

turn been responsible for the production of associated ilmenite–magnetite ores.

AMCG magmatism was widespread throughout the south-western Grenville Province and the Adirondacks during the 1180- to 1080-Ma interval, and marked an abrupt change in the nature of plutonic activity within the region. Igneous activity related to the Elzevirian magmatic arc was calc-alkaline and thus characteristic of rocks associated with modern convergent plate margins. In contrast, the AMCG suites are dominantly tholeiitic and are similar to igneous rocks associated with modern within-plate tectonic environments. Such contrasts in composition reflect the differing nature of the sources involved in petrogenesis. These differences underscore the usefulness of geochemical studies of ancient igneous and meta-igneous rocks to decipher both the modes and the environments of origin. Massif anorthosites and associated mangerites and granites have traditionally been considered to constitute an 'anorogenic trinity', the formation of which is indicative of crustal conditions in which orogenic effects are absent. However, considering the entire Grenville Province, it is clear that metamorphism and contractural deformation occurred in some parts of the province at the same time that AMCG suites were emplaced in other parts. Such contrasting styles of tectonic activity should perhaps be expected in an area that is as

large and complex as the Grenville Province is. The preponderance of available data indicate that such suites are a likely consequence of orogenesis and result from deep-seated lithospheric mechanisms that originated in response to previous orogenic processes. Viewed in this light, the Grenvillian AMCG suites represent postorogenic magmatism with origins that relate directly, albeit locally, to Elzevirian orogenesis.

Ottawan Orogeny

Final closure of the ocean bordering proto-North America occurred as a result of continent-to-continent collision beginning at 1080 Ma, involving a series of events referred to as the Ottawan Orogeny (also as the 'Grenvillian Orogeny' by some geologists; see Figure 2). Compressional tectonic forces resulted in telescoping of the continental margin and the formation of large thrust slices and ductile folds with dominant direction of transport towards the north-west (Figure 5D). Multiple lines of evidence indicate that the orogen propagated north-westward, with the youngest faults located in the vicinity of the Grenville Front Tectonic Zone (Figure 4). Metamorphic assemblages in the Grenville Province generally indicate recrystallization at granulite-facies conditions in the hinterland (region located towards the south-east), where large-scale, recumbent nappes were developed and most plutonism occurred. Thrusting was more common in the foreland (region located towards the north-west), where metamorphic conditions were more variable, ranging from greenschist to high-pressure amphibolite facies, with eclogite-facies conditions developed locally. Many of the north-west-directed thrusts were of large scale, including the Allochthon Boundary Thrust (Figure 3) that is associated with high-pressure mineral assemblages. Results from many studies throughout the province indicate that metamorphism and deformation were ongoing across the entire province during the interval from 1080 to 980 Ma. However, disagreement exists regarding the nature of these tectonic processes. Some researchers suggest that metamorphism and deformation occurred in at least three discreet pulses (see column C in Figure 2) separated by intervals of extension. The first two pulses of crustal shortening were concentrated in the hinterland, whereas the latest pulse caused north-westward propagation of the orogen into its foreland. Abundant magmatism, including emplacement of anorthosite complexes, occurred during the intervening extensional periods. A different interpretation of the tectonic evolution of the eastern Grenville Province suggests that orogenesis peaked at different times in different places, and that, as the crust yielded through thrusting in one location, stresses were transferred elsewhere. Thus, instead of separate orogenic pulses, this model proposes that the 'pressure point' responsible for compressional tectonism shifted periodically as the crust yielded locally to the forces of deformation.

The effects of the Ottawan Orogeny were generally more intense in the Adirondack massif, where evidence of ductile structural fabrics and high-grade metamorphism is widespread. Studies of metamorphic rocks indicate that the crust beneath the Adirondack Highlands nearly doubled in thickness as a result of Ottawan orogenesis. Such extreme crustal thickening, considered together with the general lack of Ottawan-age calc-alkaline rocks of possible arc affinity in the Adirondacks and contiguous Grenville Province, constitutes important evidence in support of continent–continent collision as the primary cause of the orogeny. The Adirondack Highlands represent one of the world's classic granulite-facies terranes, where peak metamorphic conditions during the Ottawan Orogeny reached temperatures of 750–800°C at pressures of 6–8 kbar. Metamorphism was associated with recumbent isoclinal folding and development of intense penetrative fabrics that typically overprinted features developed during the Elzevirian orogenesis throughout the Adirondacks (Figure 7A and B). However, the anorthosite complexes in this area, all of which were emplaced at about 1150 Ma, are typically not structurally overprinted (Figure 7C) because these generally anhydrous, dominantly monomineralic plutons acted as rigid bodies that deflected stresses during Ottawan orogenesis. Such kinematics that result in plutonic rocks that appear to be relatively undeformed when, in fact, they predate deformation, must be considered carefully by field geologists interested in deciphering the sequence of local structural events. Field relations and isotopic ages of plutonic rocks indicate that Ottawan orogenesis occurred from 1090 to 1030 Ma in the Adirondacks. The occurrence of thick allochthons and associated high-pressure belts in the Canadian Grenville Province stands in contrast to the dominantly ductile deformation and large recumbent fold nappes of the Adirondacks. Such differences are attributed to variations in style, metamorphic grade, and ductility between the mobile interior of the orogen to the south-east and the foreland to the north-west, as well as to the thermal softening effects imparted on the Adirondacks by intrusion of the Hawkeye granitic suite at 1100 Ma.

Post-Ottawan Activity

Tectonic activity associated with the Ottawan Orogeny did not terminate simultaneously throughout

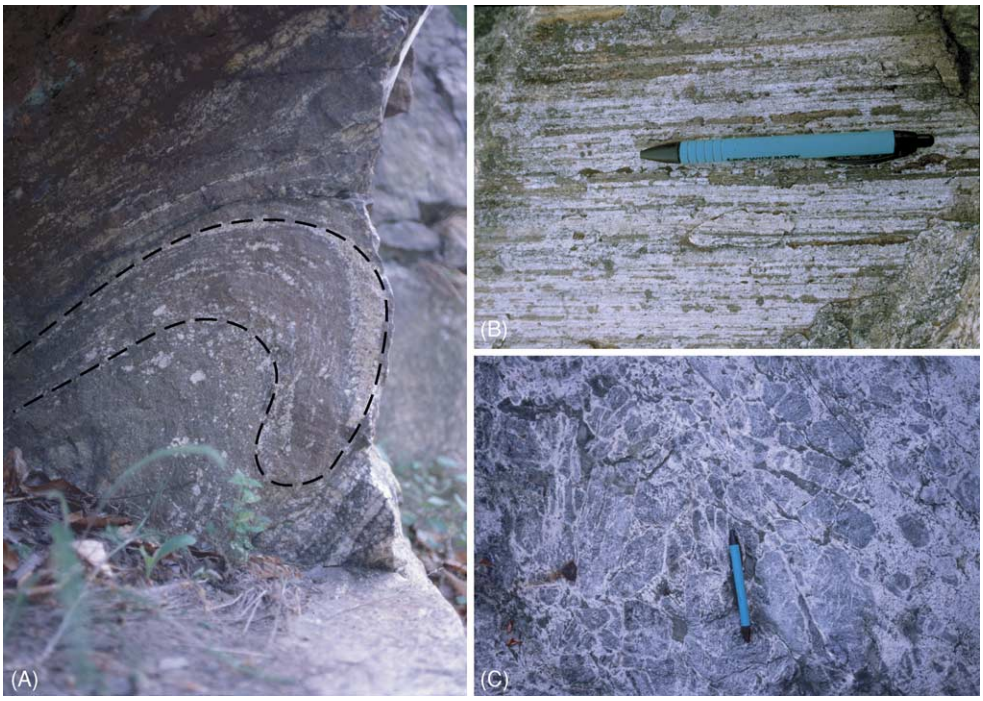


Figure 7 Deformation features developed during Ottawan orogenesis in the Adirondack Highlands. (A) Refolded isoclinal fold developed in Elzevirian-age migmatitic metapelite. Dashed line outlines the fold, which is about 1 m in length. (B) L-Tectonite (also referred to as 'straight gneiss'), developed through extreme grain-size reduction accompanying ductile deformation in migmatitic metapelite located between the Piseco anticline and Glen Falls syncline in the south-eastern Adirondacks. This is the deformed equivalent of the rock shown in A. Pen is oriented parallel to the tectonic transport direction and is 14 cm in length. (C) Inequigranular anorthosite from the southern edge of the Oregon dome in the south-eastern Adirondacks. Large blue–gray andesine crystals are typical of Grenvillian massif anorthosites. Pen is 14 cm in length.

the Grenville orogen, but rather ended with a series of sporadic events. Within Canada, north-west-directed thrusting and shearing along the Grenville Front Tectonic Zone (Figure 4) occurred as late as 1000–980 Ma, and weakly deformed 980-Ma AMCG plutons document the waning stages of orogenesis in the north-eastern Grenville Province. In contrast, undeformed 975-Ma granitic intrusions that spread across the southern part of the eastern Grenville Province appear to be postkinematic, signalling a likely end to contractional tectonics. Unmetamorphosed and undeformed 1045-Ma fayalite granite, part of the late synkinematic to postkinematic Lyon Mountain granite suite, provides a minimum age for the end of widespread Ottawan activity in the Adirondacks, although localized high-temperature events occurred until about 1015 Ma. Rapid uplift followed Ottawan orogenesis in the Adirondacks and neighbouring Grenville Province, possibly caused by a second cycle of lithospheric delamination (Figure 5E). This uplift was accompanied by extensional collapse, especially along the Carthage–Colton shear zone, where amphibolite-facies rocks of the Adirondack Lowlands were displaced downwards to the north-west to become juxtaposed against the granulite-facies Adirondack Highlands.

Appalachian Inliers

Grenville-age rocks within the Appalachians preserve evidence of geological processes that are grossly similar to those documented in the Canadian Grenville Province and Adirondacks. Rocks of AMCG affinity are present in some massifs, and a major period of metamorphism and deformation that is roughly contemporaneous with Ottawan orogenesis in south-eastern Canada has been documented in most areas. However, direct evidence of accreted arc terranes and orogenesis that might have been contemporaneous with Elzevirian activity in Canada is lacking in most massifs. Indeed, some researchers have used results from Pb and Nd isotopic studies to propose a non-Laurentian origin for the Appalachian inliers, in spite of the similarities in lithologies and tectonic evolution observed between these massifs and areas of the Adirondacks and Canadian Grenville Province.

Plate Tectonic Setting

The tectonic truncation of Precambrian geological belts in the south-western United States led to the search for possible continuations of the Grenville Orogen on other continents. Such investigations ultimately resulted in palaeogeographic reconstructions of a possible supercontinent named Rodinia (from the Russian verb 'rodit', meaning 'to beget' or 'to grow'), the amalgamation of which was marked in part by the collisional Grenvillian Orogeny. Two such reconstructions, the south-west United States–East Antarctic (SWEAT) and Australia–south-west United States (AUSWUS) hypotheses, extend the Precambrian geology of the south-western United States to the Australia–Antarctic shield region, but differ in the proposed position of the Australian craton relative to the conjugate margin. Another reconstruction, based largely on sedimentological patterns, proposed a Proterotozoic connection between Siberia and western Laurentia. Such palaeogeographic reconstructions, which depend on many lines of multidisciplinary geological evidence, are constantly refining geologists' view of Mesoproterozoic Earth.

Figure 8 Plate tectonic reconstruction of Rodinia at about 1000 Ma, with Laurentia located in its present-day North American coordinates. Orogenic belts are depicted in random-line pattern and include the Grenville belt (1), Llano belt (2), Namaqua belt (3), Albany/Fraser belt (4), and East Ghats belt (5). Labelled cratons include Baltica (B), Amazonia (AM), Rio de Plata (RP), Siberia(S,), Australia (AUS), East Antarctica (EANT), India (IND), Madagascar (M), Kalahari (K), and Congo (C). Modified with permission from Dalziel IWD, Mosher S, and Gahagan LM (2000) Laurentia–Kalahari collision and the assembly of Rodinia. *The Journal of Geology* 108: 499–513.

The sequence and style of structural and magmatic-thermal events, extended time-span of development, and considerable degree of crustal thickening that characterized Ottawan orogenesis are consistent with the case histories of other major mountain belts, including the Himalayas. Although the cause of the Ottawan Orogeny is not yet rigorously identified, there is general agreement that it involved collision with a large outboard continental block located to the south-east. Recent palaeogeographic reconstructions indicate that Grenville orogens defined a segmented belt that resulted from convergence of multiple cratons at 1.2–1.0 Ga. The type-Grenville belt of eastern Laurentia and its possible southern continuation, the Llano-Namaqua Orogen of Texas and south-west Africa, formed a major part of this orogenic global network (Figure 8), resulting from accretion of the Amazonian and Rio de Plata cratons to the large nucleus of Rodinia (consisting of Laurentia, East Antarctica, and the Grawler Craton of south-eastern Australia) at *ca.* 1.1 Ga. The Llano-Namaqua Orogen formed nearly synchronously as a result of collision with the Kalahari Craton (Figure 8). These tectonic events were part of a series of cratonic collisions that took place on a global scale at 1.2–1.0 Ga during a relatively short period in which most of most of Earth's continental lithosphere was amalgamated to form the supercontinent of Rodinia, the continental shelves of which were later to give rise to many of Earth's earliest organisms and the eventual dismemberment of which produced all subsequent continents.

See Also

Europe: Scandinavian Caledonides (with Greenland). **Igneous Processes**. **Metamorphic Rocks:** PTt - Paths. **North America:** Precambrian Continental Nucleus; Northern Appalachians; Southern and Central Appalachians. **Plate Tectonics**. **Precambrian:** Overview. **Tectonics:** Mountain Building and Orogeny.

Further Reading

Carr SD, Easton RM, Jamieson RA, and Culshaw NG (2000) Geologic transect across the Grenville orogen of Ontario and New York. *Canadian Journal of Earth Sciences* 37: 193–216.

Condie KC (1997) *Plate Tectonics and Crustal Evolution.* Oxford: Butterworth-Heinemann.

Dalziel IWD, Mosher S, and Gahagan LM (2000) Laurentia–Kalahari collision and the assembly of Rodinia. *The Journal of Geology* 108: 499–513.

Davidson A (1998) An overview of Grenville Province geology, Canadian Shield. In: Lucas SB and St-Onge MR (eds.) (coordinators) *Geology of the Precambrian*

Superior and Grenville Provinces and Precambrian Fossils in North America: Geological Survey of Canada, Geology of Canada, No. 7, chapt. 3, pp. 205–270. Dartmouth, NS: Geological Survey of Canada. (Also published as *The Geology of North America* (1998), vol. C-1. Washington, DC: Geological Society of America.)

Gower CF and Krogh TE (2002) A U-Pb geochronological review of the Proterozoic history of the eastern Grenville Province. *Canadian Journal of Earth Sciences* 39: 795–829.

Hynes A and Ludden JN (eds.) (2000) *Canadian Journal of Earth Sciences, Special Issue: The Lithoprobe Abitibi-Grenville Transect*, vol. 37, nos. 2 and 3. Ottawa: NRC Research Press.

McLelland J, Daly S, and McLelland JM (1996) The Grenville Orogenic Cycle (ca 1350–1000 Ma): an Adirondack perspective. *Tectonophysics* 265: 1–28.

Moore J and Thompson P (1980) The Flinton group: a late Precambrian metasedimentary sequence in the Grenville Province of eastern Ontario. *Canadian Journal of Earth Sciences* 17: 1685–1707.

Moores EM and Twiss RJ (1995) *Tectonics*. New York: WH Freeman.

Rankin DW, Drake AA Jr., and Ratcliffe NM (1990) Geologic map of the U.S. Appalachians showing the Laurentian margin and Taconic orogen. In: Hatcher RD Jr., Thomas WA, and Viele GW (eds.) *The Appalachian–Ouachita Orogen in the United States: The Geology of North America,* vol. F-2, plate 2. Boulder, CO: Geological Society of America.

Rivers T (1997) Lithotectonic elements of the Grenville Province: review and tectonic implications. *Precambrian Research* 86: 117–154.

Wardle RJ and Hall J (eds.) (2002) *Canadian Journal of Earth Sciences, Special Issue: Proterozoic Evolution of the Northeastern Canadian Shield: Lithoprobe Eastern Canadian Shield Onshore-Offshore Transect*, vol 39, no. 5. Ottawa: NRC Research Press.

HERCYNIAN OROGENY

See **EUROPE: Variscan Orogeny**

HIMALAYAS

See **INDIAN SUBCONTINENT**

HISTORY OF GEOLOGY UP TO 1780

O Puche-Riart, Polytechnic University of Madrid, Madrid, Spain

Introduction

Ancient civilisations in contact with nature inquired about their origins and about particular geodynamic phenomena. In most cases they satisfied themselves with empiric explanations; they even used deities in order to understand inexplicable situations.

Little by little humans learnt how to observe their environment and arrange processes. During the Renaissance the first geologic principles were born and this knowledge spread rapidly. Natural phenomena were understood in terms of dynamic cause–effect, although many dogmatic and magic interpretations persisted.

Many authors agree that geology, began to be structured as a science in the second half of the eighteenth century with Abraham Gottlob Werner (1749–1817), father of Neptunism (Figure 1). However, some geologic paradigms such as diluvialism existed before neptunism; all of them contained countless mistakes and ambiguities.

This article outlines the period up until 1780, which thus incorporates the work of James Hutton (*see* **Famous Geologists:** Hutton). His ideas were important in the development of geology, more specifically relating to the origins and dating of rocks. Geology was not completely defined till the birth of Stratigraphy at the end of the eighteenth century and Palaeontology at around 1830.

The Dawn of Geology

Thinking about the Earth first occurred when man, faced with natural phenomena such as earthquakes and volcanoes, posed questions about such phenomena and sought to provide answers in naturalistic terms. Practical matters, such as the task of prospecting for mineral resources, also stimulated interest in the Earth.

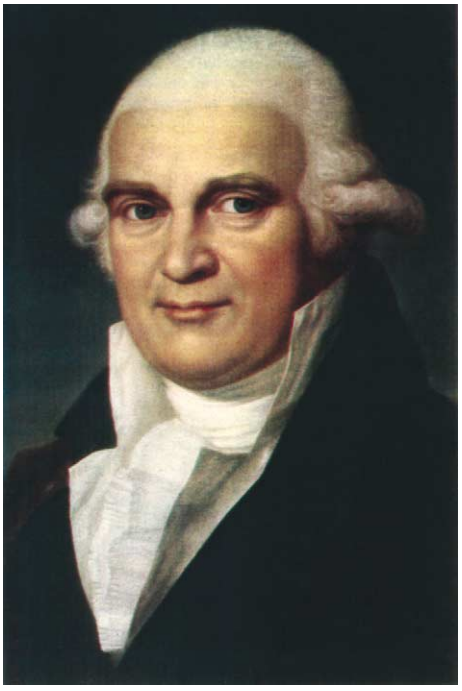


Figure 1 Abraham Gottlob Werner (1749–1817). Father of Neptunism.

The Greek philosophers thought that the universe was governed by unchanging principles and with intelligible and discoverable natural laws. This contrasted with the mythopoeic or magical explanations of nature found more generally in the ancient world and in non-scientific cultures today.

In his *Histories*, Herodotus of Halicarnassus (Ca 484–425 BC) spoke of the sedimentary loads of the Nile and of the slow growth of its delta. This was perhaps the first recorded statement based on observation indicating an awareness of the magnitude of geological time. But myth and naturalistic explanation were intertwined in Greek thought.

Plato (427 or 428–348 BC) (in *Phaedo*, 111–112) described the Earth as having internal passages carrying "a vast tide of water, and huge subterranean streams of perennial rivers, and springs hot and cold, and a great fire, and great rivers of fire, and streams of liquid mud, thick or thin", as well as a great internal chasm, Tartarus. Water moved with a 'see-saw' motion within the Earth, like the tides, and produced springs that fed rivers and streams, and returned to the sea and thence to Tartarus. The surging waters also generated great winds inside the Earth. Volcanoes were produced by the escape of rivers of fire from within. These speculations were naturalistic, but also explicitly said by Socrates to be 'myth'. Such ideas were to endure until the eighteenth century.

Along similar lines Aristotle (384–322 BC), a pupil of Plato, suggested that earthquakes were caused by subterranean winds passing through cavities within the Earth. Fossils were nature's failed attempts in the creation of living beings (the theory of *vis plastica*).

Although some authors consider Theophrastus of Ephesus (*ca* 371–*ca* 287 BC) to have written the first mineralogical treatise, *Perilithon*, there are references to a work, now lost, written by his teacher Aristotle. In the surviving *Meteorologica*, Aristotle ascribed the origin of minerals and metals to dry/smoky or moist/vaporous exhalations from within the Earth.

Minerals' curative purposes were considered in Dioscorides' *De materia medica* (*ca* 77 AD). Processes such as saline crystallization or exfoliation were remarked upon, and origins of substances of supposed medicinal value were mentioned. This medical tradition was to continue in the attempted mineral/chemical cures advocated by Paracelsus in the sixteenth century.

The Romans were less interested in abstract knowledge than were the Hellenes, but were practical, and skilled in the use of stone for building. The most notable Latin 'scientific' text was Pliny the Elder's *Historia naturalis* (first century), consisting of 37 books, uncritically compiled from 2000 works of antiquity. The last five books dealt with the mineral kingdom, with mining and smelting practices, and with the characters, occurrences, and uses of many mineral substances.

On observing sea shells in the mountains, Ovid (43 BC–17 or 18 AD) inferred that those lands had formerly been covered by the sea. He also realised how fluvial valleys could be formed and how water gradually reduced relief. The materials swept along would be deposited, lower down, in flooded areas, where on drying and hardening they would become rocks. We have for the first time the pattern: erosion, transportation, sedimentation, and lithification.

The idea of the regeneration of minerals and ores in mines was advanced by Pliny's teacher, Papirio Fabiano, an idea still maintained in the seventeenth century, as in the case of Alvaro Alonso Barba in *El arte de los metales* (1637). In this work, the Earth supposedly had the ability to 'reproduce', as envisaged in antiquity.

During the Dark and Middle Ages, Aristotle's influence continued in the West, but linked with Christian viewpoints. Thus, for example, St Isidore of Seville's (560–636), in *Etymologies* (a work considered to be the first encyclopedia), pointed to the organic origin of fossils, but connected them with the Flood.

Alchemy coming from Persia (eighth–nineth centuries) influenced the works of Ibn Sina (Avicenna) (930–1037) and subsequently Christian authors like Alfonso X (1221–1284), Raymond Lully (1235–1315), Arnaldo Vilanova (*ca* 1238–1311), Ulisse Aldrovandi (1527–1605), Andreas Libavius (1560–1616), and Alonso Barba (1569–1662). There developed the so-called theory of the opposites whereby things combined or repelled one another according to their 'sympathies' or 'antipathies'. Some spoke of the gender of minerals. For example, the word 'arsenic' derives from the Greek word for male. Minerals supposedly formed from the appropiate combinations. Alchemy was the forerunner of inorganic chemistry.

Another feature of the Middle Ages was the proliferation of 'lapidaries': list of stones, etc., with descriptions of their properties, uses, etc. Ibn Sina wrote *De lapidibus*, in which minerals were classified according to the quadrichotomy: stones/earths; metals ('fusibles'); sulphurous fossils (combustibles); salts ('solubles'). Ahmad Al Biruni (973–after 1050) mentioned more than 100 minerals and metals in his treatise on gems (*Kitab-al-Jamahir*), and accurately determined specific gravities for several types. Also Alfonso X of Castile (1221–1284) (Alfonso the Wise), translated numerous Arab lapidaries, where the properties of minerals supposedly varied according to the positions of the heavenly bodies. Al Biruni, born in Uzbekistan and a great traveller, was also notable for his studies of rivers. He recorded evidence for

changes in the course of the Amu Darya River, and the decrease of sediment size down the Ganges. So (anachronistically) he could be called a fluvial geomorphologist.

The greatest Mediaeval author on the mineral kingdom was Albertus Magnus (St Albert of Cologne), Bishop of Ratisbon and doctor of the Church (1193–1280). Anticipating Renaissance authors, he stated that experience alone was the source of knowledge of physical things. He tried to link faith and reason when he pointed out that the sea could never have covered the whole Earth by natural causes. In *De mineralibus* he recognised about 100 mineral species. Both minerals and rocks were thought to have formed from molten masses.

The First Geological Principles: The Observation Phase

With the Renaissance, the geocentric Aristotelian and Thomist universe collapsed in the 'Copernican Revolution', and observation rather than 'authority' became central to science. For example, Bernard Palissy (1510–1590), pointed out "I have never had any other books than the skies and the Earth whose pages are open to all". Systematic ordering of the observations facilitated the establishment of the first geological principles. Information also spread faster thanks to the printing press. This was particularly true of great natural catastrophes, such as the eruption of Vesuvius in 1538, which prompted interest in the Earth.

Leonardo da Vinci (1452–1519) visited the Alps and realised that the geological structure was the same on both sides of the fluvial valleys. The rivers carried away materials to the sea, where they might bury shells. When land rose up it formed hills that, on being cut by rivers, reveal layers or strata. Shells in such strata were not carried there by the Flood. We have one of the first visions, albeit incomplete, of the geological cycle.

Palissy showed that what are today called rudist lamellibranchs are 'lost' species. This recognition of extinction was an important contribution towards recognition of the Earth's antiquity. Interest in fossils developed little by little, as when Father Jeronimo Feijoo y Montenegro (1676–1764) also cited discoveries of lost species. There was still a long way to go to before fossils were used to determine the relative chronology of the landscape.

In his *Principia philosophiae* (1644) the French philosopher René Descartes (1596–1650) considered the Earth as an old cooling star. There was incandescent material in its interior, around which there was a layered structure (metallic, heavy material, air–water, and outer crust). As the globe cooled, the crust cracked and collapsed, thus creating mountains and seas. In this speculative theory we have the first attempt to explain the internal structure of the eEarth in mechanical terms (i.e., in terms of the 'mechanical philosophy' according to which all natural phenomena were explained in terms of matter and motion). Descartes also saw the planet as a great 'still', heated by its internal material. So sea water penetrating into the Earth was distilled in the interior, leaving the salt there.

Descartes' theory of a central heat re-appeared in the work of the Jesuit Athanasius Kircher (1601–1680), *Mundus subterraneus* (1665), which proposed a great central *pyrophylacium* or repository of heat, linked also with ideas going back to Plato. The main repository was connected by channels to other lesser fires, and the network of interconnected channels served as conduits for volcanoes at various places on the surface. In addition to the *pyrophylacium*, there were *aerophylacia*, through which circulated the subterranean winds that supposedly caused earthquakes; and *hydrophylacia*, or water-containing caverns, which were fed from the sea and sustained springs. (The model had similarities to that in Plato's *Phaedo*.) Earthquakes gave rise to the formation of mountains. Kircher also revived the organicist theories, speaking about the uterus of the globe, and *vis petrifica* and *vis seminalis* (petrifying and seminal powers). The Earth was a living organism with a capacity for reproduction and the other functions of a living being (so inside the Earth salt water becomes fresh through a quasi-'metabolic' process). Thereby both external and internal 'geodynamic' phenomena were explained.

Niels Stensen (1638–1686) (Nicolaus Stenonis or Steno) (*see* **Famous Geologists:** Steno), a Danish physician in the service of the Medici family in Florence, was less speculative and more original. He authored *De solido intra solidum de naturaliter contento dissertationis prodromus* (1669), in which, from the study of quartz crystals, the law of the constancy of interfacial angles was first recognized. With Steno, we also have what might be called the first 'stratigraphic diagram'. Sediments accumulate, forming horizontal layers in which marine or terrestrial fossils were buried, the oldest layers being below and the younger ones above. These layers could be undercut by erosion, fracturing and collapsing. Then new horizontal layers were deposited, at an angle to the earlier ones. One of geology's main problems, to establish a chronological order of events, had begun to be resolved. (Steno's principle of superposition was relatively trivial: the lowest layer of bricks in a wall is put in place before the upper ones. But it required imagination to apply this idea to the easily observed layered rocks.)

But stratigraphy had still to be put together. Giovanni Arduino (1714–1795) made the first chronostratigraphic division (with geological plan and section included) when he divided the rocks of the Alpine landscape into: 1) Primary: formed by quartzites, and slates; 2) Secondary: formed by limestones, sandstones, and shales; 3) Tertiary: formed by limestones, sandstones, gypsums, and clays; and 4) Alluvium. The idea was set forth in two letters addressed to Antonio Vallisnieri (published 1760).

The seventeenth century was also characterized in the West by attempts to reconcile observations of natural history with the Bible, aligning 'faith and reason', in what was called 'physico-theology', or the attempted interweaving of natural philosophy (science) and religion. Such work continued well into the nineteenth century, and even to the present.

Thomas Burnet's (1636–1715) work *Telluris theoria sacra* (1681) provides a good example. For this Anglican cleric, the Earth's initial chaotic material was ordered by gravity, with the heaviest parts in the centre and the lightest parts at the surface. The result was a concentric structure: 1) a Kircherian or Cartesian igneous core; 2) liquid; 3) an oily layer; and 4) an outer crust hardened by the sun (*ossatura telluris montium*). When the central vapours acted on the outer crust it cracked and broke, giving rise to the Flood ("all the fountains of the great deep [were] broken up": Gen. 7, 11). If the Earth had not been flat, it could not have been covered by the waters (here reason and design were introduced). After the Flood, the waters supposedly withdrew, taking much with them, thus causing the Earth's relief. Such theories, connecting the Noachian Flood with geological observations, came to be called 'diluvialist'. An antecedent of 'classical' diluvialism was perhaps the Spaniard José González Salas (1588–1651) who, in 1650, stated that the Flood changed the face of the Earth.

For Isaac Newton's successor at Cambridge, William Whiston (1671–1752), a comet caused water escape from the Earth's interior, while for John Woodward (1665–1728), the waters supposedly dissolved the Earth, which was then converted to its present layered state as matter separated out according to the law of gravity (*Essay Toward a Natural History of the Earth*, 1695).

The age of the Earth was calculated in accordance with the biblical records, as Alfonso X the Wise had done in his *General History*. This was likewise done by the Cambridge classicist John Lightfoot (1642–1644); the Anglican Primate of Ireland, James Ussher (1650); and William Lloyd, Bishop of Worcester (1701). They arrived at various values between 3928 and 5199 years old. These authors, who today may seem detached from reality, were in fact careful scholars who sought to reconcile Jewish, Christian, and pagan historical records. Ussher's date for the Earth's creation (4004 BC) became the best known, as printed in the margins in Lloyd's edition of the 'King James' Bible (1701).

Outside the religious arena, Gottfried Liebniz's (1646–1716) *Protogea* (1684) proposed that rocks had been formed by two processes: 1) the cooling of fused material to give an Earth with a 'glassy' surface; 2) the action of waters on this hard surface and the concretion of solid elements contained in aqueous solution. Leibniz's ideas thus anticipated the late 18th century debate between 'Plutonists/Vulcanists' and 'Neptunists'.

José Vicente del Olmo (1611–1699), in his *New Description of the Orb* (1681), stated that the mountains were raised up due to internal exhalations. The elevated areas were then eroded by rain, wind, and river floods. Thus a balance was established in nature, rather than a single progression of change such as Burnet envisaged. But let it be remembered that Seneca (*ca* 3 BC–65 AD) and the Epicureans had long before envisaged a 'balance of nature'.

Geology as a Science is Born

The eighteenth century was characterized by the economic development of the western nations and the development of democratic ideals. Inspired by the accomplishments of science and technology, the Enlightenment world-view, which saw things as essentially intelligible with problems being capable of solution by rational beings with minds unclouded by superstition, was to be driven on by the idea of progress, which was born with it. New centres of teaching such as Göttingen University were founded, and scientific publications for the technical and educated bourgeoisie, such as Denis Diderot's *Encyclopédie*, appeared. Experimentation also acquired greater importance, even in the study of the Earth. It was a period of glorification of the rational, where it seemed that the only things that mattered were those that could be counted, measured, weighed, or rationally calculated.

The end of the eighteenth century coincided with the Industrial Revolution (England) whose foundations were iron, coal, steam, and textile manufactures. The need for additional natural resources boosted mining, and between 1766 and 1788 the mining academies of Freiberg, Chemnitz (now Banksa Stiavnica), St Petersburg, Almaden, and Paris were founded in that order. 'Subterranean geometry' and mineralogy were taught and mineralogy began to develop into petrography, stratigraphy, palaeontology (later), and, eventually, geology (around the

end of the eighteenth century). Curiously, Britain was backward in such centralized technical education.

Abraham Gottlob Werner (1750–1817) was appointed to the Mining Academy of Freiberg in Saxony in 1775 where he developed his Neptunist theory (1777), which proposed that all rocks, even basalts, were formed by chemical precipitation from a primordial ocean or *allgemeines Gewaesser*. (There were, however, precursors of this theory, such as the Frenchman Benôit de Maillet [1755].) According to Werner, by successive sedimentation onto an irregular terrestrial core, four types of formations were supposed to be deposited: 1) Primitive: crystalline rocks such as granite and gneiss; 2) Transitional: limestones, slates and quartzites; 3) *Flöetz*: formed from what we consider today to be the layered rocks from the Permian to the Cenozoic; and 4) Alluvial: (superficial) deposits. (The 'Transition' category was absent from initial exposition of Werner's theory.) These 'chronostratigraphic' divisions had previously been adumbrated in Germany by others such as Johann Lehmann (1719–1767) and Georg Christian Füchsel (1722–1773), and also by the German traveller, Peter Simon Pallas (1741–1811), in the Urals (1768). The Primitive formations would be found in the central parts of mountain ranges, from which the water would have withdrawn first.

Werner's theory gave an approximation to the order of rocks observed in the field. But there were questions that Werner's theories couldn't solve:

1. Where did the water of the supposed primordial ocean go to?
2. Is the Earth inactive? (For Werner, sloping strata corresponded to margin sedimentation.)
3. How were rocks such as basalt, found on the tops of hills, to be explained? (Werner thought that basalt was also precipitated from his ocean, the level of which supposedly rose again for some unexplained reason.)
4. How were mineral veins and dykes to be accounted for? (Werner thought that material might have precipitated from above filling rents in the crust.)
5. How were volcanoes to be explained? (Werner thought that they might be due to the combustion of subterranean coals, etc.)
6. How could the universal ocean dissolve so much siliceous matter? (This question was never answered satisfactorily, though the occurrence of siliceous springs and quartz veins in some rocks suggested precipitation from solution.)

Not everything that came from Werner was wrong. He praised observation and the use of scientific method and he assisted into the emergence of geognosy or geology, 'oryctognosy' or mineralogy. With Werner (and before him in Russia with Lomonosov [1711–1765]), mineralogy acquired its own body of doctrine. He classified minerals according to their external characteristics (physical properties), as Linnaeus (another important Enlightenment figure) had done with plants and animals, between 1735 and 1760. The observation of crystalline forms was to lead to the birth of crystallography. This science had taken its first steps beyond Steno thanks to the Swiss naturalist Moritz Anton Capeller (1685–1769) with his *Prodomus cristallographie* (1723) and the Frenchman Jean Baptiste Louis Romé de l'Isle's (1736–1790) *Essai de cristallographie* (1772), soon to be developed further by René-Juste Haüy.

Werner was to have many disciples who would write important pages in the annals of geology during the nineteenth century, such as Guyton de Morveau (1737–1816), Horace Bénédict de Saussure (1740–1799), Déodat Gratet de Dolomieu (1750–1801), Juan José Elhuyar (1754–1896), Fausto Elhuyar (1755–1833), Andrés Manuel del Río (1765–1849), Alexander von Humboldt (1769–1859), Leopold von Buch (1774–1853), Robert Jameson (1774–1854), etc. They tried to use his stratigraphic order, worked out in Saxony, as a 'paradigm' for examining and interpreting rocks in other parts of the world.

Another notable eighteenth-century authority was the keeper of the *Jardin des Plantes* in Paris: Georges Louis Leclerc, Comte de Buffon (1707–1788), author of a great 36-volume *Histoire naturelle*. In a supplement of this work entitled *Époques de la nature* (1778) he put forward three basic ideas: 1) a longer duration of geological time (compared with the Biblical account); 2) organic evolution, preparing the way for transmutationism and evolutionism; and 3) palaeogeography.

Like Descartes, Buffon thought that mountains were formed by contraction during the Earth's cooling. He also examined the problem of the age of the Earth *experimentally*, heating spheres of different sizes and measuring how long they took to cool until they could be touched; and by analogy he estimated the possible age of the Earth. He arrived at the conclusion that it would have taken 74 832 years to have cooled to its present temperature (and privately speculated on the possibility of a much greater age).

Through further experimentation Buffon obtained silicates by melting clays. Nevertheless, he held to some older ideas, such as the view that earthquakes were caused by explosions of gases in the Earth's cavities or that volcanoes were produced by the combustion of sulphur and bitumen.

The hydrological cycle was also quantified, in accordance with the calculations of Edmé Marriotte

(1690), Pierre Perrault (1674), and the suggestion of Edmund Halley (1714–1716) that one could measure the rate of increase of salinity in lakes that had no discharge rivers, and then gauge how long it might have taken for the oceans to acquire their salinity. Nevertheless, eighteenth-century geology was not obviously an experimental science. It was not then possible to reproduce variables such as pressure, temperature, or time, which reach very high values in many of the processes occurring in nature.

In the eighteenth century there were still authors who regarded fossils as 'figured stones', as did the French physician Pierre Barrère, author of *Observations sur l'origine et la formation des pierres figurées* (1746). The diluvialist school was also active. Thus, for the Spaniard, Father Antonio Torrubia, in his *Aparato para la historia natural* (1754), fossils were represented as remains of the Flood. Nevertheless, a significant rejection of diluvialism occurred in the mid-eighteenth century, mainly in central Europe, with authors, such as the Göttingen professor Samuel Christian Hollmann (1753), while the Swiss cleric Johan Georg Sulzer (1762), pointed out the marine origin of fossils. Numerous examples were described and the natural history cabinets were filled with specimens, but without an agreed system for their cataloguing. Although the influential Werner rejected fossils as the basis for the study of stratification, they began to gain in importance, and increased knowledge began to pave the way for the birth of stratigraphy, at the end of the century, and of scientific palaeontology, which entered at the end of the eighteenth century.

Practical matters were also important in the Enlightenment. Between 1778 and 1782, Jean Étienne Guettard (1715–1786) and Inspector General of Mines Antoine Grimoald Monet (1734–1817) jointly published their *Atlas minéralogique de la France*, which showed the distribution of deposits of economic significance across their country. In Sweden, the chemist Torbern Bergman (1777) initiated general methods of mineral analysis in the 'humid' way, bringing mineral substances into solution by the action of acids or alkalis and then identifying components by a sequence of precipitation reactions. Prospecting for coal was enhanced by boring techniques, but without palaeontological control the results were not always useful through misidentification of strata.

The great catastrophe of the Lisbon earthquake on 1 November, 1775, sowed pessimism in the scientific world. There were many, including Buffon, who thought of the progressive degradation of the cooling globe. But in Spain, the naturalist Brother Benito Feijoo y Montenegro, in his *Cartas eruditas y curiosas* (1760), tried to calm things down by pointing out the greater the force of the previous one and that repetitions of earthquakes are less likely. In Germany, Immanuel Kant argued that earthquakes had natural causes and had nothing to do with the moral condition of mankind. But they could remind us not to try to find happiness in worldly goods. Old earthquake myths endured nevertheless, and it was only at the end of the nineteenth century that geologists began to suspect the main causes of tremors.

See Also

Biblical Geology. **Famous Geologists:** Hutton; Steno. **Geomythology**. **History of Geology From 1780 To 1835**. **Minerals:** Definition and Classification. **Stratigraphical Principles**.

Further Reading

Ellenberger F (1988) *Histoire de la Geologie Tome 1 Des Anciens à la première moitié du XVIIe siècle*. Paris: Technique et Documentation – Lavoisier.

Ellenberger F (1994) *Histoire de la Geologie Tome 2 La grande éclosion et ses prémices 1660–1810*. Paris, London, and New York: Technique et Documentation – Lavoisier.

Faul H and Faul C (1983) *It Began with a Stone: A History of Geology from the Stone Age to Age of Plate Tectonics*. New York, Chichester, Brisbane, Toronto and Singapore: John Wiley & Sons.

Gaudant G and Bouillet G (2000) La genèse et l'interprétation des 'fossiles' dans la science classique: de la Renaissance aux Lumières. *Bulletin de la Société Géologique de France* 171: 587–601.

López Azcona JM (1985) Los jheólogos. *Revista de Materiales y Procesos Geológicos* 3: 179–187.

Mather KF and Mason SL (1939) *A Source Book in Geology*. New York and London: McGraw-Hill.

Oldroyd DR (1996) *Thinking about the Earth: A History of Ideas in Geology*. Cambridge (Mass): Harvard University Press.

Pelayo F (1996) *Del Diluvio al megaterio. Los orígenes de la paleontología en España*. Madrid: Consejo Superior de Investigaciones Ciéntificas.

Rappaport R (1997) *When Geologists were Historians, 1665–1750*. Ithaca and London: Cornell University Press.

Rossi P (1984) *The Dark Abyss of Time: The History of the Earth & the History of Nations from Hooke to Vico*. Chicago and London: The University of Chicago Press.

Sequeiros L (2003) Las raíces de la Geologia: Nicolas Steno, los estratos y el diluvio universal. *Enseñanza de las Ciencias de la Tierra* (10,3): 217–242.

Wagenbreth O (1999) *Geschichte der Geologie in Deutschland*. Stuttgart: Georg Thieme Verlag.

Wendell E Wilson (1994) The history of mineral collecting. *Mineralogical record* 25(6): 1530–1799.

HISTORY OF GEOLOGY FROM 1780 TO 1835

D R Oldroyd, University of New South Wales, Sydney, Australia

Introduction

The years 1780–1835 mark the period when geology emerged as a science *sui generis*, distinct from mineralogy. It began to have an institutional base in the form of scientific societies and began to be taught at universities (though there was earlier tuition in cognate subjects at mining academies in Europe). 1835 saw the establishment in England of the world's first national geological survey (though there were a few earlier 'private' surveys, or national 'mineral surveys'). The period concerning us here has been called the 'hinge of history', with its shift from the eighteenth century Enlightenment to the Romantic Movement, the Gothic Revival, and extensive industrialization. It was also a time of political upheaval, with the French Revolution and the Napoleonic Wars. There was a great expansion of European horizons, both in space and time. It included notable explorations, such as those of Flinders and Baudin around Australia, d'Orbigny and von Humboldt in South America, and Darwin's *Beagle* voyage. America was extending westwards, and Russians were consolidating their hold in Siberia. Much of India was under British rule and exploration was getting underway in Africa. These explorations had economic and military imperatives, but were also of great scientific significance. By 1835, the bounds of time of earlier theological constraints were truly burst by geologists, and during this 'hinge period' the Earth began to be seen as an object whose history could be revealed by empirical examination of rocks, fossils, and strata. The cultural movement known as 'historicism' came to the fore, according to which things (like the principles of law) could supposedly be best understood by studying their history. In this sense, geology fitted into the 'spirit of the times'. Understanding of the Earth became 'historicized'.

That is one way of looking at the period of geology's emergence. But geology can also be seen as having emerged within the context of the Industrial and Agricultural Revolutions: it was the practical men concerned with mining, quarrying, surveying, agriculture, etc., who were responsible for many of the discoveries on which rested the intellectual achievements of the 'geological elite', who established learned societies and developed high theory. The 'little men', on this view, did as much to found the science as did the 'gentlemanly geologists' whose writings have since become well known to historians.

The period 1780–1835 is also interesting for the appearance of important rival geological theories, and the contests between supporters of the competing doctrines. The so-called Neptunist and Vulcanist theories, associated particularly with the names of Abraham Gottlob Werner and James Hutton (*see* **History of Geology Up To 1780**, **Famous Geologists:** Hutton), offered radically different geological theories and effectively functioned as competing paradigms. Likewise, there were strong differences of opinion (amounting to substantial philosophical differences), summarized by the terms 'catastrophism' and 'uniformitarianism'. Debates on these matters ranged into the issue of the age of the Earth; the question of whether geological discoveries could or could not be reconciled with, or support, theological beliefs; and questions about the appropriate methodology for the geosciences. Rocks and minerals were first clearly distinguished by Alexandre Brongniart (1827).

The 'Little Men' and a Geological Map

The name of William Smith has long been remembered for his discovery that different strata could be recognized and discriminated by means of their fossil contents (*see* **Famous Geologists:** Smith). Smith, who came from south-west England, was a surveyor and engineer who worked on road and canal projects, land drainage schemes, coal and ore prospecting, etc. Living near Bath, he came to realize that there were two kinds of oolitic limestones with similar appearances but different fossils. Further afield, chalks and greensands showed similar distinctions. Accordingly, as his work gave him ample opportunity for observing and collecting, he began to assemble and arrange fossils according to what we would call their stratigraphic horizons. He drew up a table (1799) listing the lithologically different strata of southern England (from coal to chalk), along with their characteristic fossils. He then embarked on the huge project of trying to identify and enter on a coloured map the strata of the whole of southern Britain. This single-handed, hand-painted, map was issued in 1815. But the intellectual elite of the Geological Society (founded 1807) hardly gave credit where credit was due, and set about making their own map, which was issued in

1819, making some use of Smith's results. But in 1831 Smith was dubbed the 'Father of English Geology' by Adam Sedgwick (*see* **Famous Geologists:** Sedgwick).

Apart from his map (and sections), Smith should be recognized for his practical and economic contributions. On the basis of his biostratigraphic knowledge, he knew where coal might or might not be expected to be found and could advise entrepreneurs accordingly. He was not much interested in *why* the rocks and fossils were distributed as they were or how old they were. Smith was only one of the many practical men who contributed to the emergence of geology, but most of them left few historical traces and are little known. There were also skilled artisan fossil collectors, the best known of whom was Mary Anning working on the south coast of England, who extracted specimens of ichthyosaurs and lesser fossils for sale to connoisseurs and museums.

Neptunism

In Germany, lithostratigraphy was chiefly deployed during the late eighteenth century, according to the system of Abraham Werner, which was taught at the Freiberg Mining Academy in Saxony. It postulated the existence of a standard 'formation sequence' of rocks: Primitive rocks (especially granite); Transition rocks (schists, greywackes, etc.); *Flöetz* (layered) rocks; Basalt; and Alluvial deposits. His observational work was chiefly undertaken in Saxony and Silesia, and through the influence of his many distinguished pupils this area of Europe came to be regarded as an exemplar (paradigm) for studies further afield. Werner was a gifted mineralogist and teacher, but his lithostratigraphic theory was based on the misconception that vast amounts of rock could formerly have been dissolved in a hypothetical universal ocean. He thus denied the igneous origin of rocks like granite or basalt, and because basalt caps were found overlying sand beds on hills near Freiberg he imagined that basalt could have been deposited from water (an idea perhaps encouraged by its frequent columnar jointing). This meant that the supposed mineral-bearing ocean would have had to have risen to deposit the hilltop basalts. So the theory of a falling universal ocean was implausible, both chemically and physically. Nevertheless, Werner's mineralogical exactitude and lithostratigraphy were appealing and influential.

But already in the eighteenth century visitors to central France such as Jean-Étienne Guettard had recognized the existence there of former volcanoes, and their associated basalts were interpreted as igneous products by Nicholas Desmarest, who deduced the historic sequence of lava flows in the area, the older ones being substantially eroded while the younger ones were fresh and had run down previously formed valleys. Pupils of Werner such as Leopold von Buch started off their careers as Neptunists, but turned against their master's ideas when they reached the Auvergne.

From about 1820, there were few who still advocated Neptunism, but later in the nineteenth century there was renewed interest in the role of water in the formation of rocks, and again in the migmatist/magmatist debates of the twentieth century: Werner's ideas had intellectual descendants. In Edinburgh, they had a strong advocate in the Professor of Natural History, Robert Jameson.

Vulcanism

Vulcanism (to be distinguished from the modern term volcanism) was the sobriquet for the theory that ascribed primacy to the agency of heat in geological processes. Such ideas were strongly argued by the Scottish Enlightenment geologist, James Hutton. As a farmer, Hutton saw soil being washed into the sea, and he supposed that if weathering and erosion continued indefinitely all the fertile land would be lost, and what he regarded as a world 'devised in wisdom' for human habitation would eventually be destroyed (he took a very long view of geological history). He therefore contemplated a cyclic theory that would allow for replenishment of soil.

In Werner's theory, volcanoes were 'weakly' explained by the combustion of underground coal deposits. But Hutton advocated a great central source of heat, in a way similar to Kircher's '*Pyrophylaciorum*' (*see* **History of Geology Up To 1780**), and analogous to the fires of the machines of the Industrial Revolution, which made the wheels of industry turn. Hutton envisaged the Earth as a kind of great machine. After his death, his 'Plutonist' theory achieved support from the experimental work of Sir James Hall, and from measurements of temperature gradients within mines, such as those made by Louis Cordier, which showed increased temperatures with depth supporting the idea of the Earth having a hot interior.

The Vulcanist–Neptunist Dispute

This controversy was prosecuted with greatest vigour in Edinburgh, between supporters of Hutton and Jameson. From today's perspective, it would seem that Vulcanism would be the easy winner, but the results of Hall's attempted experimental vindication of Hutton's doctrine were somewhat ambiguous. When granite was fused and slowly cooled the product did not look exactly like the starting material.

It is interesting to remark that a place like the Isle of Arran, near Glasgow, had exposures that were broadly compatible with both the Huttonian and Wernerian theories (see Figure 1). But unconformities such as the famous one examined by Hutton at Siccar Point (see Figure 2) could only be well explained in Huttonian terms. And granitic veins, such as those famously examined by Hutton at Glen Tilt, and anastomosing mineral veins, could hardly be accounted for by the Wernerian theory.

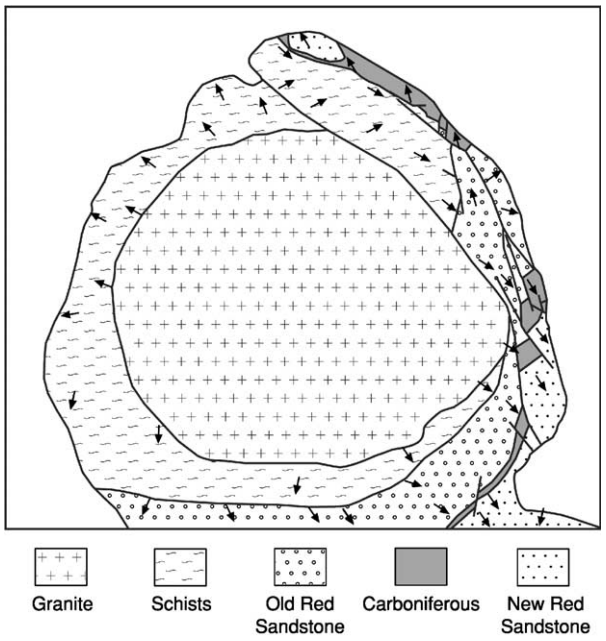


Figure 1 Sketch of the geology of the northern part of Arran. The exposures can be accounted for in terms of either the deposition of 'transitional' rocks (schists) and *Floëtz* sediments round a central granitic core; or the intrusion of granite into sediments, forming a dome structure with partial metamorphism round the granite.

Reconstruction of Past Environments

The chemist Antoine Lavoisier accompanied Guettard during his travels for the purposes of compiling a mineralogical map of France and acquired some knowledge of geology. Lavoisier became interested in the conditions that might be expected to occur at the bottom of the sea, at different depths and different distances from the shore, with cobbles near the edge of the sea, sand further out, and fine sediment some distance offshore. An idealized profile was published (1789) showing the occurrence of pelagic and littoral deposits. Such different sediments could also be seen in the strata observed inland, and the analogies were clearly understood. Thus we have the beginnings of attempts to reconstruct past environments on the basis of what was visible in the strata. Work of a similar kind was undertaken in the Paris basin in the early nineteenth century by Alexandre Brongniart and Georges Cuvier (see below). They recognized marine and fresh-water sediments on the basis of the fossil shells they contained (fresh-water forms generally having thinner shells). The arguments were based on analogies made with modern forms. Later, Charles Lyell (see below) realized that the processes of lime formation in freshwater lakes in his property in Scotland resembled similar depositions occurring in modern France. Examination of coal deposits suggested hot and humid swamp conditions. Or desert sandstones could be recognized for what they were. The beginnings of palaeoecology occurred during the 'hinge of history'.

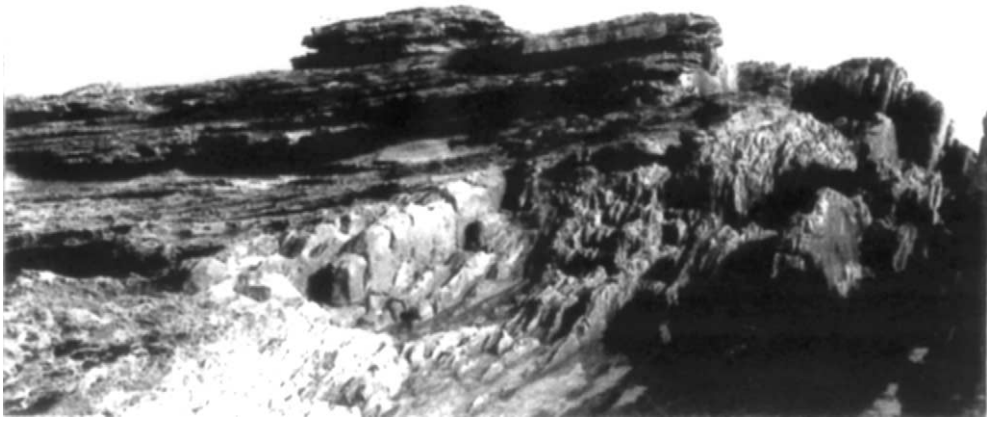


Figure 2 Hutton's Unconformity at Siccar Point, south-east of Edinburgh. 'Old Red Sandstone' with basal conglomerate lying almost horizontally on upright 'schistus' (Silurian greywackes). After Secord JA, *Controversy in Victorian Geology*. Copyright © 1986 by P. U. Press. Reprinted by permission of Princeton University Press.

Palaeontology and Comparative Anatomy

Organic fossils had been known for centuries and by the eighteenth century few doubted that they were former living organisms. Fossil collecting became a popular, and even lucrative, pastime, especially when large specimens of ammonites, ichthyosaurs, ancient fishes, mammoths, etc., were discovered. Many collections, in private or public hands, were initiated. It was Georges Cuvier, Professor of Zoology at the *Muséum d'Histoire Naturelle* in Paris, who began to establish some fundamental principles for the study of fossils (*see* **Famous Geologists:** Cuvier). He started by trying to reassemble collections of mastodon bones from America. They were manifestly creatures rather similar to modern elephants, so he used the known arrangement of elephant bones for purposes of comparison, in order to reconstruct the mastodon skeleton. In this work, he recognized that the total structure must be one in which all the parts fit together to form a 'working whole'; and the structures must be such as to enable the creatures to live successfully in their 'conditions of existence'. This 'goodness of fit' necessarily applied to the lost soft parts as well as those that had become fossilised. So Cuvier enunciated his principle of 'co-ordination of parts'.

However, most fossils were shells, etc., and it was Cuvier's colleague Jean-Baptiste Lamarck who was responsible for collecting and studying these. He was the first to make a formal distinction between vertebrates and invertebrates and state the major divisions of the latter. Only then could invertebrate palaeontology make a proper start. It was less easy to make comparative anatomical studies for fossil invertebrates than for vertebrates, though Charles Darwin (*see* **Famous Geologists:** Darwin) later famously did so with barnacles.

In examining the fossil record, Cuvier found that there were substantial and apparently sudden changes in the fossil record as one ascended the stratigraphic column. This finding largely holds to this day for macrofossils, though smooth trends are known for foraminifera. The biostratigraphic 'breaks' were, in fact, convenient for stratigraphic investigations and mapwork such as that conducted by William Smith, but their theoretical explanation was difficult, using modern analogies as Cuvier had done for his mastodon work. Cuvier's explanation was couched in terms of sudden and catastrophic events (which we might call mega-tsunamis), of unknown cause, perhaps of global extent (Cuvier was inconsistent in his statements as to the universality or otherwise of his catastrophes). Famously, he wrote that in these supposed events "the thread of operations is here broken; the march of Nature is changed". His doctrine was later dubbed 'catastrophism' by William Whewell, and there was the suggestion that the very laws of nature were broken or suspended during such catastrophes; or past events were quite unlike anything occurring today.

Seventeenth- and eighteenth-century writers such as Steno (*see* **Famous Geologists:** Steno) had sometimes propounded 'catastrophist' theories of the Earth because they had a greatly compressed timescale for Earth history, supposedly in line with the 'Biblical' age of the Earth, of some 6000 years. This was not Cuvier's view. He wrote: "Genius and science have burst the limits of space;... Would it not also be glorious for man to burst the limits of time, and, by means of observations, to ascertain the history of this world, and the succession of events which preceded the birth of the human race?". This was the voice of modern geology. His theory envisaged geological direction and progress, taking place over unspecifiably great periods of time. However, he was unable to account for the origin of new living forms (unlike Lamarck, he was no evolutionist). He thought there was a catastrophe some 5000–6000 years ago.

Geology and Religion/Theology

It is easy to overplay the connections between geology and religion in the early years of the nineteenth century. But they certainly existed and were interesting and important. One line of approach, taken by the Swiss naturalist Jean-André de Luc, who was greatly concerned to 'harmonise' geology and Biblical history, was to suppose that there had been a great catastrophe, perhaps 5000 years ago, which separated history into antediluvian and postdiluvian eras, the division being associated with the Noachian Flood. The period since the Flood could be estimated by the rate of lake infilling, or thickness of peat accumulations. Such evidences almost allowed the date of the Flood to be quantified empirically. Antediluvian time could be regarded as virtually limitless, the 'days' of Creation being of unspecified duration. De Luc's empirical work was, in fact, leading him towards a date corresponding with what we regard as the end of the last ice age. Nature was playing him a cruel trick!

At Oxford, there had been lectures in geology for many years. The field fell into desuetude in the late eighteenth century but the Reverend William Buckland gave popular lectures there from 1814 to 1849, and particularly in his early years he endeavoured to show linkages between Cuvierian geology and Biblical history, identifying Cuvier's last catastrophe with the Noachian Flood. Buckland specially interested himself in cave deposits, supposing that the animal bones found therein might be the remains of animals that had sheltered in the caves or had been washed

into them. His lectures were popular, in that they seemed to show that Biblical history could find support in the latest findings of science. The convenient occurrence of coals, limestones, and ore bodies supposedly evinced divine wisdom. Regarding the Flood, Buckland believed that it was an historical event of divine origin and humans were of recent origin, as described in Mosaic history (though, following Cuvier, he denied that there were any evidences of fossilized antediluvian humans, such as might be expected if Mosaic history were correct). But, along with this 'natural theology', Buckland was a dedicated scientist, doing much fieldwork. He showed that the marks on his cave bones were compatible with those left by the teeth of modern hyenas. He even let a tortoise walk across pastry, claiming that the marks it made were comparable to those found on sedimentary rocks from Scotland, suggesting that the latter were imprints of an ancient reptilian analogue. This was a kind of use of Cuvier's 'comparative method'. Buckland was one of many who made geology a highly popular science in the early nineteenth century.

Lyell and Uniformitarianism

But Cuvier's and Buckland's geology was an anathema to the followers of Hutton, notably Charles Lyell, in his *Principles of Geology* (1830–1833) (*see* **Famous Geologists:** Lyell). Lyell, a lawyer turned writer and geologist, envisaged an open-ended time-scale, without direction to geological change. Land could rise and fall in an irregular fashion, with consequent climatic changes, as high land happened to be chiefly near the poles or the equator. With changing conditions of existence, new species would be formed (by an unknown process – perhaps God-caused?) or become extinct, the forms approximately tracking the changing conditions. Hence, the proportions of extant forms would gradually decrease into the past, so that the percentage of present types could be used to provide divisions of Tertiary time. Cuvier's great palaeontological breaks were thought by Lyell to be due the loss of strata by erosion; or not all living forms having been preserved. Lyell was a gradualist and a non-progressionist. He was also dubbed a 'uniformitarian' or 'actualist', i.e., he thought that the proper mode of reasoning in geology was to explain the past by using analogies with the knowable circumstances and constant laws of nature of the present. (Use analogies from knowledge of circumstances as they *actually* are; or use 'the present as the key to the past'—an aphorism not coined by Lyell himself.)

An important example of Lyell's actualist geology is provided by his examination of Etna to argue for the great age of the Earth. He could observe recent eruptions and lava flows and historical records of eruptions. Hence he had an idea of the rate of accumulation of lava and the build-up of the volcano. Knowing the height of the mountain and its approximate rate of formation (*assumed* to be constant), he could estimate its approximate age: some hundreds of thousands of years, which was very ancient in human terms. But the lava flows lay on sediments containing fossils of *recent* appearance, similar to organisms in today's Mediterranean. Therefore, the sediments at the *top* of the stratigraphic column were very *old* in human terms. Therefore, the Earth itself must be exceedingly ancient. In this manner Lyell argued for the Earth's enormous antiquity. Geological time had been 'proved', on the basis of uniformitarian argument. It should be remarked that Cuvier, Buckland, and Lyell all used 'actualist' arguments, even though the first two were 'catastrophists' while Lyell was a 'uniformitarian'. Thus, there was some unity within geological thinking, even though many have thought that the early nineteenth century was divided into two hostile camps (as was undoubtedly true to an extent). Lyell was a religious man, and believed that humans were divinely created. But he had no truck with such ideas as universal catastrophes or the 'thread of operations' being broken.

Mountain Building

Lyell was a Huttonian, and accepted the idea of a hot interior to the Earth, and Hutton's cyclic geohistory. But he did not have much to say about the formation of mountains other than the build-up of lava-cones such as Etna. In 1829–1830, however, an influential theory of mountain formation was published in France by Léonce Élie de Beaumont ('*Recherches sur quelques-unes des revolutions de la surface du globe*'). He envisaged a slowly cooling and contracting Earth, with the contraction of the fluid interior leading to occasional bucklings of the solid external crust (like the wrinkles on the skin of a drying apple, we might say), and hence the formation of linear mountain ranges and ocean trenches. It was suggested that the several mountain ranges, running in different directions, might represent different 'wrinklings' of the crust (or orogenies), occurring at different times. The idea was that parallel mountain ranges might have formed at about the same time. Consequently, one could develop a kind of 'tectonostratigraphy'. The theory was subsequently developed, with the idea of a vast geometrical pattern of foldings eventually forming around the globe. It became especially influential in France, and was the forerunner of tectonic theories subsequently developed in the nineteenth century, such as that of Eduard Suess (*see* **Famous Geologists:** Suess), which envisaged

cooling and contraction as the major cause of mountain formation. In essence, cooling/contraction theories held the field until the arrival of 'drift' and plate tectonic ideas in the twentieth century.

Minerals, Rocks, and Crystals

Werner and his school concentrated on the study of the external features of minerals, and developed elaborate schemes for their description and classification. The late eighteenth century saw the emergence of chemistry successfully applied to minerals. The older methods of pyro-analysis with the help of the blowpipe, though useful in the field, could give little quantitative information. But in Sweden Torbern Bergman (1784) published a general method for the chemical analysis of gems. They could be brought into solution by fusion with alkali and then, by a sequence of precipitation reactions, and heating and weighing the several products, the different constituent 'earths' (silica, magnesia, alumina, lime, etc.) could be ascertained as percentages. Bergman's results were inaccurate, but the principles of his procedure were valid and were soon applied more successfully by chemists such as Richard Kirwan, Nicholas Vauquelin, Martin Klaproth, and Jons Jacob Berzelius, both to minerals and rocks. Aided by Lavoisier's theory of elements as simple substances, obtained as the last terms of chemical analysis, mineralogy had a satisfactory theoretical and practical basis for chemical understanding. But old problems remained. A substance of one chemical composition could have many different mineral forms and substances of similar crystalline form could have numerous different chemical compositions. The question of the best way to characterize mineral species remained contentious. Geology *per se* did not take a great leap forward through the progress in chemical mineralogy before 1830.

In petrology, the distinction between bedding and cleavage was understood by the English geologist Adam Sedgwick by the 1820s, but he probably learnt it from quarrymen. Following Ami Boué (1819), the category of metamorphic rocks was introduced by Lyell (1833): '*altered* stratified' rocks—the alteration being due to heat and pressure. He referred to 'hypogene' (formed-at-depth) rocks, instead of 'primary' or 'primitive', and he divided them into those that were 'unstratified' (plutonic, e.g., granite) and 'stratified' (metamorphic, e.g., gneisses or schists).

In crystallography, the most important contributions came from the Frenchman René-Juste Haüy, the Englishmen William Wollaston, William Whewell, and William Miller, and the German Eilhard Mitscherlich. Haüy (1784, 1801, 1822) supposed that crystals were made up of a small number of fundamental '*molécules intégrantes*' (tetrahedron, triangular prism, and parallelipipedon), which could be revealed by crystal cleavage and the 'conceptual analysis' of crystals. From these starting points, he hypothesized the 'building' of many different crystals forms from similar basic building blocks, according to assumed rules of decrement for the addition of the 'integrant molecules'. His reasoning was in part circular, but it gave intelligibility to crystallography. Haüy's 'integrant molecule' foreshadowed the modern chemical concept of molecule.

Haüy used contact goniometers, which were of limited accuracy. Wollaston (1809) devised the more accurate reflecting goniometer, and its increased precision led him to question Haüy's methods and results. But Whewell (1824), developing Haüy's concepts, was able to use co-ordinate geometry to describe crystals, arriving at the equations $x/h + y/k + z/l = 1$ or $px + qy + rz = m$ to represent crystal faces, all coefficients being integers. The indices p, q, and r are now known as the Miller indices, being reciprocally related to the co-ordinates of a vector perpendicular to the plane of a crystal face. By such analysis, crystallography could become mathematized and quantifiable, while geology remained in an 'historical' and largely qualitative mode. Mitscherlich was responsible for introducing the concepts of isomorphism, dimorphism, and polymorphism, which assisted understanding of the complexities of empirical mineralogy.

Volcano Theory

Chemistry also offered ideas about the Earth's internal heat. With the discovery of the alkali metals by Humphry Davy (1807), the suggestion was made that the heat might be generated by the action of water penetrating into subterranean stores of these metals, sufficient to produce volcanic eruptions. This accorded with the idea that volcanoes might be produced by the expansion of gases within the Earth, causing localized 'swellings' of the crust (theory of 'craters of elevation' as advocated by Alexander von Humboldt and Leopold von Buch). There was extensive controversy concerning this issue, but Lyell's theory of volcanoes being produced by successive accumulation of lava flows (or ash emissions) eventually prevailed. Chemical theories of the Earth's heat gradually declined in the nineteenth century, but improved suggestions were not really forthcoming until the twentieth century.

See Also

Biblical Geology. **Famous Geologists:** Cuvier; Darwin; Darwin; Hutton; Lyell; Sedgwick; Smith; Steno; Suess. **History of Geology Up To 1780**.

Further Reading

Ellenberger F (1994) *Histoire de la géologie Tome 2: La grande éclosion et ses prémices 1660–1810*. London and New York: Technique et Documentation (Lavoisier).

Laudan R (1987) *From Mineralogy to Geology: The Foundations of a Science, 1650–1830*. Chicago & London: The University of Chicago Press.

Lewis CLE and Knell SJ (eds) (2003) *The Age of the Earth: From 4004 BC to AD 2002*. London: The Geological Society.

Oldroyd DR (1979) Historicism and the rise of historical geology. *History of Science* 17: 191–213, 227–257.

Oldroyd DR (1996) *Thinking about the Earth: A History of Ideas in Geology*. London: Athlone Press; Cambridge (Mass): Harvard University Press.

Oldroyd DR (1998) *Sciences of the Earth: Studies in the History of Mineralogy and Geology*. Aldershot, Brooklands, Singapore and Sydney: Ashgate Variorum.

Rudwick MJS (1969) Lyell on Etna, and the antiquity of the earth. In: Schneer CJ (ed.) *Toward a History of Geology*, pp. 288–304. Cambridge (Mass) & London: MIT Press.

Rudwick MJS (1997) *Georges Cuvier, Fossil Bones, and Geological Catastrophes: New Translations & Interpretations of the Primary Texts*. Chicago & London: The University of Chicago Press.

Torrens HS (2002) *The Practice of British Geology 1750–1850*. Aldershot and Burlington: Ashgate Variorum.

HISTORY OF GEOLOGY FROM 1835 TO 1900

D R Oldroyd, University of New South Wales, Sydney, Australia

National Geological Surveys

The year 1835 is important in the history of geology, being the date when the first national survey was initiated in England and Wales by Henry De la Beche. There had been earlier 'private' surveys, such as that of John Macculloch in Scotland and Richard Griffith in Ireland, but De la Beche's was the first national survey to receive direct government funding. So with De la Beche, geology (as opposed to mining engineering) became a paid profession rather than a gentlemanly pastime or learned avocation. His enterprise was possible because by 1835 Britain had a good set of 6-inch 'Ordnance Survey' maps, produced for military purposes. Mapping was extended to Scotland in the 1860s, a Scottish Branch being established there (1867) under Archibald Geikie (*see* **Geological Surveys**).

The early British survey was an almost single-handed effort, but by the 1840s new staff were being taken on and the band of 'hammerers' gradually spread their work across the country from the south-west. By the end of the century, England, Wales, and Ireland had all been covered, and much of Scotland. The objective was to identify different divisions of the stratigraphic column and represent them on cut-up portions of the ordnance maps (field-slips), and then enter the information on full-sized maps, before reduction to 1-inch maps, which were issued hand coloured until the end of the nineteenth century. The survey thus had the task of *standardizing* nomenclature and colour symbolism for the nation's geology. Large collections were amassed and specimens exhibited at the fine Geological Museum in Jermyn Street, London, with which was associated the Royal School of Mines.

By 1900, most American states had established surveys, as had the main British colonies and the leading countries of Europe, Argentina, and Japan. The US Federal Survey was established in 1879. To an extent, stratigraphy worldwide was dominated by the ideas of British geologists, building on the work of William Smith (*see* **Famous Geologists:** Smith). However, there was confusion in nomenclature and difficulty in international correlations, for there was no reason in principle why, if Lyellian geology were correct, the stratigraphic columns should correspond in different parts of the world. The International Geological Congresses, the first of which was held in Paris in 1878, had as one of their main goals the rationalization and co-ordination of international stratigraphic nomenclature. But this project was hardly successful in the nineteenth century.

Stratigraphy

The major subdivisions (eras) of the stratigraphic column (Palaeozoic, Mesozoic, Kainozoic) were proposed by John Phillips (1840). The periods were proposed as follows: Carboniferous (Conybeare/Phillips in 1822); Cretaceous (d'Omalius d'Halloy in 1822); Eocene, Miocene, and Pliocene (Lyell in 1833);

Triassic (Alberti in 1834); Silurian (Murchison in 1835); Cambrian (Sedgwick in 1835); Devonian (Sedgwick and Murchison in 1839); Jurassic (von Buch in 1839); Pleistocene (Lyell in 1839); Permian (Murchison in 1841); Oligocene (von Beyrich in 1854); and Ordovician (Lapworth in 1879). The Precambrian was suggested by Jukes in 1862. The 'Tertiary' (Arduino in 1760) survived for the units Eocene–Pliocene. Some of the units (e.g. Cretaceous) had previously been recognized by their lithologies, but were not formally 'introduced', with palaeontological criteria, before the foregoing dates.

The introduction of several of the Periods involved well-known geological controversies. Notably, there was a battle between the 'professional' De la Beche on the one hand, and the 'amateur' gentlemanly geologists Adam Sedgwick (*see* **Famous Geologists:** Sedgwick) and Roderick Murchison (*see* **Famous Geologists:** Murchison) on the other, over the establishment of the Devonian. The Old Red Sandstone (ORS) had long been recognized as a distinctive red sandstone unit, but there were marine rocks in Devonshire that seemingly had a similar age. De la Beche argued on lithological and structural grounds that the Devon rocks were not relatable to the ORS, but his opponents successfully argued otherwise using palaeontological criteria.

But then Sedgwick and Murchison fell out, with even greater rancour, over the Welsh strata. Sedgwick studied the rocks of north-west Wales and envisaged a Cambrian system there. It was not, however, well characterized by distinctive fossils. Murchison started from the fossiliferous Welsh Border region and worked towards Sedgwick's territory. They failed, however, to establish a clear section, or boundary, between the two systems and Murchison began to extend his Silurian downwards, eventually extending it to the time when shelly fossils first appeared. Sedgwick, on the other hand, maintained the integrity of his Cambrian and tried to extend his Cambrian upwards into the Silurian domain. The controversy was only resolved after their death by the schoolmaster geologist, Charles Lapworth, who proposed (1879) a threefold subdivision of the Palaeozoic, by analogy with a similar threefold division described for Bohemia by the palaeontologist Joachim Barrande. The 'debatable' ground in the Cambrian and Silurian became the new System, the Ordovician.

Finer biostratigraphic subdivisions (stages) were established by Alcide d'Orbigny in France and Albert Oppel in Germany. D'Orbigny was a grand traveller in France and South America, and amassed huge collections; while Oppel particularly made a detailed survey of the available literature to establish what fossils occurred in which parts of the various systems, notably the Jurassic. It became evident that the Jurassic (for example) could be subdivided into stages, each with its own characteristic suite of fossils; with even finer subdivision possible into zones. D'Orbigny's 27 stages were named according to the localities chosen to 'define' them (by standard sections). Zones, with their characteristic fossil assemblages, were named by Oppel according to their most characteristic (or index) fossil species. Ideally, index fossils should be of wide geographical and short temporal range. Ammonites served this purpose well for the Jurassic.

The work of d'Orbigny and Oppel was fundamental for stratigraphy (and the aforementioned mapping), but it depended on the existence of 'breaks' in the stratigraphic sequences, such as Georges Cuvier had previously envisaged (*see* **Famous Geologists:** Cuvier). In fact, d'Orbigny willingly accepted this 'catastrophist' stratigraphy, but did not attribute metaphysical significance to the 'breaks'. They were simply useful for the practical purpose of stratigraphic subdivision and delineation of strata. It did seem, however, that the fossil record manifested some kind of 'progress', with organisms in the stratigraphic column gradually becoming more like those found alive today.

Darwinism and Evolution

As is well known, Darwin published his theory of evolution in 1859 (*see* **Famous Geologists:** Darwin). It explained the nature of the fossil record well in some respects ('progress', extinction, and the appearance of new forms). But it did not lead one to expect that fossil transitions would be quite 'jerky'. To account for the observed fossil record, Darwin took the view that there were innumerable gaps, where organisms had not been preserved in the first place, or had been lost by subsequent erosion. This point was neither provable nor disprovable, and thus to some degree the catastrophist/gradualist distinction remained metaphysical in character. Even now, smooth 'trends' in fossil forms are rare, but have been found for foraminifera, where numerous specimens may be found in small thicknesses of sediment. (D'Orbigny made important collections of forams but did not use them to discuss this theoretical issue.) In his later work, Darwin increasingly deployed Lamarckian ideas, supposing that the environment 'caused' increased variation (as apparently occurred under domestication), and that acquired characters were inherited. In the late nineteenth century, a significant number of writers turned away from 'classical' natural selection theory and supposed that the stratigraphic record revealed 'directedness' towards apparent goals (e.g., greater size) in a process that apparently occurred independently of natural selection. Where organisms were

seemingly 'trying' to adapt themselves to the conditions of existence, as it were taking charge of their own destinies, this could be construed as a version of Lamarckism. So there were neo-Lamarckians, such as the American palaeontologist Edward Drinker Cope. A variant of their theory was 'orthogenesis' (straight-line evolution), advocated by the German biologist Theodor Eimer. It suggested that evolutionary trends might continue until they became maladaptive, as perhaps in the evolution of the Irish elk to the point that its horns became so heavy as to drive the species to extinction. The empirical background to such ideas was the discovery of gigantic dinosaur skeletons in America by (among others) Cope and his bitter rival, Othniel Marsh.

Glacial Theory

Lyell's theory allowed for periods of terrestrial warming or cooling according to where the highest mountains happened to be on the Earth's surface at any given time. But there was no 'direction' to the process. In the early nineteenth century, some little-known Swiss men, observing moraines, etc., drew attention to the former extent of the Swiss glaciers. This suggestion was picked up by the Neuchâtel professor Louis Agassiz (1837) (*see* **Famous Geologists:** Agassiz). His idea was that the Earth was cooling, but did so in such a way as to fall, at times, below the temperature of an ordinary cooling curve, and then reverted to the 'normal' temperature of a cooling body. So not long before the present (in geological terms) the temperature could have been significantly lower than today, low enough to produce an 'ice age', with widely extended glaciers. The hypothesis could explain many of the curious phenomena of the superficial deposits of northern Europe: the spread of 'boulder clay' over the northern plains; vast gravel deposits in the valleys running north from the Alps; scratches on rock surfaces now without ice cover; boulders distant from places where such rocks occurred *in situ*, etc. So Agassiz promoted the idea of an *Eiszeit* or Ice-Age, which could account for phenomena formerly explained by the action of Noah's Flood. Agassiz attended a British Association meeting in 1840 and attracted some converts, notably William Buckland and even (temporarily) Lyell.

But the land-ice theory did not receive immediate acceptance. Icebergs were also proposed to transport 'erratic' boulders. So arose the 'glacial submergence' theory: that there was simultaneous global cooling and lowering of land surfaces *or* rise in sea levels. The iceberg hypothesis gave rise to the notion of 'drift' deposits. Loose marine shells were found atop some hills in North Wales, and some submergence seemed necessary for them to have got there. Agassiz's land-ice theory seemed incredible and incompatible with uniformitarian doctrine.

However, the land-ice theory began to make more progress after about 1860 when it was taken up by the British Surveyor Andrew Ramsay. For several years it contended with the glacial-submergence doctrine, eventually winning out over the latter. In Switzerland, the stratigrapher Adolf Morlot (1856) noted what appeared to be multiple glacial deposits near Lausanne, occurring in the 'Quaternary' (so named by Paul Desnoyers in 1829 for deposits of the Seine Basin, thought to be younger than Tertiary). The idea of four major glaciations (Günz, Mindel, Riss, and Würm) became almost paradigmatic through the publications of Edouard Brückner and Albrecht Penck (1901–1909) on the outwash gravels of the Alps and the Pyrennees, having been given an attractive astronomical explanation by James Croll (1875) in terms of the changing ellipticity of the Earth's orbit and precessional motion. The fourfold Quaternary glaciation was repudiated in the twentieth century, to be replaced by the more complicated theory of Milankovich cycles. But already in the nineteenth century glacial theory had largely solved the riddle of 'diluvial' deposits and put paid to a *global* Flood as a geological agent.

Geomorphology and Landforms

Attention to landforms and river patterns in the nineteenth century, coupled with ideas about slow land elevations and denudations, allowed explanation of many geomorphological peculiarities. Joseph Jukes in Ireland (1862) emphasized the role of rivers (rather than the sea) in eroding away its former Carboniferous cover, producing river patterns that could not have been caused by the sea (though the rivers supposedly eroded a surface that had undergone marine peneplanation). He also enunciated the principle that rivers transverse to a geological structure are generally older than their longitudinal or strike side branches. Rivers excavate their own valleys, but adjust their courses to fit underlying geological structures. At about the same time Ramsay successfully argued that glaciers could have excavated rock basins in relatively soft strata, forming lakes that are presently being filled with sediment after the end of the glacial epoch.

Notable advances in geomorphological understanding were made by the geological explorations of the American West by the likes of J.W. Powell, G.K. Gilbert, C.E. Dutton, and W.M. Davis. Powell in 1875 had the idea of 'base-levelling', and that drainage patterns could be older than the mountains through which the rivers run. Thus rivers could cut down at a rate equivalent to that at which land was

rising, explaining otherwise anomalous drainage patterns. Gilbert introduced the idea of graded rivers and tried to educe laws governing the sculpture of the Earth's surface. Dutton in 1882 drew attention to the influence of differences in the hardnesses of strata on the cross-sections of river valleys and hence the development of canyon profiles. Davis, considering evidence from the Appalachians (1889), wrote about the 'life-cycles' of land-forms, from 'youth', through 'maturity', to 'old age'. But rivers could be 'rejuvenated' by land elevation. He spoke of 'antecedent drainage' and in general considered the *evolution* of landscapes almost as if they were living entities.

In 1859, back in England, Darwin had given greater emphasis to marine erosion, proposing an excessive estimate of the age of the Weald valley near his home on the assumption that it was cut by the sea. Earlier, during his *Beagle* voyage, he had successfully explained the origin of coral atolls by supposing that coral could grow upwards at about the rate that land was subsiding, hence explaining the peculiarities of fringing reefs.

Mountain Formation and Isostasy

The idea of the major features of the Earth's surface being due to cooling, contraction, and wrinkling of its crust dominated the nineteenth century from the work of Léonce Élie de Beaumont and Eduard Suess particularly. But the work of the American James Dwight Dana was probably more influential worldwide than that of Élie de Beaumont. Well travelled through his participation in the Wilkes expedition, in 1847 Dana recognized a fundamental difference between continents and ocean basins, thinking that they formed early in the planet's history. North-west and north-east trending island chains supposedly marked 'cleavage lines' originating back in Archaean (primaeval) times, which still influenced the evolution of the crust. The continents cooled and solidified first, whereas ocean basins were situated where subsequent cooling and contraction were concentrated and where volcanoes were still chiefly active. Basin subsidence caused lateral pressure, folding, and uplift of the continental margins.

In 1856 Dana envisaged growth of the North American continent, starting from the V-shaped ancient core or Azoic nucleus of the metamorphic rocks of the Hudson Bay region, to which additions were successively made from the south-east and south-west. (This was the forerunner of the twentieth century concept of 'cratons'.) Further, he thought of continental interiors as relatively stable, so that folding and faulting were concentrated at their margins, as exemplified by the Appalachian range. The whole process of contraction and accession of new land was supposedly divinely guided or teleological.

Dana was challenged by his countryman in 1859, James Hall (*see* **Famous Geologists:** Hall), who saw vertical movements of the crust as responses to gravitational loading. Sediment could be deposited in long trenches (later called geosynclines), parallel to the continental margins. Global contraction caused crumpling of the upper sediments, while the downward-bulging trench bottoms would be fractured and intruded by igneous matter. Linear mountain ranges such as the Appalachians might have accumulated their sediments in geosynclinal structures, but the process of their uplift following sediment deposition was obscure. Élie de Beaumont's lines of mountain elevation were Hall's lines of original accumulation. But, as Dana in 1866 complained, Hall offered a "theory of the origin of mountains, with the origin of mountains left out".

In 1873 Dana coined the term 'geosynclinal' (later geosyncline), and its complement 'geanticlinal'. The evolution of the two, and concomitant growth of a continental margin as envisaged by Dana, is illustrated in Figure 1.

Meanwhile, in Switzerland, and especially in the Glarus Canton, geologists such as Arnold Escher had, since the 1840s, been finding substantial evidence for lateral earth movements, and seeming inversions of the usual order of strata, according to the palaeontological evidence. In 1878 Albert Heim proposed a great double fold for the Glarus region with two mountain masses moving together. He was addressing a real structural problem, but his solution was mechanically implausible. The great synthesis for the Alps, and, indeed, worldwide, was provided by the Austrian geologist Suess's *Das Antlitz der Erde* (1883–1909) (*see* **Famous Geologists:** Suess).

By Suess's theory, global contraction gave rise to subsidence in parts of the Earth, with generation of tangential forces, manifested as thrust faults. Thus, the Alps might thrust northwards over the 'foreland' region of Germany, while in the 'backland' of the Mediterranean and Adriatic seas there could be further collapse and volcanic activity. Likewise the Carpathians could ride over the Russian foreland. But in China the lateral movement was southward, as in Yunnan. Suess's theory was also linked to stratigraphy. Collapses of ocean floor regions would cause worldwide marine *regressions*. But an oceanic collapse would stimulate erosion of the more exposed land surfaces, and the increased sediment supply would fill up the basins and produce marine *transgressions*. So there would be cycles of erosion and deposition, and since all the oceans interconnected the

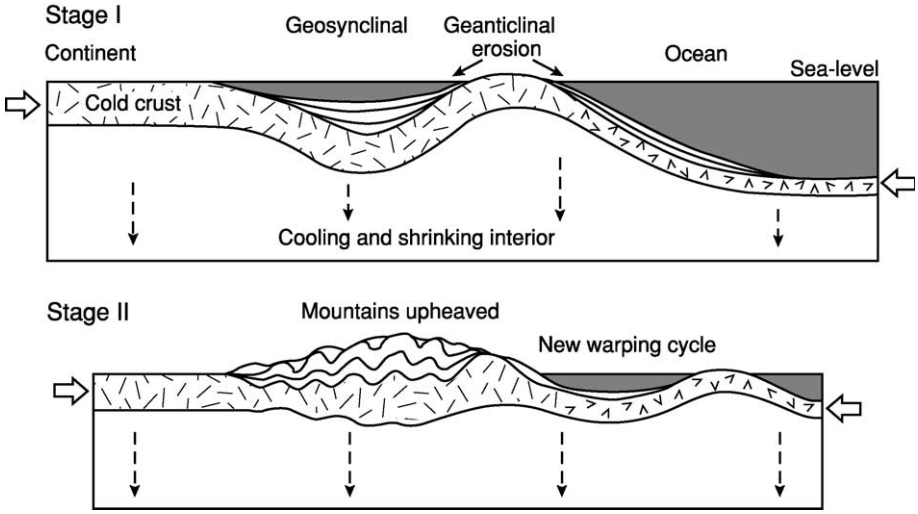


Figure 1 Representation of Dana's 1873 theory of mountain building, as reconstructed by Robert H. Dott (1997, p. 299). Reproduced by courtesy of the *American Journal of Science*.

worldwide nature of the major divisions of the stratigraphic column could be explained. In 1888, Suess thus introduced the notion of 'eustatic movement' or eustasy.

Additional ideas were furnished by the trigonometric survey conducted in India in the nineteenth century, when it was discovered that there was a discrepancy between the results of astronomical and geodetic determinations of latitude along the line of meridian being determined. In 1855, the mathematical Archdeacon of Calcutta, John Henry Pratt, ascribed this to the Himalayas' attraction of the plumb-line used in the astronomical determinations. The point was noticed by the Astronomer Royal, George Airy, who thought that mountain ranges might have underlying 'roots' in the Earth's fluid interior, with mountains being held up rather like floating icebergs. Pratt also assumed a balance between Archimedean upthrust and the weight of crust, but the two models were different. Pratt assumed a common 'depth of compensation', with the less dense parts of the crust standing higher, while Airy's model had deeper roots for the higher mountains.

The general idea was taken up by Dutton in 1871 and in 1892, who introduced the term 'isostasy' or 'equal standing'. An inhomogeneous spinning earth would not be perfectly spheroidal but would have depressions where the crust was more dense and bulges where it was less dense. The form of the Earth (geoid), determined for North America by the US Coast and Geodetic Survey and eventually published early in the twentieth century by John Hayford, suggested that the Earth was generally in a state of isostatic balance. But the Pratt model was assumed for ease of computation, and it entered the thinking of William Bowie (who succeeded Hayford) and subsequently stood in the way of acceptance of continental drift theory.

The Formation and Age of the Earth

After Lyell, most geologists were willing to accept an indefinitely large age for the Earth, and the Biblical age was only upheld by theologians and non-geologists. Darwin thought he could assume as much time as he required for the processes of erosion and deposition and organic evolution. His figure of 1859 for the time taken for the marine erosion of the valley of the Weald was some 300 million years, a figure with which he and other geologists were comfortable. However, in 1862 the physicist William Thomson (Lord Kelvin) argued that the sun acquired its initial heat from meteoric impacts, and had subsequently been cooling. It might be somewhere between 10 and 500 million years old. He estimated that a cooling Earth with temperature gradient and heat loss as at the present could be 20 to 400 million years old. In 1871 he also suggested that if the Earth acquired its spheroidal shape when it was still molten, one could estimate its rotation rate at the time of solidification. Subsequently, it had been slowing its rotation, due to tidal friction, to the present value. By estimation of the decrease of rotation rate and thermal contraction it seemed that the age might be of the order of 100 million years. This was just about acceptable to evolutionists and geologists, but both would have preferred more time, while being unable to argue with Kelvin's calculations. The matter rested thus until 1909, and the work of John Joly, where the effect of radiometric heat was considered and the

supposed rate of terrestrial cooling was seen to need revision. Thus, physical objections to the great age of the Earth, which seemed substantial in the second half of the nineteenth century, subsequently receded.

The eighteenth-century theory of the Earth's origin was that of Laplace and Kant, which supposed a coalescence of matter under gravity from a primaeval spinning 'cloud' of nebular matter. Thus, the Earth originated as a spheroid of hot gas, which cooled to a liquid and then to a solid state. This had dynamical problems, and at the end of the nineteenth century the influential American geologist, Thomas Chamberlin, in collaboration with the astronomer Forest Moulton, proposed their theory of 'planetesimals' (miniscule planets), according to which matter dragged out from the spinning sun by some other star could have accreted to form small solid bodies, which could have collided to form the several planets. This was the beginning of planetary geology, or the conflation of geology, cosmology, and astronomy.

Rocks and their Formation

There were few major contributions to sedimentary petrology in the period here under review. The Irishman Patrick Ganly's 1830s discovery of the 'wayupness' criterion offered by current bedding was not utilized until the twentieth century. In 1839 Christian Ehrenberg published his microscopic studies of chalks and limestones. Formal distinction of sedimentary, igneous and metamorphic rocks was made by Henri Coquand in 1857.

Various igneous rocks such as basalts, granites, gabbros, syenites, or porphyries had been recognised since antiquity and many classifications were proposed in the nineteenth century, according to chemical and/or mineral composition, texture, or supposed mode of formation, but the field was confused. Examination of rocks in thin section by Henry Sorby assisted in a sense, but the proliferation of information also added to the confusion. While Huttonian theory was triumphant as regards Werner's original theory, there was continued interest in the role of water in the formation of igneous and metamorphic rocks. Notably, in 1857 and 1859 Gabriel Auguste Daubrée of the French Mines Department subjected materials to high temperatures and pressures, with or without water, and concluded that new minerals could crystallize without wholesale melting. He thought that past conditions could have been radically different from those at present and that there might have been an *ocean primitif*. Granite was thought to be produced by 'aqueous plasticity', not igneous melting. At Freiberg, Bernhard von Cotta held analogous views. But in 1860 Daubrée thought that foliations were due to pressure during 'regional metamorphism'.

The variety of igneous rocks raised the question whether there were different kinds of subterranean magma, or whether some processes of differentiation occurred from essentially the same starting material. In 1844 Darwin had the idea of differentiation of magma by gravity settling of first-formed crystals; and in 1846 he distinguished cleavage, foliation, and stratification (while regarding gneisses as stratified rocks). Dana thought that differentiation of magma might precede crystallization. By contrast, on the basis of observations in Iceland, in 1851 and 1853 the chemist Robert Bunsen proposed that there were two separate magma chambers under the island, producing 'trachytic' and 'pyroxenic' rocks or intermediate mixtures. In 1853 Wolfgang Sartorius von Waltershausen hypothesized the existence of different subterranean zones; and the eruption of more siliceous types preceded the more basic, thus relating igneous compositions to age. In 1857 Joseph Durocher asserted the liquation model and priority for the idea over Bunsen. Metallic lodes were ascribed to '*émanations*'. Following work in Hungary and California, Ferdinand von Richthofen envisaged a succession of magmas, the earliest being more siliceous, the great outpourings of Tertiary basalts being due to earlier depletion of siliceous magma. In 1878 in America, Clarence King thought pressure release could facilitate fusion. In 1880 Dutton suggested that fusion could follow pressure release, local temperature elevation, or water absorption.

The master petrologists of the period were Harry Rosenbusch in Strasbourg, Ferdinand Zirkel in Leipzig and Ferdinand Fouqué and Auguste Michel-Lévy in Paris, who specialized in the study of feldspars. All were adept with the use of the petrographic microscope. In 1873 Rosenbusch published a catalogue of all then known magmatic and metamorphic rock types. Rosenbusch's 1877 study of metamorphism around the Barr–Andlau granite in the Vosges was important for his recognition of zones of contact metamorphism (schists, knotted schists, hornfelses), seemingly without feldspars. But Michel-Lévy found feldspars in the contact aureole of the Flamandville granite in Normandy and he and Fouqué thought there was no fundamental distinction between contact and regional metamorphism. Throughout this period Continental petrologists continued, in the Wernerian tradition, to try to find relationships between age and 'hard-rock' composition, whereas their British counterparts chiefly concerned themselves with biostratigraphy.

However, in 1893 the British surveyor George Barrow, working on metamorphic rocks in the southern Scottish Highlands, found characteristic

metamorphic minerals (sillimanite, kyanite, and staurolite) around a granitic mass, and this gave him mappable subdivisions of the region. These were developed in the twentieth century as 'Barrovian zones', but the useful idea was not initially followed up.

Experimental petrology was undertaken in the nineteenth century, but until high-pressure and pressure techniques were developed in the twentieth century for simulating rock formations, work on phase diagrams was developed, and ideas about the structure of the Earth's interior could be pursued through seismology, petrological understanding remained speculative and somewhat at the level of natural history.

See Also

Famous Geologists: Agassiz; Cuvier; Darwin; Hall; Lyell; Murchison; Sedgwick; Smith; Suess. **Geological Maps and Their Interpretation**. **History of Geology From 1780 To 1835**. **Metamorphic Rocks:** Facies and Zones. **Sedimentary Processes:** Glaciers. **Stratigraphical Principles**. **Time Scale**.

Further Reading

Berry WBN (1968) *Growth of a Prehistoric Timescale Based on Organic Evolution*. San Francisco and London: W.H. Freeman & Co.

Bowler PJ (1976) *Fossils and Progress: Paleontology and the Idea of Progressive Evolution in the Nineteenth Century*. New York: Science History Publications.

Buffetaut E (1987) *A Short History of Vertebrate Palaeontology*. London, Sydney and Wolfeboro: Croom Helm.

Burchfield JD (1975) *Lord Kelvin and the Age of the Earth*. New York: Science History Publications.

Cross W (1902) The development of systematic petrography in the nineteenth century. *The Journal of Geology* 10: 331–376.

Davies GH (1969) *The Earth in Decay: A History of British Geomorphology 1578–1878*. London: Macdonald Technical and Scientific.

Dott RH (1997) James Dwight Dana's old tectonics—global contraction under divine direction. *American Journal of Science* 297: 283–311.

Gohau G (1997) Évolution des idées sur le métamorphisme et l'origine des granites. In: Bonin B, Dubois R, and Gohau G (eds.) *Le métamorphisme et la formation des granites evolution des idées et concepts actuels*, pp. 9–58. Paris: Nathan.

Greene MT (1982) *Geology in the Nineteenth Century: Changing Views of a Changing World*. Baltimore: Johns Hopkins University Press.

Nieuwenkamp W (1977) Trends in nineteenth century petrology. *Janus* 62: 235–269.

Oldroyd DR (1996) *Thinking about the Earth: A History of Ideas in Geology*. London: Athlone Press; Cambridge (Mass): Harvard University Press.

Rudwick MJS (1985) *The Great Devonian Controversy: The Shaping of Scientific Knowledge among Gentlemanly Specialists*. Chicago and London: The University of Chicago Press.

Secord JE (1986) *Controversy in Victorian Geology: The Cambrian–Silurian Dispute*. Princeton: Princeton University Press.

Yoder HS (1993) Timetable of petrology. *Journal of Geological Education* 41: 447–489.

Young DA (2003) *Mind over Magma: The Story of Igneous Petrology*. Princeton and Oxford: Princeton University Press.

Zittel KA von (1901) *History of Geology and Palaeontology to the End of the Nineteenth Century*. Translated by Maria M. Ogilvie-Gordon. London: Walter Scott.

HISTORY OF GEOLOGY FROM 1900 TO 1962

D F Branagan, University of Sydney, Sydney, NSW, Australia

Introduction

The period prior to the plate tectonics revolution of the 1960s has been said by some historians of science to have been a time of stagnation for geology. This supposed stagnation is based on the idea, then largely held, of the fixity of the continents and oceans, which some have extended to suggest that geologists in the main remained rather fixed in their ideas and were concerned only with mundane geological matters. Was this so? It might be partly true, in that only a few people were attending to 'large questions'. However, many unsolved geological problems were studied, and one could argue that these had to be tackled before fundamental concepts could be challenged. The first half of the twentieth century was marked by two world wars and the disruption of scientific contact for much longer than just the war years. However, even in these years two things happened that would benefit geology: techniques were developed that allowed the quantification of many aspects of geology; and geology was increasingly applied to engineering problems.

Aerial photography is a prime example. It led to more rapid geological mapping, and by the mid-1920s it was applied to the search for gas- or oil-yielding structures. Later, systematic photographic coverage contributed to the understanding of regional, and even supra-regional, problems. Likewise the geophysical instruments used to detect spatial changes in rock strata were improved through research during the First World War. Geologists during that war introduced environmental geological maps, the forerunners of many variations of the geological map later developed by engineering geologists. Attention was also given to locating strategic minerals, especially during the Second World War. In the interwar period, explosion seismology was used in oil prospecting (first in Oklahoma in 1921), and other geophysical techniques were brought into use. Although modern ocean research had begun in the 1930s, the development of radar during the Second World War quickly produced the first significant information about the ocean floors, seamounts, and deep trenches. The accumulation of data required more technical expertise, and in general the 1950s saw the rise of 'team' efforts and multiauthored publications. To some extent, this heralded the demise of the geological polymath, and few people attempted to generalize from the new information that was forthcoming, preferring to be one of a consensual group.

The Age of the Earth, and its Subdivisions

Determining the age of the Earth was perhaps the most significant achievement of geological research in the early twentieth century. The geological significance of radioactivity was recognized in 1903, when Pierre Curie and his researchers found that radium salts release heat constantly, and Ernest Rutherford and Frederick Soddy saw that energy was released by radiation from radioactive materials. These findings implied that the Earth was not necessarily cooling. Rutherford also noticed that helium was trapped in radioactive minerals and thought that measuring the content of this gas might be used to determine geological ages. Lord Rayleigh and Bernard Boltwood were the first to study the radioactivity of rocks, and in 1905 Boltwood noted that lead was invariably associated with uranium and might be an end-product of the radioactive decay of uranium. Experimentation, mostly with 'home-made' equipment, indicated that there were two uranium isotopes, which decayed at different rates and produced different lead isotopes while releasing helium. By 1907 Boltwood was working on the uranium–lead ratios, while Rayleigh dated minerals using the helium produced. However, the gas could escape, leading to errors in the ages determined. In 1913, Arthur Holmes published the first full review of the methods and became personally involved in the experimental work.

Not all geologists liked the idea that the Earth might be billions of years old, and Holmes faced opposition to his conclusions about the Earth's age. Only in the late 1920s did his work begin to be accepted, when he calculated that the Earth was about 3300 Ma old. One of the problems attending radiometric age determinations was the uncertainty as to whether the rate of breakdown was constant, an issue raised by Joseph Barrell in 1917. In 1919–1920, Frederick Aston built the first mass spectrometer, which separated atoms according to their weight, but it was nearly 20 years before consistent results were obtained, using a machine built by Alfred Nier, with newly developed vacuum pumps. In 1956 Clair Patterson calculated the generally accepted age of the Earth (4.55×10^9 years) and the age of the solar system, based on lead-isotope ratios measured from iron meteorites. Later workers took up the study of various other isotopic relations, such as rubidium–strontium (mainly in the 1950s and 1960s), potassium–argon, and potassium–calcium (between 1920 and 1943). Different methods proved suitable for determining the ages of different parts of the stratigraphical time-scale. For the most recent 20 000 years or so, the radiocarbon (^{14}C isotope) method, developed by Walter Libby in 1952, has proved to be an effective dating tool. Radiocarbon is formed when atmospheric nitrogen is bombarded with neutrons and taken into organic material by photosynthesis, radioactively decaying after death. By the late 1950s the field of geochronology was well established. However, the expensive equipment required made it a somewhat exclusive field, and it has remained so.

Petrology (Igneous and Metamorphic)

The diversity of igneous rocks led to many attempts to explain the varieties and to classify the rocks into meaningful groups. Experimental petrology as a special branch of geology began in the 1890s, as the chemical analysis of rocks and minerals became more precise and more closely linked to the identification of the relationships of rocks using the petrological microscope. Controlled high-temperature and high-pressure studies commenced at the Geophysical Laboratory of the Carnegie Institution, established in Washington in 1905. Classifications were proposed on the basis of mineral content and texture, and more particularly chemical content, as assay methods became easier and cheaper. The so-called 'CIPW normative classification', the result of research by Charles Whitman Cross, Joseph Iddings, Louis Pirsson, and Henry Washington

in 1902, was one of the first classifications to be widely accepted. It depended on the recalculation of the 10 or so oxides most commonly found in igneous rocks by analysis, and assigning the values to particular minerals. It was particularly valuable for glassy rocks, where the minerals were not easily identified. The 'norm value' contrasted with the 'mode' or actual mineral composition. Although it was an artificial system, norm calculations allowed ready comparison, based on the rocks' chemistry, and grouping and subdivision without the distraction of minor mineral or textural variations. It was thought to give clues about the order of mineral crystallization in magmas.

Variations of the normative approach were proposed by researchers such as Samuel Shand, who suggested a classification of igneous rocks based on the proportion of silica present. While this was a 'logical' classification, it did not explain the origins of such variations. Classifications were also proposed based on rock textures, such as flow-banding, porphyritic content, and crystallinity.

There was widespread acceptance that the compositional and textural variations were caused by magmatic differentiation. This was thought to result from the presence of immiscible liquids within the magma, or from the separation of minerals in crystallization order by, for example, the sinking of heavier early-formed minerals. Others placed more emphasis on crystal–liquid fractionation (which was subsequently regarded as more important). A major influence from 1910 was the experimental work of Norman Bowen at the Carnegie Laboratory; one of his results is encapsulated in Figure 1. His classic book *The Evolution of the Igneous Rocks* (published in 1928), which summarized much of his experimental work to that time, influenced several generations of petrologists.

In the 1940s, Herbert Read reintroduced the idea of granitization, rather than magmas, as the major source of 'igneous' rocks. He argued that there was a space problem in the emplacement of large batholiths and suggested that igneous rocks were often earlier-formed rocks that had been altered by the action of active fluids that caused recrystallization. Read suggested that there was a 'continuum' from regional-metamorphic rocks to igneous rocks through what he called 'migmatitic' rocks (a term introduced by Jakob Sederholm). The role of volatiles consequently began to attract more attention, and Bowen, with Orville Tuttle, studied partial melting (hydrothermal activity) in addition to magma fractionation. Laboratory work, using pressure cells, contributed to the study of metamorphism, simulating conditions of pressure and temperature deep within the Earth. So, work on metamorphism proceeded hand-in-hand with the studies of igneous rocks.

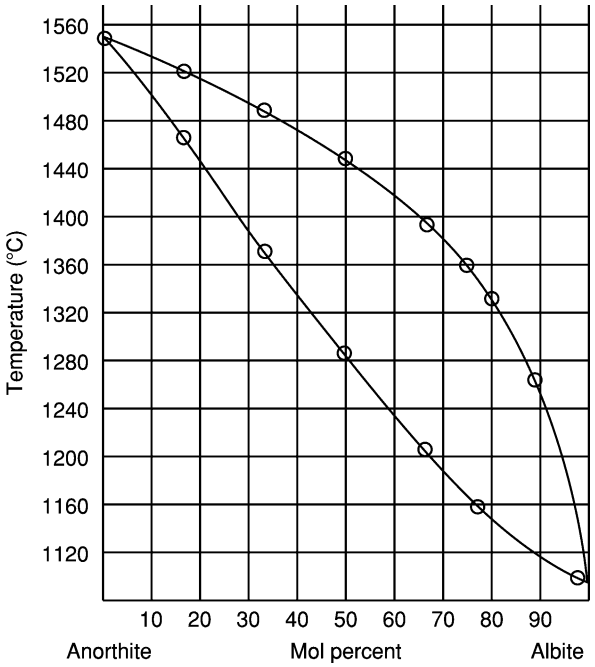


Figure 1 Phase diagram for plagioclase feldspar at one atmosphere. (Reproduced from Bowen N (1913) The melting phenomena of the plagioclase feldspars. *American Journal of Science, Series 4* 35: 577–599 [Figure p. 583].)

In 1912, Victor Goldschmidt recognized that metamorphism essentially depended on two physical variables, pressure and temperature, and that most changes occurred while the original rocks remained essentially solid. A few years earlier Friedrich Becke had clearly described different metamorphic fabrics (distinctive patterns of minerals caused by different degrees of metamorphism). While Goldschmidt's ideas were valid in most cases, there were exceptions. In 1936, Dmitrii Korzhinskii and others showed that some melting could occur and that the mobility of some fluids, particularly in low-temperature hydrous activity, was important, although it had been dismissed at the beginning of the century. Related to this was the study of diagenesis, or very low-temperature metamorphism, which occurs as sediments are converted to sedimentary rocks. Serious study of this topic began in the 1920s, when certain mineral associations were recognized as indicators of the changes taking place through time. The work of Douglas Coombs in New Zealand in the 1950s played a major role in understanding these processes.

Geochemistry

The study of geochemistry began in the second half of the nineteenth century with the collecting of analytical data from rocks, minerals, and mineral waters by

a few chemists and geologists, including, at the turn of the century, Frank Clarke. In the 1920s, attention turned to attempting to understand the distributions and abundances of the elements in the Earth's crust, led by Goldschmidt in Germany and Vladimir Vernadsky in Russia. Goldschmidt's *Geochemistry* (published in 1957) brought together the essentials of the subject, although it had already moved on with the studies of isotopes begun at the University of Chicago, immediately after the Second World War, by Harold Urey and his co-workers. An international Geochemical Society was formed in 1955, and important seminars were held in 1957 and 1958 at the Carnegie Institution. The study of organic geochemistry also developed, with practical work related to deleterious substances in coal, mine dusts, and petroleum, and their dispersal into the atmosphere.

Stratigraphy and Sedimentology

Stratigraphy and the attempt to write the history of the Earth's crust had been the essence of geology since the late eighteenth century. But the rock layers and their enclosed fossils called for further study. During the first part of the twentieth century, the field mapping of the late nineteenth century was extended to many parts of the Earth that had been previously untouched. A major improvement in the maps was the attention to structural and sedimentological detail. Around 1920, it was shown that features such as cross-bedding and graded bedding could indicate the order of superposition, particularly in vertical and highly folded non-fossiliferous strata. The difference between cleavage and bedding in metamorphic rocks also had to be recorded. Thus previous interpretations of important regions, such as the Scottish Highlands, had to be re-examined. The structural geologist Marland Billings wrote "one clear contact or key bed is more valuable than a hundred petrofabric diagrams".

To make sense of the observable history worldwide, international agreement was needed about naming certain features and presenting them on maps and cross-sections. Stratigraphical commissions were established to set down rules for such things as identifying 'type localities'. The study of facies – the total nature of a volume of strata (rock composition, fossil content, type of bedding, sedimentary structures, etc., usually reflecting the conditions of origin) – was developed by Johannes Walther and others. Walther recognized that facies relations were dynamic. He stated that only rock types that can be deposited side-by-side can overlie each other directly in a vertical sequence. This led to the study of time-stratigraphical units, with the recognition that rock boundaries did not necessarily correspond with time. In this work, and in sedimentation studies in general, Amadeus Grabau was an acknowledged leader. Statistical analysis, long neglected by geologists and first used in petrology by Paul Niggli in 1924, became widely applied in stratigraphy, with information about rock-layer thickness (isopach) and lithofacies variation (e.g. sand–shale ratios) being presented in graphical form.

Closely related to stratigraphical studies was the examination of the materials of sedimentary rocks. Pioneers were Johann Udden and Chester Wentworth, who, in about 1920, studied size distributions and the shapes of grains, devising a quantification chart that was useful in the interpretation of sediment histories. Ralph Bagnold's study of the physics of blown sand in 1941 brought together considerable earlier research. Between 1900 and the 1950s, the heavy accessory detrital minerals of sedimentary rocks were widely used for stratigraphical correlation by William Rubey, Percy Boswell, and others. Laboratory work began to supplement field studies. From the 1930s flumes were widely used by Henry Milner, Paul Krynine, and others to examine the behaviour of sedimentary materials under different conditions. One problem, not always adequately addressed but worked on by Francis Shepard, one of the first US marine geologists, and M. King Hubbert, was the scaling from the actual geological dimensions to the laboratory dimensions.

In the early 1940s, Krynine attempted to set up a sediment research laboratory in the USA, bringing together the skills of oil companies, academics, and government geologists. This did not eventuate, and separate laboratories continued. In 1950 experimental flume work by Phillip Kuenen and Carlo Migliorini showed how 'turbidity currents' occurred, carrying materials rapidly downslope from a shallow source. This led to a 'turbidite revolution' in interpreting many types of stratigraphical occurrences, particularly greywackes, graded bedding, and other sedimentary features, and to an understanding of how some submarine cables might have been cut on continental shelves following earthquakes.

Palaeontology

The study of invertebrate fossils continued apace during the first half of the twentieth century. Initially palaeontologists were concerned mostly with taxonomy and classification, but additional fields opened up with the study of the evolution of particular groups. The Jurassic ammonites provided an example of a rapidly evolving fauna, which helped to pin down time zones, while the Ordovician and Silurian graptolites facilitated correlation between sedimentary rocks from these periods in widely separated parts of the

world. Palaeoecology began to develop, with the consideration of relations between fossil groups and sedimentary facies, and researchers began to lean heavily on the study of present-day environments.

From the 1920s, the importance of previously neglected organisms, such as microfossils, in identifying sections of the Tertiary epochs with the potential for oil productivity was recognized. Palynology, the study of plant spores, was also taken up, along with the parent study of palaeobotany. The period from the late nineteenth century to 1930 has been called the 'heroic period' of vertebrate palaeontology. Led by Henry Osborn, there were major discoveries on all continents. In 1910 Osborn proposed that Central Asia was the cradle of mammalian evolution, and much research was devoted to testing this theory. Asiatic discoveries expanded the fossil vertebrate (especially mammalian) record back to the Permian. Of particular significance was the discovery of human remains near Peking in 1926. Important work in palaeoanthropology was also carried out in Europe, Africa, Indonesia, Australia, and elsewhere.

Structural Geology

Structural geology made considerable progress in the first half of the twentieth century, often as a result of studies of metamorphic rocks. In the early 1930s, Bruno Sander and Walter Schmidt initiated 'petrofabric analysis': the study of spatial relations, including those between the individual minerals making up a rock, and the movements that could have produced these relations. The methods were used to investigate rock deformation and were taken even further to consider the genesis of both sedimentary and igneous rocks. Thus a special field of structural petrology was born, dealing with deformed rocks and their tectonic history.

Analysis of thin-sections using the 'universal stage' allowed the determination of the three-dimensional orientations of mineral grains relative to the original positions of rock specimens recorded in the field, and their representation on stereograms. This led to the recognition of various phases of deformation by workers such as Coles Phillips and Lamoral de Sitter. In structural geology, bedding planes, joints, and foliations could be represented graphically, and data could be averaged by the contouring of data points on the stereograms; hence polyphase deformations could be revealed. However, there were controversies about the significance and order of particular deformation events thus interpreted. Broad aspects of folding and even mountain building were studied in laboratory experiments by Rollin Chamberlin, Bailey Willis, and David Griggs, among others, using theories of scaling from engineering, but this pressure-box work suffered to some extent from the use of unsuitable materials and scaling problems.

Geomorphology

The development of landforms was widely studied in the early part of the twentieth century. In the first half of the twentieth century William Davis's erosion-cycle concept was widely accepted. This was the idea that the landscape tended to be worn down, but with decreasing speed, to a 'base level': the peneplain. The concept of Davis's stages – youthful, mature, and old-age landscapes (followed by rejuvenation by uplift) – became widely accepted, and various topographical levels were thought to represent the end points of separate cycles of peneplanation, recognizable back to the Mesozoic.

However, workers such as Albrecht Penck (1924) suggested other possible methods of surface evolution, including scarp retreat, the preservation of original depositional surfaces, and the formation of pediplains by the coalescence of pediments below scarps. It was not until the 1940s that Davis's concepts were significantly challenged. In 1945, Robert Horton published an article on the development of drainage networks, which marked the beginning of studies concentrating on the mechanisms of denudation rather than description. Arthur Strahler encouraged 'dynamic methods' of study, in which landscape units were treated as open systems in equilibrium, where a change in the system caused an adjustment to offset the effects of the change. Measurement of modifications in river profiles, slopes, and runoff began to provide evidence of landscape evolution. A decade of research on river-channel patterns followed, by workers including Luna Leopold and Markley Wolman (1957), using methods borrowed from engineering (such as fluid mechanics). This work was linked to experimental studies in sedimentology by workers such as John Allen, Alan Jopling, and notably William Krumbein in 1963, who treated beach morphology and processes as part of a system, whose changes could be computed.

Glaciation, Climate, and Palaeogeography

While the question of the existence of an Ice Age was resolved in the nineteenth century, studies during the twentieth century in Europe, North America, and New Zealand showed that it was an event with considerable variations, involving a series of intense glaciations with intervening interglacial periods. Of perhaps greater significance were the studies of earlier glaciations, particularly the Late Palaeozoic event,

which was first recognized in India, then in Australia and South Africa, and a little later in South America. These discoveries gave considerable support to the concept of continental drift. Evidence of Late Precambrian glaciation was also recognized early in the century in China, Norway, and Australia. These studies encouraged research into climate change through geological time.

Palaeogeography was developed in the early 1900s by Auguste de Lapparent. He published maps of France, Europe, and the world. Preparation of the maps required knowledge of the three-dimensional extents of rock units and their environments of deposition. The point of time chosen depended on identification of fossils. A particular exponent was Charles Schuchert. However the maps were general and usually covered too much time (because of a lack of secure age determinations). Nevertheless they were studied by petroleum geologists, and, thanks to the detailed stratigraphical studies that resulted, in return, from oil-search drilling, maps of specific 'slices of time' became possible.

Petroleum Geology

By 1900, some of the basic concepts of petroleum geology were understood. Oil was clearly of organic origin, from both vegetable (dominantly) and animal sources, and had formed at normal rock temperatures. It occurred in reservoirs within sandstones and limestones, and permeability was important, with the limestones usually being 'tighter'. A relatively impervious roof of rock, such as shale, was a primary requisite for oil to accumulate. Many reservoirs showed a separation by gravitation into three layers: gas, oil, and saltwater. A particular problem demanding explanation was the enormous pressures sometimes encountered in oil fields. Edward Orton was one of the first to relate these pressures to artesian conditions.

It took many years to determine how the organic material was transformed into oil. In the 1930s Parker Trask argued that only certain organic material could produce petroleum, while J. M. Sanders thought that almost any organic material could be converted to petroleum, given the right conditions. It seemed that specific sedimentary deposits, such as dark marine shales, were likely source beds. Studies showed that petroleum formed slowly, with solid matter being converted to heavy and viscous fluids. In time these thick fluids were changed by heat and pressure into lighter oils. Oil 'pools', containing immense accumulations of oil, contrast with the disseminated nature of the oil forming in the source beds. Thus geologists began to study the migration from source beds to reservoir rocks. Differential pressure was recognized as the essential cause of the movement of the fluids through porous beds, and permeability was an essential condition.

A wide variety of oil 'traps' was described: structures such as anticlines or domes were important, but there were also depositional traps in which particular favourable beds thinned out between impermeable beds or were partly eroded and covered by 'tighter' beds. Faulting could also cause traps to form. Many traps were produced as a result of a combination of various causes. After the Second World War oil companies devoted considerable attention to the study of modern environments (especially deltas) to elucidate many of the fundamental aspects of the accumulation of organic materials and their conversion to oil, while also investigating the complexities of oil structures and the pressure–temperature regimes that contributed to the variation in types of oil or gas accumulations.

Exploration Geophysics

In parallel with developments in petroleum geology, there were major advances in exploration geophysics during the first half of the twentieth century, particularly between 1925 and 1929. Although concentrating on oil and mineral prospecting, some of these practical advances also contributed to the broader studies of the Earth's interior, to seismology, terrestrial magnetism, hydrology, geodesy, and meteorology. Four geophysical methods were widely used in mineral and oil exploration: seismic, gravimetric, electric/electromagnetic, and magnetic. After the Second World War other methods, such as radiometric, also began to be used. Each of these has been applied in different ways. The seismic methods used artificial shock waves: seismic reflection records the rebound from a reflecting surface, while seismic refraction records the path of waves refracted along high-speed layers. Although seismic reflection is simpler in theory, seismic refraction proved to be more useful in practice. Gravity methods used an Eötvos torsion balance or a gravimeter. Electrical methods measured natural or artificially induced earth currents. The electrical methods were combined and extended to logging the variations in resistance and electric potential down a drill hole. The resistance gave an indication of the type of bed (high-resistance oil sandstones and coal beds contrast with salt-water-bearing sandstones and shales). These electrical tests often gave consistent results within particular beds, allowing the correlation of both stratigraphy and structure (Figure 2). Two magnetic methods were used: one measured anomalies caused by ore bodies containing significant quantities of magnetic minerals; the other measured

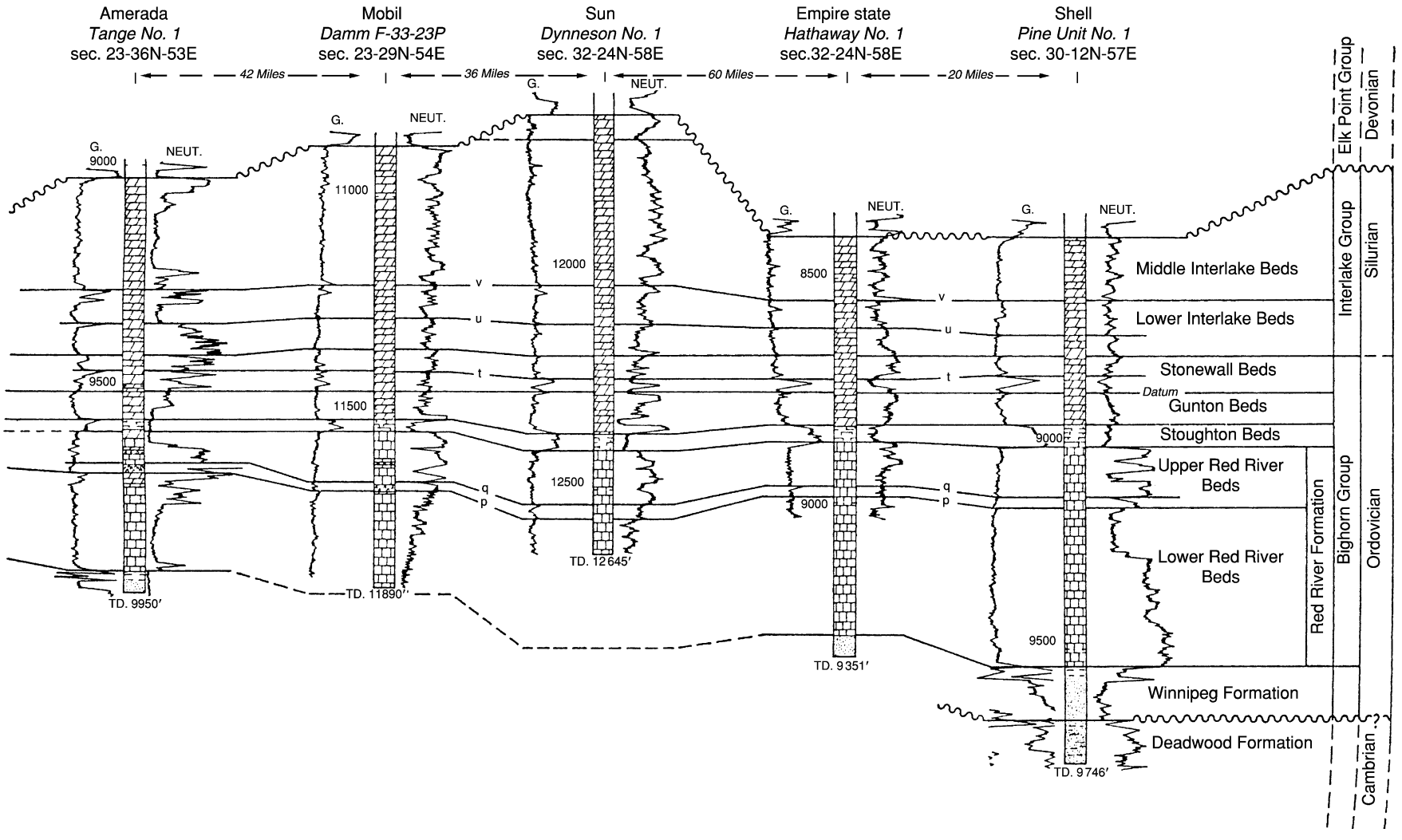


Figure 2 Williston Basin, Eastern Montana, Subsurface section (approximately 150 kms length), showing correlation achieved using Schlumberger G(amma) – Neut(ron) well-logs. (Reproduced from Krumbein W and Sloss L (1963) *Stratigraphy and Sedimentation*, 2nd ed., p. 382. San Francisco, W.H. Freeman & Co.)

slight variations in the magnetic properties of common rocks (basalt is normally more magnetic than other rocks) to construct structural maps, which had the potential to locate salt domes and volcanic plugs. A pioneer of geophysical prospecting was Conrad Schlumberger, who developed his electric methods between 1908 and his death in 1936. Surveying for structure using the Eötvos balance began in 1921 when the Anglo-Persian Company tested it in Hungary. The Royal Dutch Shell Oil Company took up the method the following year in Egypt and later in Texas and Mexico. There was similar interest in Germany and Russia at about the same time. The formation of the American Association of Petroleum Geologists in 1917 led to increased communication between geologists and spread their enthusiasm for testing geophysical methods.

Economic Geology

The considerable close study of metalliferous mineral deposits and related geological aspects early in the century led to the formation of the Society of Economic Geologists and the *Journal of Economic Geology* in 1905. Mineral deposits were often associated with igneous bodies, and Waldemar Lindgren's (1913) hydrothermal classification (hypothermal, mesothermal, and epithermal: that is, deposits formed from high-, moderate-, and low-temperature/pressure fluids emanating from igneous bodies) was widely accepted and applied and in many cases was an effective tool in the recognition of variations within an ore body and was useful in the search for new ore bodies. Important work was done by William Emmons on the secondary enrichment and zonation of ore bodies. There was general acceptance that deposits showing evidence of original sedimentary bedding were the result of 'replacement' of the bedding by later-introduced ore-bearing fluids from a nearby igneous source. In the 1950s the 'replacement' theory of the formation of certain ore bodies was challenged by Haddon King, who with his associates studied the massive Broken Hill ore body in Australia. King proposed that the lead and zinc layers were deposited as sediments and that the only major change had been later folding. This idea has become widely accepted for many of the world's largest base-metal deposits and the major iron-ore bodies. The significance of micro-organisms and their ability to concentrate and deposit metals were uncovered by laboratory experiments such as those by Lourens Baas-Becking. The use of reflectance microscopy (mineragraphy) to identify opaque ore minerals, study the relationships between such minerals, and find clues about the deposition of ores began with J. der Veen in the 1920s. It reached a high level in the 1940s and 1950s through the work of Hans Schneiderhöhn and Paul Ramdohr in Germany, and Frank Stillwell and Austin Edwards in Australia.

Engineering Geology

The failure of the St Francis Dam in California in 1928 drew attention to the need to assess foundation conditions and rock quality. After the Second World War there were numerous attempts to quantify, for example, the rate of weathering of stone, joint distributions, and other weaknesses within rock masses. The field of rock mechanics grew from these developments through the efforts of workers such as David Griggs, Karl Terzaghi, Charles Berkey, John Jaeger, J. Talobre, and Robert Legget as well as researchers in South African mines and Australian hydroelectric schemes.

In the late 1930s geophysical methods (electrical and seismic) were adapted by engineering geologists to determine the depth to solid rock at dam sites and elsewhere, while electrical methods were used to determine the depth to the water table in arid areas. Such uses continued to develop after the Second World War.

World Views

Despite the growth of specialization, there were, of course, attempts to develop global theories, and global structural patterns were particularly discussed in the early twentieth century. The most famous of these studies was Eduard Suess's *Das Antlitz der Erde*. It became influential in the English-speaking world following its translation in the early 1900s. A particular aspect of Suess's theory, based on the study of major features, such as mountain ranges and the patterns of coastlines (Atlantic and Pacific types; the former 'fractured', the latter with fold mountains parallel to the coast), was the idea of an Earth that had been contracting since its formation. Suess rejected the idea that the present continents and oceans had existed from earliest times, believing that the Pacific Ocean was the oldest, possibly formed when the Moon separated from the Earth. Hans Stille also believed in a contracting Earth, but majority opinion held to a fixist concept and the permanence of the continents and oceans. This was questioned, albeit cautiously, by some. Reginald Daly was more adventurous in *Our Mobile Earth* (published in 1926) and was supported by the seismologist Beno Gutenberg in proposing mobility of the Earth's crust, even though the processes were unknown.

The associated concepts of geosyncline (mobile fold-belt) formation and isostasy to explain the accumulation of enormous thicknesses of (mainly)

shallow-water sediments and mountain building, although formulated in the second part of the nineteenth century, continued to be developed, and were widely applied in the period up to 1950, particularly in North America. There were arguments as to how the geosynclines had contributed to continental growth. Was it by accretion on the oceanic edges of supposedly stable continental cratons? Or was a geosyncline formed within a continental mass when one side had subsided? Such subsidences were often invoked to explain the cutting off of routes of animal and plant migration between continental masses. While the geosynclinal concept as the basis of a tectonic theory has, since the acceptance of plate tectonics, been dismissed by many, the descriptive aspects of geosynclinal sedimentation are still useful (*see* **Famous Geologists:** Hall; **History of Geology From 1835 To 1900**).

The meteorologist Alfred Wegener published his famous book, Die Enstehung der kontinente und Ozeane (*The Origin of Continents and Oceans*) in 1915, proposing a tectonic theory based on lateral movements of the Earth's continental crust. Others, such as Frank Taylor, had previously made similar suggestions, but had not set out their ideas as fully as Wegener did. Wegener's theory became more widely known when his book was published in English in 1924. In the following 30 years or so, geophysicists, led by Harold Jeffreys, claimed that the forces postulated by Wegener were insufficient to cause horizontal movements of the continents, and, therefore, the theory should be forgotten. However, it was taken up by some European geologists in the 1920s, one supporter being the Swiss geologist Emile Argand, who made a 'spirited defence' at the first post-war International Geological Congress in 1922, discussing the 'Tectonics of Asia'. Argand's nappe theory, which was generally accepted by then (although general 'mobilism' was not), indicated that large horizontal movements (extending through a considerable thickness of crust) caused by strong lateral compression had occurred in parts of Europe and North Africa. However, he could not suggest how the forces required to cause these movements, or by the theory of continental drift, could be generated. He was unaware that a solution for the Alpine deformations had been suggested by Otto Ampferer in 1906 – the action of massive convection undercurrents within the upper mantle causing what Ampferer called 'subduction'.

In 1925 Ampferer told Wegener that these currents must be contributing factors in the mechanism of continental drift. Although it was accepted only in about 1950, Arthur Holmes, in the late 1920s, essentially explained how massive slow-moving convection currents could operate within the upper mantle (Figure 3). Holmes recognized that the radioactivity of continental rocks was generally greater than that of oceanic rocks, causing higher temperatures below the continents. This unequal heating of the mantle would cause ascending currents under a continental region, which would spread out at the top in all directions towards the cooler peripheral regions. The downward currents would be strongest beyond the continental edges. Thus the continental block could be ruptured, and portions could be carried apart on the backs of the currents.

While supported by some European geologists, Wegener's idea was rejected out-of-hand by the majority of North American geologists, and some

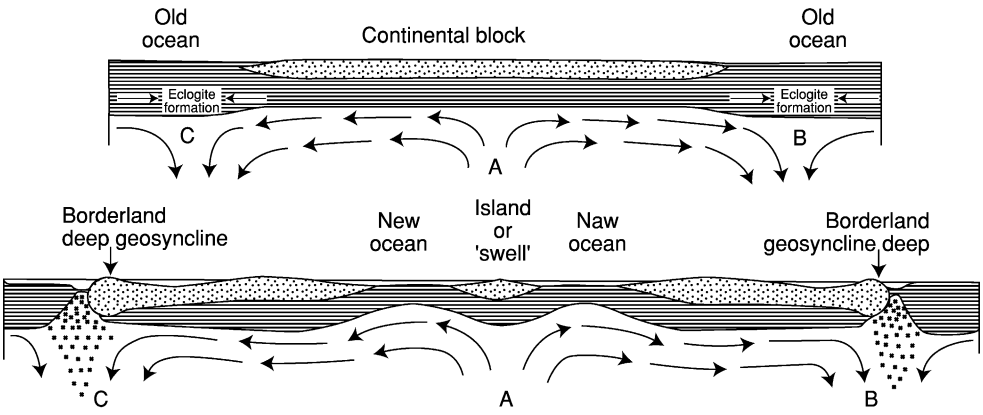


Figure 3 Holmes's convection model. The top figure shows a postulated system of convection currents in the mantle rising at A beneath continental crust and moving horizontally to edges of the continental block before descending at B and C (incidentally causing the formation of eclogite, because of high temperature and pressure). The lower figure shows the consequence of this movement. The horizontally-moving currents have split the old continental block into three, forming two new continents (and a minor remnant). The outer edges of the new continents are deformed during down-dragging. (Reproduced from Holmes A (1931) Radioactivity and earth movements. *Transactions of the Geological Society of Glasgow for 1928–29* 18: 559–606 [Figure p. 579]).

German geologists, particularly Stille, also opposed Wegener. It was left to southern-hemisphere geologists, particularly the South African Alexander du Toit, to accumulate evidence, such as the similar widespread Late Palaeozoic glaciations, Permian coals, and Mesozoic continental successions in the southern continents (and India), that substantiated Wegener's views, but this work interested few northern-hemisphere geologists. The idea of continental drift was kept alive into the 1950s by the South African Lester King and the Australian Warren Carey, who were not taken seriously, in the early 1960s, by many North American geologists. However Carey's symposium on continental drift, held in Tasmania in 1958, attracted attention at a time when ideas in many fields of geology were fomenting.

Ironically, one of the American scientists opposed to Wegener's ideas, William Bowie, was responsible for getting one of the pioneers of plate tectonics, Maurice Ewing, into ocean-floor exploration. In 1934, Bowie (with Richard Field) asked Ewing if he could study the outer edge of the continental shelf. Was it a basic geological feature, such as a fault, or just superficial – the result of outbuilding of sediment? Ewing said that the problem could be studied using seismic-refraction geophysical methods and proceeded to do so. Ewing's work in the ocean basins was to be crucial in resolving the continental-drift controversy.

Around 1900 the phenomenon of palaeomagnetism, and field reversals preserved in lavas, were recorded by Bernard Brunhes, but their significance was ignored until the 1920s. A key to change came with the publication of the 'pole-wander' studies of John Graham in 1949. These results were dismissed as 'impossible' by prominent mathematicians, but by 1955 Keith Runcorn and Paul Blackett provided evidence that the poles had indeed wandered during geological time. Runcorn maintained that movement of the magnetic poles relative to the crust had occurred, in contrast to Blackett's idea that portions of the crust had moved many thousands of kilometres. But there was a problem: the studies on different continents gave different sets of results. These sets could, however, be reconciled by rotating the continents, bringing the various sets into coincidence. There had to be something in the idea of continental drift. The resolution of the problem is discussed in **History of Geology Since 1962**.

The Inner Earth

An important breakthrough in determining the constitution of the Earth's interior and particularly the size of its proposed liquid core was made by Richard Oldham in 1906, using data from various earthquake records. Oldham showed that there were three separate records from a major earthquake, caused by three distinct forms of wave motion, propagated at different rates and along different paths, forming three distinct phases in the record of the earthquake at a distant point. The latest record was due to waves propagated on or near the Earth's surface. The other two were due to waves that had travelled through the Earth. Oldham indicated that the crust was thin and non-homogeneous and therefore did not transmit mass waves. Below this crust, the records suggested that there were two distinct zones, the outer three-fifths (mantle) differing from the inner two-fifths (core). 'The core appeared to be liquid. . . .shadow' and: 'The S waves (**Figure 4**) were delayed in arriving on the opposite side of the Earth'. Oldham, at the time thought that they travelled more slowly through the core for some unknown reason, but recognized that if the core were liquid this would prevent the S waves travelling through and they would be diverted

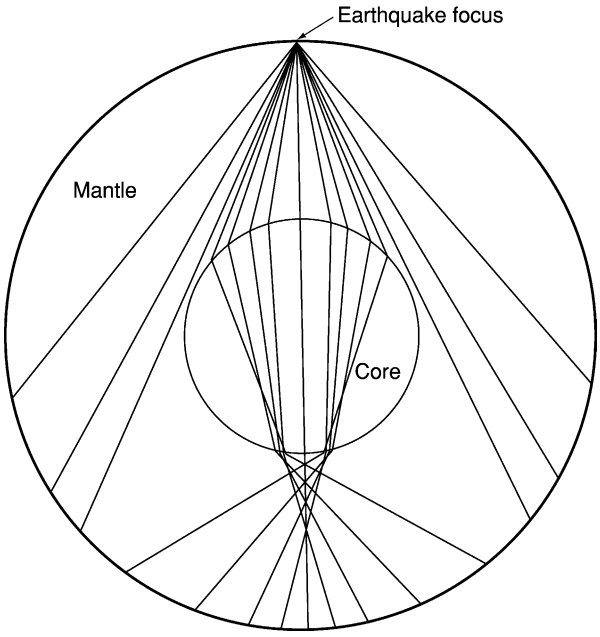


Figure 4 Richard Oldham's simplified figure showing the transmission of seismic waves through the Earth. The figure represents the deducted paths of what Oldham called second-phase waves (i.e. S waves) travelling through a central core occupying. 4 of the radius of the earth at half the original speed of propagation. He pointed out that 'there should be a zone, at about 140° from the origin where the second -phase waves would be so dispersed, and consequently so feeble, that it would amount to a shadow, and the second-phase should be absent in records from this distance, or [only] feebly marked.' (Reproduced from Oldham R (1906) The constitution of the earth as revealed by earthquakes. *Quarterly Journal of the Geological Society of London* 62: 456–475 [Figure p. 468]).

(refracted) around the core (and thus delayed). It was not until 1913 that he accepted the idea of a liquid core which caused the 'shadow' for S waves if the angular distance from the focus of the earthquake were more than 120°. Oldham and most other researchers were unaware that the Russian Leonid Leybenzon had suggested the same idea in 1911.

There then followed the recognition of the 'Moho discontinuity', discovered by Andriya Mohorovicic in 1909, separating, he thought, the crust from the mantle. However, what was the nature of this discontinuity? Was it the result of chemical or merely physical differences? Were the concepts of two separate crustal layers, the heavier 'sima' (silica and magnesium) and the lighter 'sial' (silica and aluminium), introduced by Suess, still valid? Ideas and terms remained in turmoil, as some, such as Bailey Willis, equated 'crust' with sial. Did such crustal divisions explain the major differences in magmas and the resultant intrusions and extrusions that could be observed at the surface, or were deeper levels involved in the evolution of surface geology?

Other workers suggested different models of the Earth's interior, and in about 1914 Joseph Barrell introduced the concept of the 'lithosphere' as a thin crust, underlain by the 'asthenosphere', a relatively weak 100 km or so, with an even weaker layer of unknown thickness beneath, in which adjustments to isostasy took place. Between 1940 and 1942 Keith Bullen divided both the mantle and the core into three separate zones, but there were still unresolved questions. In 1954, a conference 'The Crust of the Earth' was convened at Columbia University, New York, and many prominent geologists and geophysicists contributed, reaching some agreement on terms and directions for future research into the crust and mantle. However, the question of the Earth's inner composition remained largely unsolved. The increasing sophistication of geophysical equipment, particularly seismic, showed that the 'Moho' was not a continuous feature, but was 'replaced' in places, such as below mountain belts, by a disturbed 'mixed' zone. In the early 1950s, evidence from marine surveys showed that the ocean floors were covered in many areas by only a thin layer of sediments. With the increasing ability to drill deep within the Earth it was suggested in the late 1950s that some 'real evidence' of the nature of the Moho discontinuity could be obtained by drilling a hole 5 or 6 km through the crust on the seafloor. The 'Mohole project' was taken up with enthusiasm, but eventually failed because the ideas of the geologists involved differed from the approach advocated by the preferred drilling company, and the project was terminated in 1966.

While the records of earthquakes were used to study the Earth's interior, seismologists also devoted considerable efforts to predicting and classifying earthquakes. In the 1930s a number of seismologists began work at the California Institute of Technology. They included Harry Wood, Beno Gutenberg, Hugo Benioff, Frank Press, and Charles Richter, who were essentially the founders of modern seismology, although there were important researchers such as Fiusakichi Omori in Japan. Nevertheless the accurate prediction of earthquakes has not yet been achieved (*see* **Seismic Surveys**).

Impact Craters

While actualism (uniformitarianism) ruled during the early part of the twentieth century, there was still a suspicion that 'catastrophes' might have played a part in shaping the Earth. Late in the nineteenth century Grove Karl Gilbert argued for impacts as the origin of most of the Moon's craters. However, it was a long time before such features on the Earth's surface were accepted as being due to meteorites. Specimens from small craters at Henbury in Central Australia were described in 1932 and accepted as meteorites, but it was not until the 1950s that the larger famous Barringer (Meteor) Crater in Arizona was accepted as an impact crater, despite information obtained from drilling in the early 1900s. As early as 1942, Harvey Nininger argued that large impacts might have influenced the evolution of life. This idea was not really taken up until the 1980s, although when the Yucatán Puerto Chicxulub crater was investigated for oil in the 1950s the drill cores showed impact melt material, and, in 1970, Digby McLaren also talked about the idea. In the mid-1950s research on craters began in earnest, initially in two independent fields: photographic and field studies investigated the origin of craters and the possible significance of accretion of impact bodies in the formation of heavenly bodies; and the mode of formation of craters was investigated by nuclear and chemical explosions and detailed studies of shock-wave propagation. Important researchers in the former (geological) field included Ralph Baldwin, Robert Dietz, and Eugene Shoemaker; Donald Gault's experimental work on hypervelocity impacts was typical of the latter. By the late 1950s there was considerable interaction between these two groups, and the first international impact symposium was held in 1961. Pioneering work in the 1940s and 1950s by George Baker and others on the glassy meteorites (tektites), notably their shape, proved important in the study of the problem of re-entry by spaceships into the Earth's atmosphere.

The International Geophysical Year, 1957–1958

The years after the Second World War saw a rapid increase in air travel, which enabled researchers to meet, face-to-face, other workers with similar interests but different backgrounds, and a considerable amount of international collaborative work began. The International Geophysical Year, beginning in 1957, was a significant event. Following international cooperation in 1882–1883 and 1932–1933 in gathering data about the polar regions, it was first suggested that 1957–1958 should be the Third Polar International Year. However, it became much larger, as a result of the encouragement of the UN-supported International Union of Scientific Unions. Between July 1957 and December 1958 (a period of high solar activity) 13 scientific programs investigated the Earth, the atmosphere, and space, and their interrelations. At least 60 nations participated, and thousands of scientists were involved, most undertaking national problems, but there was, as in the earlier 'Years', a concentration on Antarctica and the Arctic, with new work along the equator and on selected meridians. Tuzo Wilson, President of the International Union of Geodesy and Geophysics, played a considerable part in helping to coordinate the results of the research. One of the best-known events of the International Geophysical Year was the launching of *Sputnik*, which produced significant scientific data and also inspired the 'space race'. A second, and happier, result was the Antarctic Treaty to protect that continent. The 18-month 'year' provided the impetus for work that has continued to the present.

Summary

The period considered here saw many advances. While broad concepts were initially tied to the idea of a generally ordered Earth with fixed continents and oceans, facts that were accumulating hinted at the need for major changes in thought. The increased ability to measure, thanks to much improved technology, meant that rapidly acquired data could be quickly analysed and disseminated through more easily accessible scientific media. Coming towards the end of the period, the International Geophysical Year was the culmination of a long phase of gestation. The plate-tectonics revolution could not have occurred until many years later were it not for the accumulation and analysis of more than 60 years of data. There was conservatism, and new ideas took time to be accepted. Nevertheless, much was achieved.

See Also

Analytical Methods: Geochronological Techniques; Gravity. **Diagenesis, Overview**. **Famous Geologists:** Du Toit; Walther; Wegener. **History of Geology Since 1962**. **Microfossils:** Palynology. **Seismic Surveys**. **Time Scale**.

Further Reading

Davis WM (1954) *Geographical Essays*. New York: Dover Publications.

Ginsburg RN (ed.) (1973) *Evolving Concepts in Sedimentology*. Baltimore: Johns Hopkins University Press.

Good G (1998) *Sciences of the Earth: An Encyclopedia of Events, People and Phenomena*. New York: Garland.

Holmes A (1944) *Principles of Physical Geology*. London: Thomas Nelson.

Lewis CLE and Knell SJ (eds.) (2001) *The Age of the Earth: From 4004 BC to AD 2002*. London: The Geological Society.

Mather KF (ed.) (1967) *Source Book in Geology 1900–1950*. Cambridge, MA: Harvard University Press.

Muir Wood R (1985) *The Dark Side of the Earth*. London: Allen and Unwin.

Müller DW, McKenzie JA, and Weissert H (eds.) (1991) *Controversies in Modern Geology: Evolution of Geological Theories in Sedimentology, Earth History and Tectonics*. London: Academic Press.

Oldroyd DR (1996) *Thinking about the Earth: A History of Ideas in Geology*. London: Athlone.

Oldroyd DR (ed.) (2002) *The Earth Inside and Out: Some Major Contributions to Geology in the Twentieth Century*. London: The Geological Society.

Pettijohn FJ (1984) *Memoirs of an Unrepentant Field Geologist*. Chicago: University of Chicago Press.

Umbrove JHF (1947) *The Pulse of the Earth*. The Hague: Nijhoff.

Wertenbaker W (1974) *The Floor of the Ocean: Maurice Ewing and the Search to Understand the Earth*. Boston: Little, Brown and Company.

White JF (ed.) (1962) *Study of the Earth: Readings in Geological Science*. Englewood Cliffs: Prentice-Hall.

HISTORY OF GEOLOGY SINCE 1962

U B Marvin, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

Introduction

The theory of plate tectonics postulates that the Earth's outermost layer, the lithosphere, which consists of the crust (*see* **Earth:** Crust) and the uppermost mantle, is about 100 km thick and is broken into rigid plates that slowly move and change their configuration in response to thermal instabilities in the mantle. The theory holds that the ocean floors are youthful due to their continual creation at spreading ridges and destruction as they plunge back deep into the mantle (*see* **Earth:** Mantle). The continents, in contrast, are buoyant sialic blocks with components of all ages, which ride passively on the surfaces of the plates. This theory, established in 1968, has unprecedented power for making quantitative calculations of past crustal motions and predictions of future ones. The immediate precursor of plate tectonics was the hypothesis of seafloor spreading proposed separately by two scientists in 1960 and 1961. Seafloor spreading was based primarily on information provided by geophysical explorations of the ocean basins during and after World War II. This article reviews the discoveries that led to plate tectonics, and the changing climates of opinion that led to a new era in the geosciences.

Post-war Explorations of the Ocean Basins

The geology of the ocean basins, which occupy more than 60% of the Earth's surface, remained largely unknown until after World War II. Despite the early hypothesis of continental drift, a majority of geologists regarded the ocean basins, particularly the Pacific Basin, as primordial features, formed when the Earth's crust first cooled. In 1946 the Dutch geologist, Philip Kuenen, calculated that if the Earth is 3 billion (3×10^9) years old, the ocean floors should be covered with a layer of sediments 5 km thick. In 1956, when radiometric dating showed the Earth to be ~4.55 billion years old, many expected the sediments to be much thicker. Indeed, as late as 1958, some of the scientists who were planning Project Mohole, to drill to the Mohorovičić discontinuity through the Pacific floor, anticipated that the cores would reveal the entire sequence of sediments deposited during the Lipalian interval from the end of the Precambrian to the beginning of the Cambrian Period – a time of severe erosion on the continents.

In the latter 1940s, the United States Navy provided generous funding for oceanographic research including ships, state-of-the-art equipment, laboratories, and scientists. Two of the four institutions credited with founding plate tectonics began sea-going explorations: the Scripps Oceanographic Institute at the University of California in San Diego, and the Lamont Geological Observatory (now the Lamont-Doherty Earth Observatory) at Columbia University in New York. The other two institutions contributing to the establishment of plate tectonics were the Department of Geophysics at Cambridge University in England, and Princeton University.

Preeminent scientists directed both of the oceanographic institutes: Roger Revelle at Scripps, and Maurice Ewing at Lamont. Both began intensive programmes of mapping the submarine topography by echo depth soundings, displaying the layering of the ocean floors by seismic profiling, and collecting samples by piston coring. Scientists aboard Scripps vessels also measured heat-flow values, while those from Lamont focused more on measuring gravity at sea. Both organisations towed magnetometers behind their ships. At the beginning, these scientists were not testing hypotheses; they were performing the first comprehensive investigations of the last unknown domain of the Earth. There were many surprises in store for them.

The ship *Horizon*, which left Scripps in July 1950, with Revelle and several other now-famous scientists aboard, made its first discovery 300 miles south-west of San Diego. There, in deep water, its companion Navy ship fired the first explosives and *Horizon*'s seismic profiler showed that the Pacific sediments were not 5 km or more thick: they were only 260 metres thick. The profiler detected three layers in the ocean floor: Layer 1, consisting of sediments ~0.25 km thick; Layer 2, of consolidated sediments or volcanics or both, 1–2 km thick; and Layer 3, the authentic bedrock of the ocean floor, which ultimately was shown to maintain a uniform thickness of ~4.5 km throughout the ocean basins.

From the beginning, the echo soundings showed the ocean floor to be unexpectedly crowded with hills, 10–30 or more km long, 2–5 km wide, and 50–500 m high. Further explorations showed that these so-called 'abyssal hills' occupy more than 30%

of the floors of all oceans, placing them among the Earth's most abundant topographic features.

Revelle and Arthur Maxwell measured heat-flow from the ocean floor and observed, to their astonishment, that the values averaged within 10% of those from the generally granitic and consequently more radioactive continents. For want of other possibilities, they suggested that the heat in the ocean floor must be carried there by rising limbs of convection cells in the mantle, an idea compatible with the earlier 'drift' model proposed by Arthur Holmes in the1920s.

In the mid-Pacific, *Horizon* dredged the surfaces of guyots – volcanic cones with flat summits, submerged 1–2 km below sea-level. These curious features had been discovered and named by Harry Hess, of Princeton University, who located 160 of them while he was on naval duty during World War II. At first, Hess assumed they were Precambrian islands, erupted long before reef-building corals evolved, that had been eroded at sealevel and then drowned by the rising of water due to the deposition of 2–3 km of oceanic sediments. The rare opportunity to study and describe the dredged materials fell to Edwin Hamilton, to whom Robert Dietz, on the *Horizon*, had assigned this topic for his PhD thesis. The hauls brought up fragments of Cretaceous corals, only about 100 my old. Suddenly, the 'ancient' ocean basins were seen to be much younger than the continents.

In the final days of the cruise, the *Horizon* steered a course to allow the youthful Henry Menard, of the US Naval Electronics Laboratory, to determine the nature of the Mendocino escarpment, a long narrow feature, striking westward from Cape Mendocino in California. On a S–N traverse of ~60 km, the echo-sounder showed that, starting in deep water the ship passed over a low ridge tilted to the south, a narrow trough, a steep scarp 18 km high, a high ridge tilted to the north, a low swale, and then it arrived in relatively shallow water. This 'fracture zone' was a new class of submarine structure of which Menard and others would find several similar examples farther south in the eastern Pacific Ocean. They appeared to be strike-slip faults with a component of normal faulting (*see* **Tectonics:** Faults), but they terminated before slicing into the American continents. Eventually, fracture zones, long and short, would be found in all the ocean basins.

Meanwhile, Lamont ships had begun producing the earliest detailed topographic charts of the mid-Atlantic Ridge and its linkage to what would prove to be a world-encircling succession of submarine mountain ranges, with various spurs, nearly 60 000 km long, 1–3000 km wide at the base, and 2 km high, with peaks rising to 4 km above the ocean floor. In 1956, Maurice Ewing and his colleague, Bruce Heezen, reported finding a deep rift valley (*see* **Tectonics:** Rift Valleys) occupying the ridge crests. Subsequently, the rifts proved to be discontinuous but, where present, they were zones of shallow-focus seismicity, strong magnetic anomalies, and higher than average values of heat-flow. The ridges are offset horizontally into segments by transverse fracture zones, which were assumed to be strike-slip faults.

The Navy prohibited publication of contoured bathymetric charts, but two Lamont scientists, Heezen and Marie Tharp, began drawing 'physiographic diagrams' showing how the ocean floors would look with the waters drained away. These diagrams, published in the 1960s, astonished both non-scientists and scientists, who found them remarkably useful for interpreting the structures and history of the ocean basins. Heezen and Tharp both won gold medals for their work.

The East Pacific Rise (or Ridge) is unique in being a huge bulge (nearly 13 000 km long, 3000 km broad, and nearly 4 km high) with no sharp crest, no central rift, and no mid-ocean position – it lies close to the margins of the American continents. In 1960, Menard proposed that convection currents had moved the northern portion of the Rise beneath western North America where its presence would account for the high plateaus of Mexico and Colorado and the Tertiary Basin and Range topography of Utah and Nevada. It followed that the floor of the eastern Pacific Ocean, cut by the great E–W fracture zones, was the western flank of the buried Rise.

Many of the observations, made in the 1950s, suggested youthful, mobile ocean floors, but contrary explanations were offered in each case, puzzling scientists who were trying to piece together the dynamic history of the Earth. Then in 1960 and 1961, Hess and Dietz independently but simultaneously, proposed hypotheses that addressed global dynamics in terms of moving seafloors driven by convection in the mantle.

Sea-floor Spreading: 1960 and 1961

In 1960, Hess circulated the preprint of a book chapter in which he described the ocean floors as the exposed surface of the mantle. He proposed that large-scale convection cells in the mantle create new ocean floor at the ridges, where the rising limbs diverge and move to either side until they cool and plunge down into the mantle at the trenches (*see* **Tectonics:** Ocean Trenches) or beneath continental margins (*see* **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges). He now explained guyots as volcanic peaks eroded on oceanic ridge crests and carried to the depths by the moving seafloor. Where

limbs rise beneath continents they split them apart and raft the granitic fragments in opposite directions until grounding them over zones of down-welling. Hess suggested that inasmuch as granite is too buoyant to sink into the mantle, some fragments of continents have survived since the beginning of geologic time, while the ocean floors have been swept clean and replaced by new mantle material every 300–400 my.

Hess further theorized that the mantle consists of peridotite, an olivine-rich rock that becomes hydrated to serpentinite at the ridges by reaction with heated waters released from depth. He favoured serpentinite partly because of its ease of recycling and partly because he believed it would be impossible to achieve the uniform 4.5 km thickness of the ocean floors with dykes and lava flows. Hess acknowledged that mantle convection was regarded as too radical an idea to be widely accepted by geologists and geophysicists, but he pointed out that his model would account for many phenomena in a coherent fashion: the formation of the ridge-rift system and the trenches, the youth and uniform thickness of the ocean floors, the thinness of pelagic sediments, and moving continents. Hess called his chapter an essay in 'geopoetry'.

In 1961, Dietz published a three-page article in *Nature* in which he, too, argued that convection in the mantle, moving at a few centimetres per year, could produce the overall structure of the ocean basins: the ridges form over sites of rising and diverging limbs; the trenches form at sites of converging and down-welling limbs, and the fracture zones are shears between regions of slow and fast creep. Dietz suggested that the mantle consists of eclogite, a dense pyroxene-garnet rock, and that the ocean floors are built of basaltic dykes and pillow lavas, formed by the partial melting of the eclogite at the ridges. But he added that the actual rock compositions were of less importance than the fact that the ocean floors must be recycled mantle material. Dietz pointed out, as had Hess, that rising convection currents rift continents apart and carry the sialic fragments *en bloc* to sites of down-welling and compression, where folded mountain ranges form on their margins. He added that the pelagic sediments ride down into the depths on the surfaces of the plunging oceanic slabs where they are granitized and welded onto the undersides of the continents – thus contributing to the persistent continental free-board, despite steady erosion of the continental surfaces toward base level.

Dietz called this process 'sea-floor spreading'. This concept, he argued, requires geologists to think of Earth's outer layers in terms of their relative strengths, so we should begin referring to the 'lithosphere', a historic name for Earth's outermost layer, which is relatively strong and rigid to a uniform depth of about 70 km (now generally taken as 100 km) under both continents and oceans. Beneath the lithosphere lies the weaker, more yielding, 'asthenosphere' on which the lithosphere moves in response to convection currents. The asthenosphere had previously been hypothesized on the basis of seismic evidence.

Dietz made the first suggestion (later confirmed) that the oceanic abyssal hills are a chaos topography, developed when strips of juvenile sea floor have ruptured under stress as the floors move outward. Finally, he referred to two papers *in press*, one by Victor Vacquier *et al*., and one by Ronald Mason and Arthur Raff, reporting the discovery of linear magnetic anomalies on the Pacific floor. Some of the magnetic lineations appeared to be offset by up to 1185 km along the Mendocino fracture zone. Dietz suggested that the lineations are developed normal to the direction of convective creep of the ocean floor. Noting that neither the fracture zones nor the magnetic lineations impinge upon the continental margins, he suggested that both lost their identity as the Pacific floor slipped beneath the American continents.

Rarely are ideas subsequently seen as basic to a grand new system of thought in science, appreciated at full value when they first appear. That was the case with sea-floor spreading, which failed to catch the attention of more than a few readers at a time when most geologists and geophysicists were unprepared to take seriously the idea of mantle convection, much less that of the new notion of spreading seafloors. However, Dietz's paper elicited a favourable letter to *Nature* from J Tuzo Wilson at the University of Toronto, and it inspired Ewing to redirect a large portion of Lamont's research into testing the seafloor spreading hypothesis. Ewing outfitted two ships with upgraded seismic reflection profilers to measure the thickness of pelagic sediments across the oceans. He was to find virtually no sediments on the ridges and a modest thickening of the layers towards the edges of the continental shelves. Ewing took the earliest photographs of deep sea sediments and observed ripple marks, which, until then, had been used as diagnostic of shallow waters.

In the next few years, as data favourable to seafloor spreading accumulated, a regrettable negative reaction developed toward Dietz. Even though his model differed from that of Hess, the belief spread that he had 'stolen' Hess's basic idea and rushed it into print under his own name. This impression still persists to some degree, even though Dietz's choice of basaltic rather than serpentinized ocean floors is the one universally accepted today.

In 1986, Menard reviewed this controversy in his *The Ocean of Truth*, a personal account of the

sequence of events that led to plate tectonics. Menard, who had refereed the pre-publication manuscripts of both Hess and Dietz, wrote that he felt certain – and said so at the time – that Dietz was unaware of Hess's preprint when he wrote his own paper. Nevertheless, Menard believed that only one person can have priority for an idea and it should be Hess, so he urged Dietz, if only for appearances sake, to add a footnote to his next publication on seafloor spreading conceding credit for the idea to Hess. Dietz did so, and the two men made their peace in print. Subsequently, Menard came to realise that, in fact, more than one person can be struck with the same idea, particularly at a time when new data are coming in and are being freely discussed. He cited several additional examples that took place during the race towards plate tectonics.

Finally, Menard remarked that the priority for this idea probably should go to Arthur Holmes in Britain, who had favoured convection in the mantle as a ruling factor in global tectonics from the late 1920s until he died in 1965. Holmes described a basaltic layer which becomes a kind of endless travelling belt as it moves from ridges to trenches carrying continental fragments along with it. Over the years, Holmes changed his diagrams somewhat, but they all show the limbs of a convection cell rising under a continental slab, which is stretched and pulled apart into two fragments that are rafted to either side leaving behind new ocean floors. Holmes did not depict the creation of new ocean floor at a spreading oceanic ridge; he showed the ridges as being wholly or partially sialic. In that important respect his idea differed from those of Hess and Dietz. Nevertheless, Holmes's basic model was, without question, a predecessor of sea-floor spreading.

History books do not necessarily aid us in resolving disputes. In 1973, Allan Cox of the US Geological Survey omitted Dietz's article from his collection of the landmark papers that led to plate tectonics. And in the entry on Hess in Volume 17 of *The Dictionary of Scientific Biography*, published in 1990, we read that his hypothesis of sea-floor spreading was the most important innovation leading to plate tectonics, but that it was given its name by Dietz who, "with Hess's preprint in hand", published the first article on it in 1961. Fortunately, others knew better. In 1966, the Geological Society of America would present its highest honour, the Penrose Medal, to Hess for his research on the petrology of ultramafic rocks and for his provocative tectonic hypotheses including that of the spreading ocean floor. In 1988, the Penrose Medal went to Dietz for his "world-class, innovative contributions in three divisions of the geosciences: sea-floor spreading, recognition of terrestrial impact structures, and the meteorite impact of the Moon's surface".

Magnetic Anomalies on the Ocean Floors: 1961, 1963

In August, 1961, the paper by Mason and Raff, of which Dietz had seen a preprint, was published by the Geological Society of America along with a companion paper by Raff and Mason that included a particularly striking diagram showing long, narrow, mostly vertical black and white 'zebra stripes' of alternating high and low magnetic intensities measured on the floor of the north-eastern Pacific Ocean (**Figure 1**). The authors suggested that the contrasting intensities might reflect structural ridges and troughs or a system of sub-parallel dykes, but subsequent topographical and gravity surveys failed to detect either one. This research project, which proved to be momentous, was undertaken when Mason, a visitor to Scripps, casually asked a seismologist over morning coffee if anyone had thought of towing a magnetometer behind a ship to gather data while the ship was engaged in other operations. Revelle overheard the question and offered Mason the assignment, then and there. Mason soon learned that Lamont already had towed a magnetometer behind a ship in the Atlantic, and he could borrow it pending acquisition of an instrument by Scripps.

One reader, Lawrence W Morley of the Geological Survey of Canada, who had conducted extensive aeromagnetic surveys over lands and seas, remained mystified by this pattern for nearly two years until he discovered Dietz's paper on sea-floor spreading. Morley immediately wrote a short article and submitted it to *Nature* in February 1963, proposing a test of sea-floor spreading. He argued that the magnetic stripes form at the crests of spreading ridges where the erupting lavas acquire the magnetization of the Earth's ambient field (a detail Dietz had not specified). Today, the Earth's magnetic field is north-seeking, but in the 1950s it had become clear that some rocks have cooled during periods of south-seeking polarity, so Morley proposed that the parallel stripes on the moving sea-floor record the periods of normal and reversed polarity. And, inasmuch as Cretaceous rocks were the oldest yet recovered from the ocean floors, he speculated that such stripes have recorded the history of the ocean basins for the past 100 my or so. Morley tentatively calculated the rates of spreading and lengths of reversal periods. *Nature* rejected his paper saying they had no space available. Morley then sent it to the *Journal of Geophysical Research*, which kept it for some time and then rejected it, with a message from one referee saying that such speculation was more appropriate to cocktail party chatter.

On 7 September 1963, *Nature* published the now famous paper by Frederick Vine and Drummond

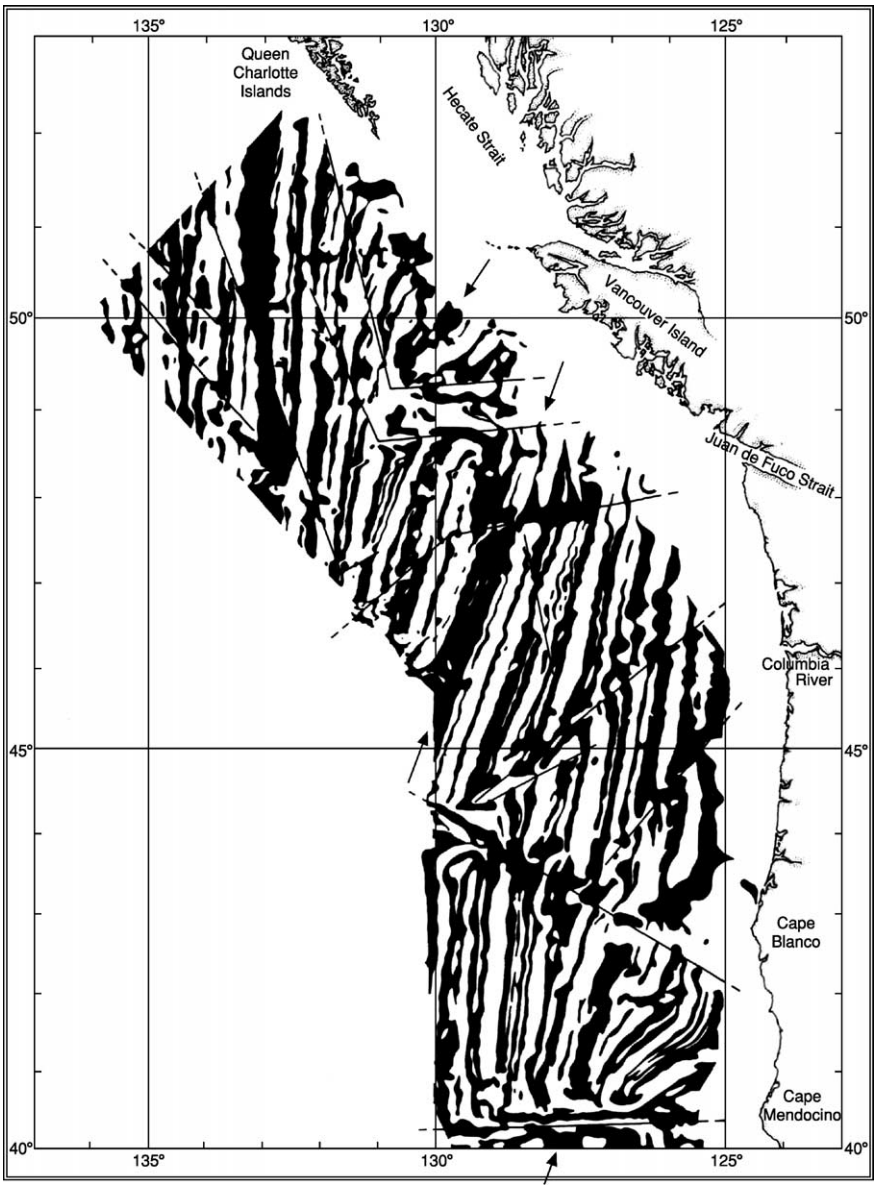


Figure 1 Magnetic anomalies offshore from Cape Mendocino, California, to the north of Vancouver Island, British Columbia. In 1961, Raff and Mason interpreted these stripes as most likely due to high and low magnetic intensities in linear patterns of mafic oceanic rocks. They drew the lettered lines as fault traces. In subsequent years, such 'zebra stripes' would be seen as evidence of magnetic field reversals with normally magnetized rocks (black) alternating with reversely magnetized ones (white). (Reprinted by courtesy of Raff and the Geological Society of America.)

Matthews of Cambridge University: 'Magnetic Anomalies over Oceanic Ridges'. Like Morley, Vine and Matthews assumed that convection, sea-floor spreading, and reversals of magnetic polarity all occur, and that the ocean floors consist of basalt that becomes strongly magnetized at the ridge crests. Today, the Vine–Matthews paper is widely seen as the founding paper of plate tectonics, but at the time it was poorly received and largely ignored for the next three years. It was not even included by their Department of Geophysics at Cambridge University in its list of important contributions for 1963! Years later, when attitudes changed, and Morley's story came out, the Earth science community began to speak of the Vine–Matthews–Morley (VMM) hypothesis. Morley then realized that he had gained more fame by having his paper rejected than he would have by its publication.

In 1963, all three of the basic assumptions listed by Vine and Matthews were suspect to geoscientists. Professor Harold Jeffreys, at Cambridge University, argued in every edition of his book, *The Earth*, beginning in 1924, that the mantle of the contracting Earth

is too stiff to allow for convection. In his sixth and final edition of 1976, he added several pages arguing against plate tectonics. In 1963 and again in 1964, Gordon MacDonald, of the Institute of Geophysics and Planetary Physics of the University of California, published a detailed paper, 'The Deep Structure of the Continents', in *Reviews of Geophysics* and a short version of it in *Science*. He intended this paper as a death-blow to continental drift and sea-floor spreading. MacDonald argued that isostasy prevails; thousands of measurements made on the surface and by satellites show that gravity over continents is equal to that over oceans – despite marked difference in their compositions and densities. This, together with the equal values of heat-flow determined worldwide indicated that the continents must have formed by vertical segregation of the mantle directly beneath them, which, therefore, must differ from the mantle under ocean floors to depths of 400 to 700 km. MacDonald concluded that no significant horizontal motion, due to convection or any other process, has occurred.

Magnetic Field Reversals, Isotopically Dated: 1964

Since early in the twentieth century, many advances had been made in techniques of measuring the remanent (permanent) magnetization of rocks and interpreting the results. In the 1950s, polar wandering curves had yielded evidence in support of continental drift. Nevertheless, inasmuch as Earth's magnetic field varies in alignment, intensity, and polarity, many geoscientists still viewed palaeomagnetism as an incomprehensible property, studied by 'black box' techniques, and scarcely to be trusted. However, by the early 1960s, measurements had begun to show that whereas recent basaltic lavas from widely spaced sources behave as north-seeking compasses, those of Early Pleistocene age behave as south-seeking compasses. Clearly, reversals of the magnetic field have taken place and the search was on for reliable means of dating them.

In June 1964, Allan Cox, Richard Doell, and Brent Dalrymple of the US Geological Survey at Menlo Park, California, published a paper entitled, 'Reversals of the Earth's Magnetic Field', in which they reported that they had dated rocks of normal and reversed magnetism by the K/Ar method. Their results documented the occurrence of two epochs of normal polarity and one of reversed polarity during the past 3.5 my. They named the epochs in honour of pioneers of palaeomagnetic studies (*see* **Palaeomagnetism**): the Bruhnes normal epoch, from the present to 1.0 my ago; the Matuyama reversed epoch from 1.0 to 2.5 my ago; and the Gauss normal epoch from 2.5 to 3.4 my ago. Both the Matuyama and the Gauss epochs were interrupted by short 'events' of opposite polarities – named, respectively, the Olduvai and the Mammoth events – each lasting 250–300 000 years. Later, in February 1966, Doell and Dalrymple revised this time-scale by shortening the present Bruhnes normal epoch to 700 000 years and adding the newly discovered 'Jaramillo normal event', which occurred about 0.85 my ago in the early part of the Matuyama reversed epoch. Since then, additional epochs and events have been dated for the past 180 my, back to the mid-Jurassic Period.

Cox and his group dated their basaltic samples using a mass spectrometer of the type designed in the late 1950s by John Reynolds of the University of California at Berkeley, which opened a new age in isotopic geochemistry. Reynolds' all-glass instruments could be heated to evacuate all traces of atmospheric argon in order to yield accurate measurements of the small amounts of argon produced by the radioactive decay of potassium. The work by Cox and his colleagues persuaded many scientists that palaeomagnetism must be taken seriously. Once again, however, another group was doing the same research. Within weeks of its appearance, *Nature* published 'Dating Polarity Geomagnetic Reversals' by Ian McDougal and Donald Tarling, who had established a palaeomagnetism laboratory at the Australian National University, and were friendly rivals of the group at Menlo Park.

Transform Faults: 1965

In June 1965, *Nature* published a paper by Wilson entitled: 'A New Class of Faults and their Bearing on Continental Drift', in which he proposed what he called 'transform faults' as a test of sea-floor spreading. These are faults with large horizontal movements that appear, along with their seismic activity, to terminate abruptly. He wrote that instead of just terminating, these faults are 'transformed', at each end, into mid-ocean ridges (*see* **Tectonics:** Mid-Ocean Ridges), mountain ranges, trenches, or island arcs (*see* **Tectonics:** Ocean Trenches), which together make up the network of mobile belts that divide the Earth's surface into large, rigid plates. Wilson cited the San Andreas as a transform fault on land with its southern end connected to the oceanic ridge in the Gulf of California and its northern end to one north of Cape Mendocino. But his clearest examples of transform faults were those that connect offset segments of spreading oceanic ridges. As shown in Figure 2, all the slabs of ocean floor on each flank of a ridge move side-by-side down-slope and across the seafloor. Nevertheless, along the transform fault plane connecting the ridge crests, and only there, the

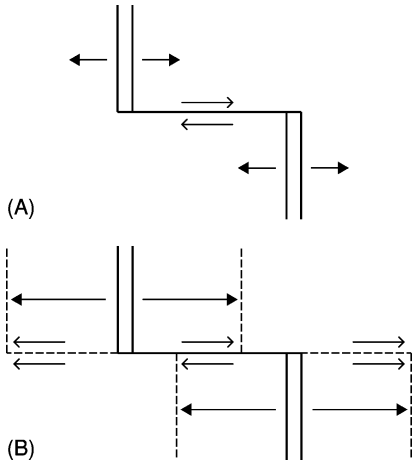


Figure 2 An evolving transform fault (A) Two expanding ridges are connected by a transform fault along which rocks moving down the eastern slope of the upper segment move in the opposite direction from those moving down the western slope of the lower segment. (B) After a lapse of time, spreading has advanced but the offset of the ridge has not changed, and movement along the fault still is confined to the ridge-ridge section. The dashed extensions of the fault have become aseismic fracture zones. Note that along a transform fault, the hot eruptive zones on the ridge crests lie adjacent to cold walls of older, subsiding rocks on the flanks. (After Wilson (1965) combined by Mason in Oreskes, 2001; reproduced by permission of Oreskes to Mason and Westview Press.)

rock walls move in opposite directions. Wilson noted that in 1963 Lynn Sykes at Lamont had shown that along the fracture zones cross-cutting the ridges, seismic activity is strictly limited to these transform fault planes. Wilson attributed this to the relative motions of the rock walls and also to the marked differences in temperature, elevation, and age of the rocks across the fault – where hot rock beneath a ridge is juxtaposed with a cold wall of much older seafloor. How did the ridges become offset into segments? Wilson and Sykes both argued that the offsets formed in place when the lithosphere first split open at the beginning of the present cycle of ridge building. No strike-slip faulting separated them, nor has there been any along the fracture zones, which now were recognized as topographic continuations of transform faults that became aseismic when spreading carried them beyond the ridge crests.

Wilson's paper led to a major change in the history of structural geology. Its reception was mostly positive as geoscientists looked at their problems from the new perspective. But in March 1965, Alan Coode, at Newcastle University, had sent a short note illustrated with a block diagram of ridge segments connected by a transform fault (although he gave no new name to it) to the *Canadian Journal of Earth Sciences*. That journal appears quarterly, so Coode's article was published, to little notice, a week or two after Wilson's paper appeared in *Nature*. Great honours would be showered on Wilson; none on Coode, although he had submitted his paper first.

The *Eltanin* Profile: 1966

Sea-floor spreading got off to a slow start partly because neither the Vine–Mathews nor the Mason–Raff papers showed obvious symmetry of linear magnetic anomalies across oceanic ridges. Early in 1966, James Heirtzler, Xavier Le Pichon, and J G Baron at Lamont published the results of an aeromagnetic survey over the Reykjanes Ridge, south of Iceland, which yielded a beautifully symmetrical pattern of magnetic stripes across a spreading ridge. They computed the spreading rate at slightly less than 2 cm per year of separation.

In December 1966, Walter Pitman III, and Heirtzler, at Lamont published the magnetic profiles recorded by the research vessel, *Eltanin*, during four passes over the Pacific–Antarctic Ridge, south of Easter Island. All four profiles gave similar results but the most southerly one, *Eltanin*-19, was so spectacularly symmetrical, both in the topography of the ridge and its record of magnetic reversals that it seems to have prompted mass conversions to sea-floor spreading, first at Lamont and then elsewhere. The profile showed each of the dated magnetic epochs of the past 3.4 my, for which it yielded a computed spreading rate of 9 cm per year of separation. Assuming a constant spreading rate within 500 km on either side of the ridge, the profile made it possible to extend the series of geomagnetic reversals back to 10 my. Pitman and Heirtzler documented a good match between the anomalies in the South Pacific and those on the Reykjanes Ridge, adjusted for the slower spreading rate in the Atlantic. Today the *Eltanin*-19 profile is ranked as one of the most important pieces of evidence in the history of geophysics. And further confirmation was at hand. Approaching the problem by an independent method, Neil Opdyke and his colleagues at Lamont, plotted the magnetic polarities of fossiliferous strata from deepsea cores of the South Pacific floor. Early in 1966, they found a definitive match with the dated record in the basaltic oceanic bedrock.

In February 1966, Vine visited Lamont where Heirtzler gave him a copy of the *Eltanin*-19 profile. Vine incorporated it into his paper 'Spreading of the Ocean Floor: New Evidence', which appeared in *Science* the following December, shortly after one by Pitmann and Heirtzler. In it, Vine presented six symmetrical profiles of magnetic anomalies across

ridges in the Atlantic, Indian, and Pacific Oceans, and showed that the linear anomalies on the East Pacific Rise match those on the Juan de Fuca Ridge offshore from British Columbia, even though they lie 11 000 km apart. His computed spreading rates for all the ridges ranged from about 2.0 to 3.0 cm of separation per year in the Atlantic and Indian Oceans to 8.8 cm per year across the East Pacific Rise. Vine speculated that the whole history of the ocean basins in terms of ocean-floor spreading must be 'frozen-in' as paired magnetic anomalies in the oceanic crust. Meanwhile, Vine had given a summary of his results in November at the annual meeting of the Geological Society of America, where it startled many geologists with their first serious introduction to sea-floor spreading.

A Matching of Continents: 1966

At the same GSA meeting, Patrick Hurley, a geochronologist at the Massachusetts Institute of Technology, presented a paper, with nine co-authors, offering radiometric evidence for continental drift. After attending the Symposium on Continental Drift, organised in 1964 for the Royal Society by Patrick Blackett, Edward Bullard, and Keith Runcorn, Hurley said: "I went to London a fixist and came home a drifter". Bullard had displayed a map that had been programmed by a computer to apply Euler's theorem to find the best fit of the continents across the Atlantic. The map showed an especially close fit between Brazil and Ghana, where Hurley knew that a sharp contact had been mapped between two rock provinces that were 600 million years old and 2000 million years old, respectively. If Bullard's map were applicable, the same two provinces should occur near São Luis at the easternmost tip of Brazil. Hurley gathered a group of Brazilian and other collaborators who obtained rock samples from the region of São Luis and dated them. The results showed an excellent match between the petrology and dates of the two rock provinces now on opposite sides of the Atlantic. Hurley's research was crucial in persuading many American geologists to accept continental drift *via* seafloor spreading. In America, the doctrine of the permanence of continents and ocean basins, founded by James Dwight Dana at Yale in 1846, finally expired in 1967.

Plate Tectonics: 1967–1968

Of the many papers published during the emergence in 1967 and 1968 of plate tectonics from sea-floor spreading, four of them often are cited as milestones. The earliest of the four, 'The North Pacific: An Example of Tectonics on a Sphere', by Dan MacKenzie and David Parker, applied the concept of transform faults to motions on a sphere on which aseismic areas move as rigid plates. They, too, applied Euler's theorem (with which Wilson acknowledged he had been unfamiliar when he wrote of the movement of rigid plates) to the effect that any displacement of a plate on a spherical surface may be considered as a rigid rotation about a fixed vertical axis. It follows that the slip vectors of moving plates must describe small circles around the Euler pole. McKenzie and Parker used fault-plane solutions for earthquakes in the North Pacific to work out the actual directions of the slip vectors and found that they wholly conformed to the sea-floor spreading model.

The remaining three papers, all of which review the evidence in broad fields, appeared early in 1968, in Volume 73 of the *Journal of Geophysical Research*. In his paper entitled 'Rises, Trenches, Great Faults, and Crustal Blocks', Jason Morgan at Princeton presented the first depiction of the entire surface of the Earth divided into rigid blocks, each having three types of boundaries: ridges where new crust is formed; trenches or mountain ranges where crust is destroyed by sinking or shortened by folding; and great faults. His diagram showed twenty rigid blocks, large and small, identified by the seismicity at their boundaries.

Morgan's interpretation required that the seafloors must move as rigid blocks, for which the thin oceanic crust lacks the required strength. Therefore, he envisioned the outer layer of the Earth as a rigid 'tectosphere' about 100 km thick, sliding over the weak asthenosphere. He argued that the location of an oceanic ridge is determined, not by the activity of some deep-seated system of thermal convection, but by the motions of the blocks themselves: wherever tensional forces fracture the tectosphere to a depth of 100 km, hot mantle material wells up between the blocks and serves as a zone of weakness. Subsequently, each new intrusion of mantle material is injected into the centre of the most recent one, thus maintaining a median position for a spreading ridge in most cases. He argued that the emplacement of new ocean floor would be reflected in the symmetry of its magnetic anomalies and from these he charted the history of motions that brought the blocks to their present positions.

In an article titled: 'Sea-floor Spreading and Continental Drift', Xavier Le Pichon, then a visitor at Lamont, asked whether large crustal blocks do, in fact, move for significant distances without any deformation of the sea-floor sediments or the continental interiors. He adopted a simple model of six large

blocks (rather than Morgan's twenty), analysed their slip vectors, and found the results in reasonable agreement with the globe's physiographic, seismic, and geological data. Le Pichon then extended his analysis back over the past 65 my, looking for patterns of continental drift and sea-floor spreading. He reported evidence for three major episodes of spreading – one in the Late Mesozoic, one in the Early Cenozoic, and one in the Late Cenozoic – with a major reorganization of the global pattern occurring at the beginning of each cycle. He correlated the slowing down of each cycle of spreading with the onset of mountain-building. Le Pichon's analysis of surface motions provided a coherent and intelligible history of interrelated plate motions over the entire globe.

Lamont's group of seismologists contributed 'Seismology and the New Global Tectonics', by Bryan Isacks, Jack Oliver, and Lynn Sykes. Sykes previously had shown that not only was seismicity limited to transform fault planes but that first motions on spreading ridges also confirmed Wilson's predictions for transform faults. Now the group reviewed the full range of seismological evidence that supports plate tectonics. The spreading oceanic ridge-rifts are loci of shallow-focus normal faulting. The East Pacific Rise, which has no rift and few earthquakes, appears to be spreading rapidly by ductile flow. Deep-focus earthquakes occur only on planes dipping beneath continents and into trenches. At first the group had been surprised to observe that these quakes generate high-frequency seismic waves which travel faster and more efficiently along the seismic zones than had been expected. This led to their realisation that the quakes are generated within slabs of rigid lithosphere as they descend into the yielding asthenosphere – the process called subduction (*see* **Tectonics:** Ocean Trenches). These seismologists concluded that the sinking of cold, heavy plates plays a dominant role in generating plate motion. Their task of analysing earthquakes globally was aided immeasurably by the establishment by the US Coast and Geodetic Survey of the World Wide Standardized Network (WWSSN) of 125 stations from which data was continually transmitted to a central location for digitising and archiving. By the mid-1960s, the data were available on microfilm. The original purpose had been to distinguish between earthquakes and underground nuclear explosions, but the Network was an invaluable source of high-quality data for seismologists. The mid-1960s also saw the development of high-speed computers and of programs, capable of reducing reams of data from continuous recordings of many kinds. These, along with other technical advances, played an essential role in developing the theory of plate tectonics.

Plate Tectonics Today

Plate tectonics rapidly became the dominant model of geoscience. It simplifies our understanding of the three principal igneous rock types that make up the Earth's crust: the basalts of the ocean floors are derived by partial melting of the peridotite mantle and erupted at the spreading ridges; the more siliceous andesites are derived by partial melting plus dehydration of descending slabs of lithosphere and are erupted over subduction zones, where they continuously contribute to island arcs and continental margins; the still more silica-rich continental rocks arose from the growth of andesitic landmasses and their subsequent reworking by metamorphism and metasomatism to form granitic plutons and rhyolitic volcanics, with no contribution from the mantle. Beginning with the first andesitic volcanoes, the buoyant continents have grown in area throughout geologic history while the oceans floors have lost area to subduction. By one estimate, the rapidly moving Pacific floor may disappear about 200 million years from now, when North America will collide with Asia.

Our understanding of the mechanism driving plate tectonics remains incomplete. Many scientists have abandoned the idea of rolling convection cells, which would not form a coherent system beneath the segmented oceanic ridges. However, we know that the asthenosphere is close to the melting point, so some scientists argue that global stresses rent the lithosphere into the pattern of mobile belts resembling the seam on a tennis ball. To them, the cracks came first and the basalts of the spreading ridges arose from decompressional melting of the mantle. They also see the gravitational pull of the cold, heavy slabs of sinking lithosphere as a prime factor in maintaining plate motion.

But the mantle is by no means a quiescent domain. It includes very deep-seated 'hot spots' from which vertical plumes arise and erupt as volcanoes. Each hot spot persists for millions of years. A dramatic example is the hot spot beneath the big island of Hawaii, which has remained fixed-in-place for 60 my while the Pacific plate has moved over it carrying a long line of islands and seamounts with ages steadily increasing toward the north-west. There is even a bend in the line showing that about 46 my ago the plate shifted its motion from northwest to almost due north. We now know of more than 100 mantle plumes under continental areas, oceanic ridges, and islands (**Figure 3**). Surely these long-lived sources of magma make a powerful contribution to the thermal instabilities involved in plate motions. Today, with or without our having to specify details of the causal mechanism, the theory of plate tectonics seems secure.

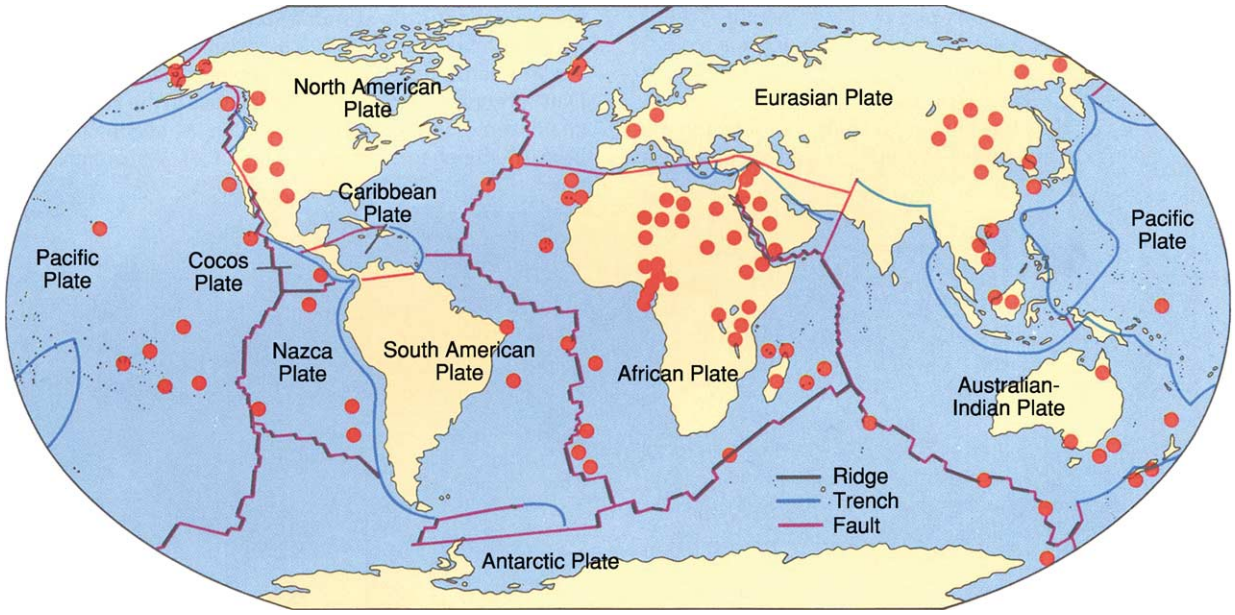


Figure 3 Sketch-map showing the Earth's seven major and three minor plates bounded by oceanic ridges, trenches, and faults, and 101 fixed hot-spots from which plumes of magma arise from deep within the mantle. Note that Africa has an unusually high number of hot-spots for a continental area. This, together with the fact that Africa is almost surrounded by spreading ridges, suggests that this continent has remained fixed-in-place for a very long time. (From Skinner BJ and Porter SC (1995) *The Blue Planet*; reproduced by permission of Skinner and John Wiley and Sons.)

Is Plate Tectonics Unique to the Earth?

Only two of the more than 100 planets that we have discovered outside our solar system have actually been seen as black spots passing in front of their suns. All the rest are inferred from irregularities in the orbits of their suns. Therefore, a search for plate tectonics outside the Earth must be limited to our four rocky neighbours in the inner Solar System – our Moon, Mercury, Venus, and Mars – and some of the moons of the outer planets. We have analysed samples collected on the Moon, and analysed those rare meteorites that have been projected to Earth by impacts on the Moon and Mars. In addition, we have landed analytical instruments on the surfaces of the Moon, Venus, and Mars. All our results show that the crusts of these bodies consist of basalts and related rocks that are akin to those of our ocean floors. As yet we have found no evidence of granitic continents.

Scientists have searched in vain on images of these bodies for patterns indicative of plate boundaries (*see* **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges): long ridges like spreading zones, chains of folded mountain ranges, or lines of volcanoes. Our Moon and Mercury are volcanically dead and pockmarked by impact craters. Venus, beneath its thick layer of clouds, appears to have been resurfaced with fresh basalts about 5 my ago, but it shows no linear pattern of features resembling ridges or subduction zones. Mars has the largest volcanoes in the Solar System, but they are scattered over a large area, and we see no evidence there of linear mountain ranges or trenches. Even less promising are the icy moons of the giant planets. We conclude that plate tectonics is unique to the Earth.

In this, we are fortunate, because plate tectonics is one of the essential factors that enable our planet to support complex forms of life. Volcanic eruptions continually renew Earth's waters and its atmosphere, including ozone, that shields us from lethal solar radiation, and the greenhouse gases that moderate the temperature and maintain it within a livable range for mammals. The creation of new islands and splitting and rearranging of continental areas generate a great variety of habitats which, in turn, minimizes the effectiveness of mass extinctions from whatever cause. Were plate motion to cease, mountain building (*see* **Tectonics:** Mountain Building and Orogeny) would cease, and uninterrupted erosion would begin to wear down the continents. Ultimately, the deposition of sediments in the oceans would presumably raise sea-level enough to flood much of the land surface. We literally owe our lives to plate tectonics.

See Also

Earth: Mantle; Crust. **Famous Geologists:** Du Toit; Wegener. **Gondwanaland and Gondwana**. **History of Geology From 1900 To 1962**. **Palaeomagnetism**. **Plate Tectonics**. **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges; Faults; Mid-Ocean Ridges; Mountain Building and Orogeny; Ocean Trenches; Rift Valleys.

Further Reading

Blackett PMS, Bullard E, and Runcorn SK (1965) A symposium on continental drift. *Philosophical Transactions of the Royal Society* London, Series A, 258, No. 1088.

Cox A (ed.) (1973) *Plate Tectonics and Geomagnetic Reversals*. San Francisco: WH Freeman and Company.

Marvin UB (1973) *Continental Drift: The Evolution of a Concept*. Washington, DC: Smithsonian Institution Press.

Menard WH (1986) *The Ocean of Truth: A Personal History of Global Tectonics*. Princeton: Princeton University Press.

Oliver JE (1966) *Shocks and Rocks: Seismology in the Plate Tectonics Revolution*. Washington, DC: American Geophysical Union.

Oreskes N and Le Grand HE (eds.) (2001) *Plate Tectonics: An Insider's History of the Modern Theory of the Earth*. Boulder: Westview Press.

Skinner BJ and Porter SC (1995) *The Blue Planet: An Introduction to Earth System Science*. New York: John Wiley and Sons, Inc.

Ward PD and Brownlee D (2000) *Rare Earth: Why Complex Life is Uncommon in the Universe*. New York: Copernicus, Springer-Verlag.

IGNEOUS PROCESSES

P D Asimow, California Institute of Technology, Pasadena, CA, USA

Introduction

All phenomena that involve rock in the liquid (molten) state are igneous processes. Molten or partially molten rock at depth in the earth is called magma; if erupted onto the surface at a volcano, it is then called lava (*see* **Lava**). Igneous processes involve making magma from previously solid source rocks, moving magma from one place to another, and freezing magma to produce igneous rocks. Although all of these steps may overlap, it is convenient to separate a discussion of igneous processes into three separate processes: melting processes, such as decompression melting, flux melting, conductive heating, and impact melting; magma transport processes, such as porous flow, dike injection, pluton emplacement, and volcanic eruption; and differentiation processes, such as fractional crystallization, magma mixing, and assimilation.

Igneous processes are distinguished from metamorphic processes, in which minerals react to form new rocks but melting does not occur, and from aqueous or hydrothermal processes, in which the composition of the fluid phase is dominated by water or other highly volatile components rather than by molten rock-forming minerals, although both metamorphism and hydrothermal activity are closely associated with igneous activity. Igneous processes are an essential part of the rock cycle; they lead to the formation of the wide diversity of observed igneous rocks and they constitute the fundamental process by which Earth and the planets differentiate into crust and mantle. The surface of the Earth reflects sources and processes at depth; by delivering material to the surface, igneous processes provide geologists and geochemists with essential information from which to constrain the compositional, mineralogical, and thermal character of Earth's interior. On Earth, igneous processes are closely related to plate tectonics, in that igneous activity is concentrated at divergent and convergent plate boundaries, although intraplate magmatism occurs as well. The types of igneous processes that occur in particular tectonic environments and the types of igneous rocks that result are generally characteristic of those environments. Therefore, distinctive associations of ancient igneous rocks, together with distinctive structural, metamorphic, and sedimentary features, can be used to identify ancient tectonic environments by comparison to modern examples.

Igneous processes are closely linked to the igneous rocks that are their final products. Basaltic magmatism (creating rocks such as basalt, gabbro, and komatiite) results from partial melting of planetary mantles of ultramafic composition; on Earth, this is seen at mid-ocean ridges, back-arc basins, large igneous provinces, continental rifts, intraplate hotspots, and subduction zones. Intermediate and silicic magmatism (creating rocks such as andesite, diorite, rhyolite, obsidian, and granite) is restricted to continental crust and is most common above subduction zones (*see* **Igneous Rocks:** Granite; Obsidian). Many economically significant ore deposits are associated with silicic magmatic activity; these are known as magmatic ores (*see* **Mineral Deposits and Their Genesis**).

Melting Processes

An important aspect of the modern understanding of igneous processes is that the interior of Earth is predominantly solid. This has not always been recognized; early geologists supposed that volcanic eruption of lava resulted from tapping of some permanently molten or glassy layer within Earth. Seismology, however, shows that Earth's crust and mantle transmit shear waves, which requires them to have rigidity. It is also known that the source regions of magma in the interior of Earth are too hot for glass to persist over geologic time. The shallow part of the mantle magma is less dense than solid rock and will rapidly separate and ascend rather than be stored for long times, whereas the pressure in the deeper part of the mantle is high enough to keep rocks in the solid state despite their high temperatures. Therefore, regions of volcanic activity are not simply areas where magma is tapped, but, first and foremost, are areas where magma is being created by melting of previously solid rocks.

In common experience, a solid material (e.g., an ice cube) is melted by first increasing its temperature to the melting point and then supplying the latent heat of fusion to transform the solid into a liquid. Intuitively, then, it might be imagined that melting of rocks is accomplished by increasing the temperature of the source rocks or by heat flow into the melting region. In this case, intuition is misleading; the dominant melting processes inside Earth are decompression melting and flux melting, both of which occur without increase in temperature or flow of heat, because the energy necessary to drive melting under the right circumstances is

already present in the source material. Furthermore, the phase relations of rocks are somewhat more complicated than can be understood by thinking of a single melting point, such as that of ice at ambient conditions. First, even for a pure substance such as ice, or for certain model mineral systems, the melting temperature is a function of pressure; for most substances, the melting temperature increases with pressure. Also, natural rocks are not pure substances but are mechanical mixtures of several minerals, each of which is a solid solution of many chemical components. This causes the melting temperature to widen into a range of partial melting temperatures, even at constant pressure. The point at which melting of a given rock begins with increasing temperature is called the solidus; the locus of such points at various pressures is the solidus curve. The point at which melting is complete and the last crystal disappears with increasing temperature is called the liquidus; the locus of such points at various pressures is the liquidus curve.

The solidus and liquidus curves for the most common composition in Earth's mantle are shown in Figure 1. These curves provide a context for much of the discussion of melting, because they show the pressure and temperature conditions at which mantle rocks partially melt, as well as the temperatures that actually occur in Earth's upper mantle in various tectonic regions. The gradient of temperature over the first several hundred kilometres in depth below a given place on Earth is divided into two intervals. The lithosphere generally encompasses the entire crust and that part of the upper mantle that is cold enough to act as a rigid material. The lithosphere can deform by faulting, but vertical motions are limited and hence heat transport through the lithosphere is dominated by conduction. Transportation of the geothermal heat flow by conduction causes the lithosphere to sustain a large thermal gradient; temperature increases with depth in the lithosphere by $500\,\mathrm{K\,GPa^{-1}}$ ($16^{\circ}\mathrm{C\,km^{-1}}$) or more. With increasing temperature, the plastic strength, or viscosity, of mantle rocks decreases to the point at which it becomes more efficient to transport geothermal heat flow by convection (in other words, by upwards vertical flow of hot material and downwards vertical flow of colder material at plate-tectonic rates of up to tens of centimetres per year). Heat transport by convection tends to drive the vertical temperature gradient towards the value that a parcel of material follows when its pressure changes without addition or removal of heat, the adiabatic gradient, which for mantle minerals is about $10\,\mathrm{K\,GPa^{-1}}$. The part of the mantle dominated by convective transport and displaying a nearly adiabatic temperature gradient is called the asthenosphere.

The temperature within the asthenosphere is characterized by a quantity called potential temperature, which is the temperature obtained by extrapolating the adiabatic gradient to the surface as if there were no lithosphere (and no melting). The potential temperature of the upper mantle sampled by mid-ocean ridges is thought to vary between about 1250° and 1400°C, whereas hotspots such as Hawaii may be underlain by hot plumes with potential temperature up to 1500°C. The temperature gradient and heat flow in the lithosphere are governed by the age and thickness of the lithosphere. For a normal potential temperature of 1350°C, Figure 1 shows a mid-ocean ridge thermal profile, wherein the zero-age lithosphere is only 10–30-km thick and the asthenosphere temperature profile extends nearly to the surface. Note that this temperature profile crosses the dry solidus of peridotite. Also plotted are an old oceanic thermal profile with a roughly 100 km-thick lithosphere and a stable continental thermal profile with a 150-km-thick lithosphere, both of which cross the water-saturated solidus of peridotite but stay well below the dry solidus.

Decompression Melting

Decompression melting drives basaltic volcanism at mid-ocean ridges, hotspots, back-arc basins, and may contribute to some extent in subduction-related island arcs and continental arcs as well. To understand decompression melting, it is necessary to consider the temperature structure of Earth, the phase relations of mantle rocks, and the existence of vertical flows in the asthenosphere that are fast enough to move material around without significant heat exchange with the surroundings.

The surface of Earth is cold, too cold for molten rock to be stable. At depths of more than a few hundred kilometres, the interior of Earth is everywhere at high enough temperature that, at ambient pressure, mantle rocks would be at least partially molten. However, at these depths, the pressure is high enough to raise the solidus temperature well above modern asthenosphere temperatures. In other words, although the temperature increases with increasing pressure in the asthenosphere by about $10\,\mathrm{K\,GPa^{-1}}$, the dry solidus temperature of mantle rocks increases by about $130\,\mathrm{K\,GPa^{-1}}$. Conversely, then, if solid material in the asthenosphere can flow upwards more quickly than heat can be conducted out of it, it will cool by only $10\,\mathrm{K\,GPa^{-1}}$ while the solidus temperature decreases by $130\,\mathrm{K\,GPa^{-1}}$ and the upwelling material will intersect the solidus and begin melting. Therefore, the expectation is to find basaltic magma anywhere that horizontal extension of the lithosphere allows the asthenosphere to

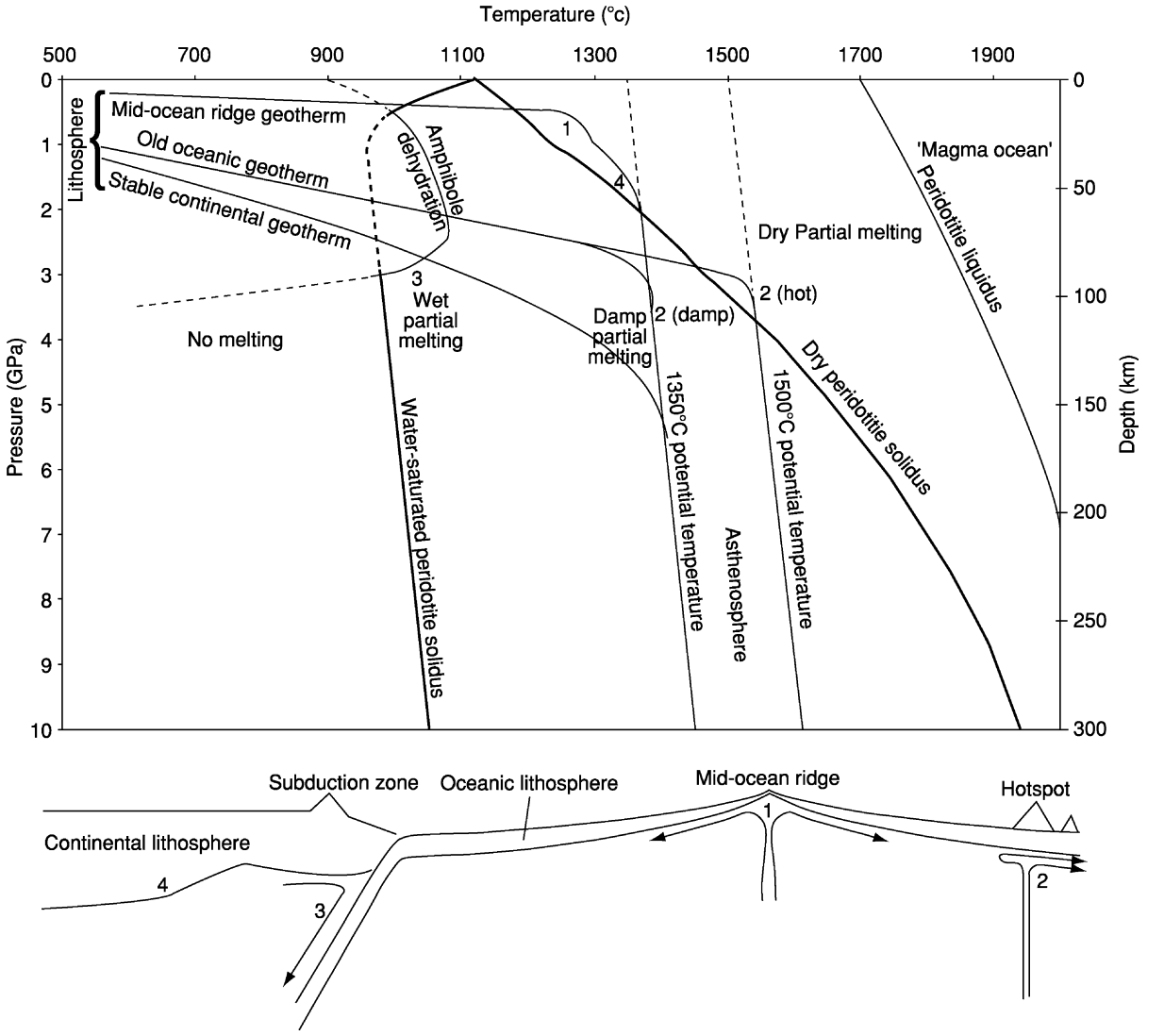


Figure 1 The pressure–temperature structure and physical locations of major igneous processes on Earth. Curves shown in pressure–temperature space include the liquidus of peridotite, the major rock composing the Earth's upper mantle; the solidus of peridotite under dry conditions; the solidus of peridotite in the presence of excess water; and the breakdown of the hydrous mineral amphibole, which can store water and prevent wet melting as long as it is stable. Compared to these underlying phase relations, the temperature structures of lithosphere and asthenosphere underlying mid-ocean ridges, hotspots, and continents are shown. This shows that melting of peridotite is expected at low pressure and dry conditions under mid-ocean ridges (1), at high pressure and either hot or damp conditions at intraplate hotspots (2), and at wet and cold conditions above the water-saturated solidus and amphibole stability limit within subduction zones (3).

flow upwards with minimal cooling and fill the resulting gap. This is the case at mid-ocean ridges (region 1 in Figure 1) and in well-developed continental rifts (region 4 in Figure 1). At higher potential temperatures or higher mantle water contents, the asthenospheric temperature profile will cross the peridotite solidus at higher pressure, so that upwelling flow can drive melting even beneath a thick lithosphere. This is the case at hotspot localities such as Hawaii (region 2 in Figure 1) and in large-scale igneous provinces such as the Ontong-Java plateau or the Columbia River flood basalt (*see* **Large Igneous Provinces**). Whether decompression melting occurs in any particular environment depends on the potential temperature, the water content of the source rocks (see later), and the extent to which the lithosphere is thinned or removed to allow the asthenospheric temperature gradient to extend close to the surface.

It is important to realize that decompression melting is always a partial melting process; there is always a solid residue left behind in the mantle. Although a completely liquid magma can separate from the residue, this is the result of magma transport

(see later), not of complete melting. With continued decompression above the point of intersection with the solidus, partial melting continues and reaches higher extents of melting for longer melting columns. Thus, the volume or flux of magma generated depends on the same variables that determine whether melting occurs in the first place (lithosphere thickness, potential temperature, and water content) as well on as the flow rate of mantle source rocks through the melting region.

Flux Melting

At subduction zones, a cold slab of old oceanic lithosphere flows down into the mantle and induces downwards corner flow of the mantle wedge above the slab and below the thick lithosphere of the overlying plate. The shape of the dry solidus of peridotite and the energetics of decompression melting show that melting is to be expected under thin lithosphere, where the mantle is hot, and where the flow direction is upwards. Subduction zones presumably show none of these characteristics, yet most of Earth's most recognizable volcanoes and nearly all of the hazardous ones form above subduction zones. Clearly, a great deal of melting occurs in such settings. This primarily results from the introduction of water into the mantle wedge via subduction of oceanic sediments and hydrothermally altered oceanic crust. At high pressures, when water can readily dissolve in magma, it acts as a flux; Figure 1 shows that the water-saturated solidus of peridotite is several hundred degrees below the dry solidus. At intermediate water contents, melting begins on a damp solidus curve between the two limits shown. The process of melting by addition of water to material below its dry solidus, but at or above the wet solidus, is called flux melting, which is the second most efficient source of magmatism on Earth.

Subduction zone magmatism is complex, but generally, there is consensus on the essential elements. The subducting slab remains relatively cold, and only the sedimentary component is thought to melt directly, creating a mobile liquid that ascends into the shallow part of the overlying mantle wedge and refreezes, modifying the composition of the wedge. At somewhat greater depth, the basaltic component of the slab undergoes a series of dehydration reactions that create a water-rich fluid. The stability limits of hydrous minerals in peridotite are different from those in basalt, so that as this fluid ascends out of the slab and into the immediately overlying cold part of the mantle wedge, it may also freeze in place and generate an enriched, hydrated mantle source. However, with further downflow along the slab, or if the fluid can migrate far enough into the hot interior of the mantle wedge, this material crosses the hydrous solidus (region 3 in Figure 1) and partially melts to create primary arc basalt.

Heating by conduction At locations remote from plate boundaries and hotspots (for example, in continental interiors), volcanism and magmatic intrusion do occur, requiring a mechanism other than decompression or flux melting. These locations include large rhyolitic caldera-forming systems such as Long Valley, in California. Their source materials are embedded in the lithosphere (which is both too cold and too rigid for decompression melting to be effective) and, on a stable long-term geotherm, sit below even the water-saturated solidus curve. Direct heating by conduction is therefore the most likely mechanism for bringing these sources into their melting range. Why should conduction deliver more heat in one place than in another? The answer generally goes back to decompression melting in the underlying mantle. Geochemical and geophysical evidence typically shows that, although the principal source of intraplate volcanism may be within the crust, there is often a mantle component. Ponding of basaltic magma and subsequent underplating of basaltic rocks at the base of the crust are the most efficient ways to focus a large heat flow into a specific region of the crust. Because basalts crystallize at temperatures above 1000°C and the rocks of the continental crust can begin melting (in the presence of water) near 700°C, it is clear that crustal melting is a likely consequence of the arrival of a large mass of basalt at the base of the crust. If the basalt actually intrudes the crust and assimilates or mixes with the resulting melts of surrounding crustal rocks, the process may be described as a differentiation process (see later) affecting the basalt, as well as a melting process affecting the wallrock.

Magma Transport

The transport of magma from regions of melting to regions of emplacement or eruption is a fundamental aspect of igneous phenomena; indeed, the outpouring of lava or explosive eruption of volcanic ash is the most obvious and hazardous manifestation of igneous activity. Most of the melting processes that occur inside Earth are partial melting processes, producing a mixture of liquid and residual minerals. Somehow the liquid component of this mixture is physically separated from the residue and is transported, generally to shallower depths. Thus, for example, a mid-ocean ridge basalt created by 10% melting of its mantle source may erupt as a 100% liquid, because the liquid and residue have been separated by melt

migration. Indeed, in the absence of separation between melts and residues, igneous processes would be unable to drive differentiation of Earth into mantle and crust, and would be unable to produce many of the eruptive phenomena associated with volcanism.

There are several mechanisms of magma transport to discuss (Figure 2), but all of them are driven essentially by the same force: gravity acting on the buoyancy of melts or melt–gas mixtures relative to residual minerals or wallrocks. That is, when a rock partially melts, at least at the pressures at which most melting takes place inside Earth today, the liquid is less dense than its surroundings are. Gravity will therefore cause the melt to rise, if a pathway is available that allows it to do so. Such a pathway can be established by intergranular porous flow, if the liquid

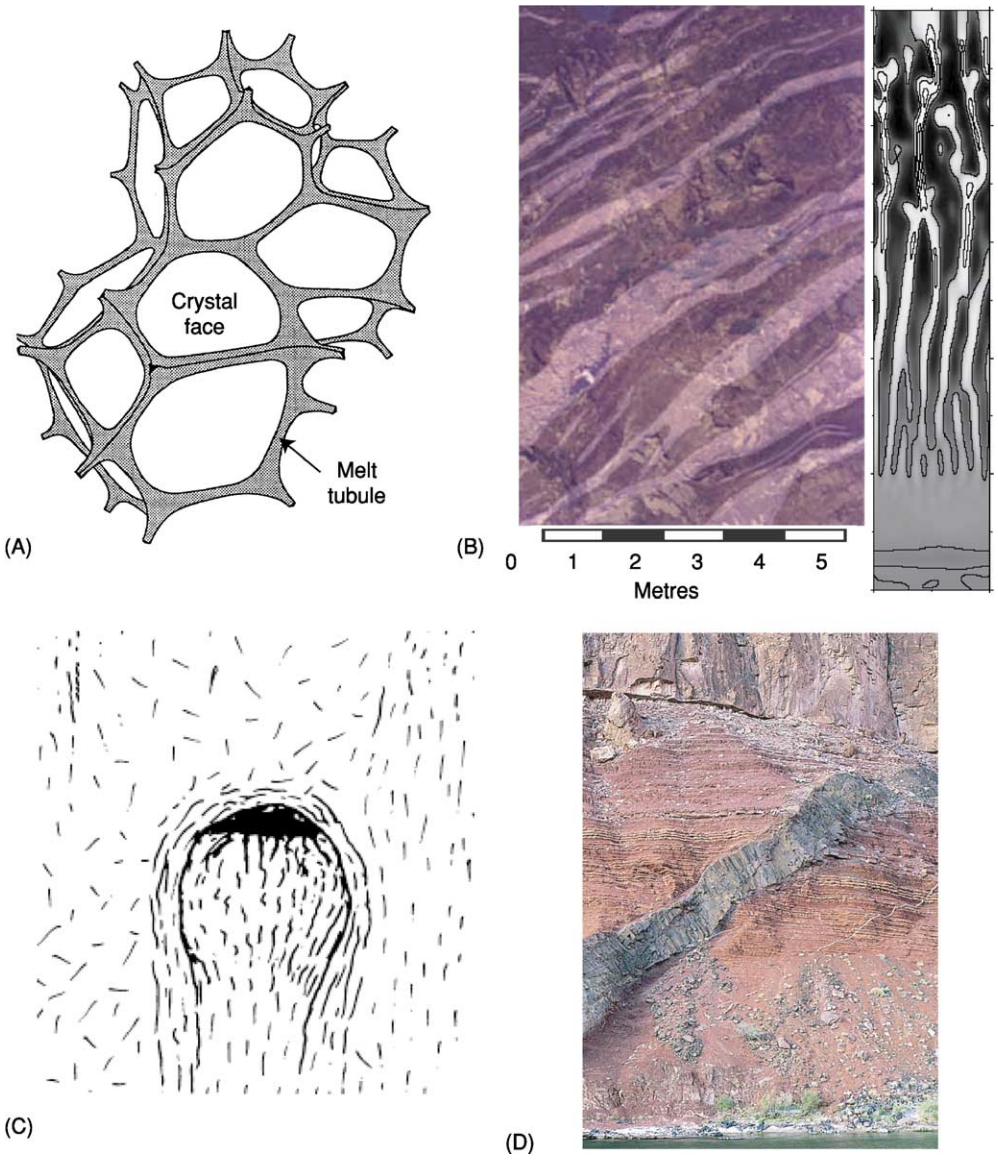


Figure 2 Illustrations of magma transport mechanisms. (A) During partial melting of mantle rock, the melt phase organizes into an interconnected network of tubules along the intersections where three or four solid crystals of the residual rock meet. This allows the melt to move by porous flow. (B) Bands of dunite (light colour, olivine only) and harzburgite (dark colour, olivine plus orthopyroxene) in a section of former oceanic upper mantle exposed in the Sultanate of Oman; the adjacent diagram is a numerical model of the development of such structures by porous flow, in which melt is able to develop preferred flow pathways by dissolving orthopyroxene. (C) Sketch of a buoyant diapir of partially molten rock rising through denser asthenosphere containing less melt. (D) A dyke of basalt intruded into sedimentary rocks in the Grand Canyon, Arizona, USA. (C) Reprinted with permission from Basaltic Volcanism Study Group (1981) *Basaltic Volcanism on the Terrestrial Planets*, p. 499. (B) Reprinted with permission from Braun MG and Kelemen PB (1992) Dunite distribution in the Oman ophiolite: Implications for melt flux through porous dunite conduits, with permission from American Geophysical Union. (D) Reprinted with permission from Kelemen PB, Hirth G, Shimizu N, Spiegelman M, and Dick HJB (1997) *Philosophical Transactions of the Royal Society of London, Series A* 355, p. 296.

forms an interconnected network along the grain boundaries of the partially molten rock, or by opening a crack or dyke in the overlying rock to create a conduit for melt flow. Also, the partially molten assemblage is less dense than is the unmelted rock of the same bulk composition, and therefore melts can rise by diapirism, or upward flow of the entire partially molten region, so long as the surroundings are weak enough to allow such flow. Finally, as melts containing dissolved volatiles reach low pressures in the crust, the dissolved gas components form bubbles. This leads to a runaway drop in density, increase in buoyancy, faster uprise, and continued bubble growth; the end result may be fragmentation of the magma as bubbles begin to touch each other, and the explosive eruption of tiny shards of volcanic glass called ash.

Porous Flow

The initial stage of the melt separation process is presumably porous flow. Experiments on the texture of partially molten mantle-like rocks shows that, at mantle temperatures and small degrees of partial melting, the melt phase organizes into a network of tubules along the boundaries of crystals that make up the solid residue (Figure 2A). The melt thereby establishes an interconnected network through which it can migrate relative to the solid. Such porous flow is driven by pressure gradients, principally due to gravity but also due to shearing forces as the host rock deforms, and resisted by the viscosity of the magma and the permeability of the host rock. Although the exact relationship is unknown, permeability is an increasing function of the fraction of melt present, such that more melt can flow through a region where more melt is present. Porous flow is a rather slow melt migration process and is thought to allow continuous chemical equilibration between melts and solids, under mantle conditions. Such chemical equilibration can, in some cases, lead to extra melting and an increase in local melt fraction. Because this increases the permeability, increased flow can lead to yet more extra melting and an instability can develop whereby a porous flow system evolves into a set of high-porosity, high-flux conduits embedded in a low-porosity, low-flux matrix. This process is thought to be important in the rapid extraction of basaltic magma from the mantle, in determining the chemical characteristics of mid-ocean ridge basalt, and in explaining the rock types and distributions seen in outcrops of former oceanic mantle rocks (Figure 2B).

Diapirism

In addition to relative flow between melt and solids in a partial melt, the bulk partially molten assemblage can migrate relative to surrounding regions (Figure 2C). When a mass of bulk partial melt flows upwards because of its thermal or melt-induced buoyancy, this is called a diapir. The significance of diapirism in many magmatic settings is unclear. There may be a component of active, buoyancy-driven flow beneath mid-ocean ridges, but, in general, this is not necessary to explain the existence of ridges or the eruption of magmas on-axis. There may be buoyant upwellings of hydrated and/or partially molten material from near the slab in subduction zones, and this may help to explain the large degrees of melting and large volumes of melt emplaced in such settings, but, again, this is controversial. The most likely setting in which diapirism is essential to explaining observed field relations is in the emplacement of granitic plutons within continental crust.

Dyke Injection

In the lithosphere, porous flow becomes an inefficient means of moving melts. The low temperatures and conductive heat flow imply that melts migrating slowly through the lithosphere will begin to freeze. This reverses the permeability feedback that occurs in the asthenosphere, and chokes off melt conduits where melt is able to react with wallrocks. Also, the plastic strength of minerals increases with decreasing temperature to the point at which the host minerals are unable to compact or expand to accommodate changes in melt fraction. On the other hand, this allows differential stresses to accumulate rather than relax. As pressure decreases, the differential stress needed to cause brittle failure and crack propagation decreases as well. Hence, at shallow depths, melt migration becomes dominated by the formation of cracks and the flow of melt through the resulting conduits. This process is called dyke injection, and the tabular bodies of igneous rock that end up frozen in such crack-related conduits are called dykes (Figure 2D). By its nature, dyke injection is an episodic process; it requires stress to build to the point of failure and it requires a large pool of melt to flow suddenly into the dyke and maintain stress at the crack tip. On the other hand, porous flow is a continuous process. The transition from porous flow to dyke injection, which is associated in some way with the transition from asthenosphere to lithosphere, is therefore a likely location for melt to pond and accumulate in some temporary storage reservoir. This is one mechanism for developing a magma chamber, the primary site where differentiation takes place (see later).

Eruption

When the products of igneous activity within Earth, in the form of lava or ash, flow or explode onto the surface, the site of eruption is called a volcano.

This is a type of igneous process, clearly involving melt transport. (*see* **Pyroclastics, Volcanoes**).

Differentiation Processes

Much of the diversity of igneous rock compositions found on Earth results from the evolution of magmas once they have migrated away from the melting region of their source and have arrived at a shallow enough depth to begin crystallizing. The sets of processes that act on a primary magma after separation from its source are collectively called differentiation, and they include partial crystallization, assimilation of foreign material, and mixing of various magmas.

Just as the melt phase composition is distinct from the residue composition during melting, allowing separation of chemical components by partial melting, so the crystals that form during differentiation differ in composition from the liquid, and so by mass balance cause the liquid to evolve in composition as it cools and crystallizes. During crystallization, the liquidus is the temperature at which the first crystals appear and the solidus is the temperature at which the last bit of liquid freezes.

Fractional Crystallization

The dominant process relating a suite of igneous rocks within a given locality is often fractional

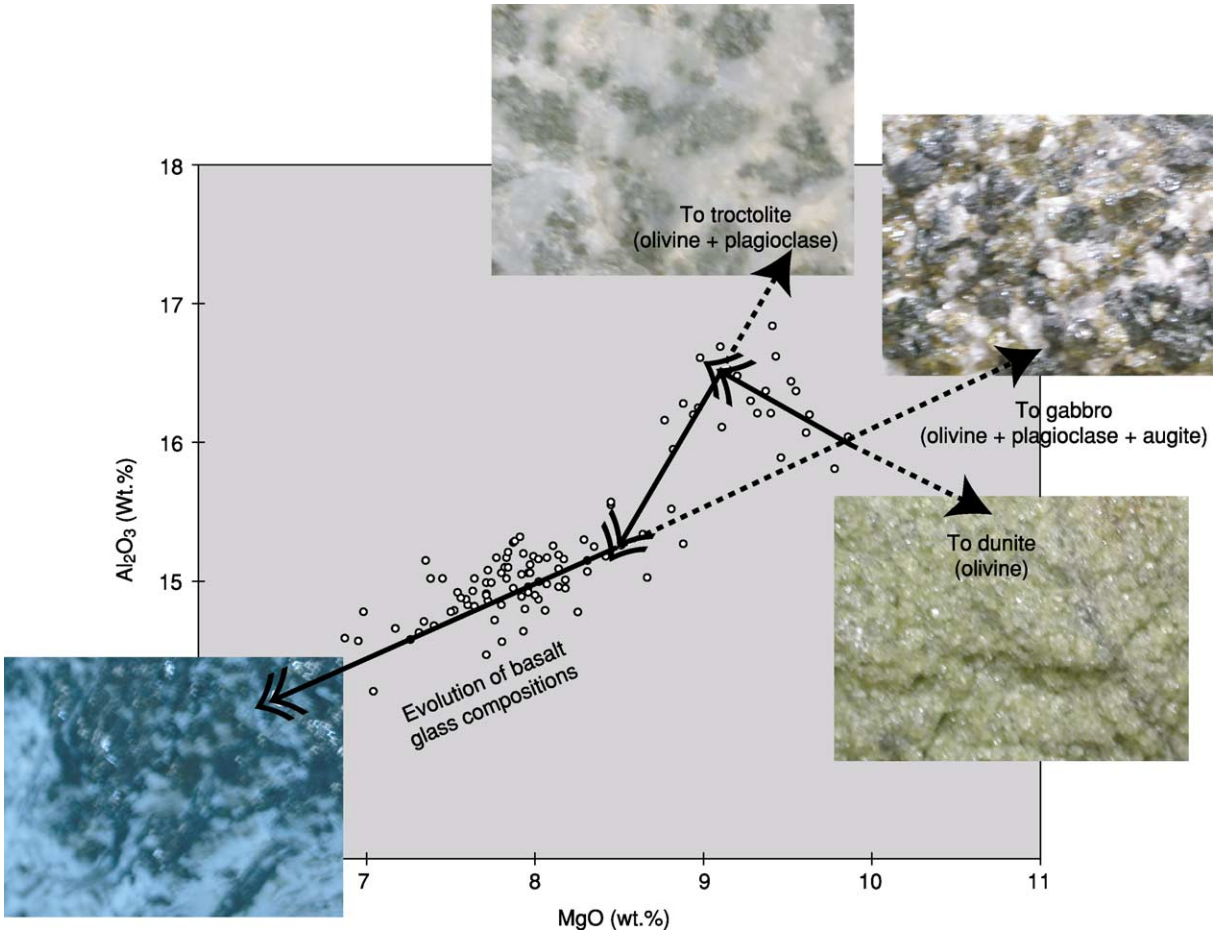


Figure 3 Illustration of complementary formation of plutonic and volcanic rocks by fractional crystallization. The graph shows MgO and Al_2O_3 contents measured in a suite of samples of volcanic glass recovered from a small area of the Mid-Atlantic Ridge. The solid double-headed arrows show the evolution of magma composition that results from the withdrawal of particular combinations of minerals, and the dashed arrows point towards the rocks that form by accumulation of the fractionated minerals. The first mineral to crystallize from the parental liquid is olivine (pale green). Olivine is high in MgO and low in Al_2O_3, and its removal causes the liquid to evolve towards lower MgO and higher Al_2O_3. A rock consisting entirely of olivine is called dunite. The second mineral to join olivine in the fractionating assemblage is plagioclase (white), which is high in Al_2O_3; removal of olivine and plagioclase causes the rock to decrease in MgO and Al_2O_3. A rock consisting of olivine and plagioclase is called a troctolite. Finally, a composition of clinopyroxene, called augite (black), joins the fractionating assemblage, driving the liquid along a shallower trend of decreasing MgO and Al_2O_3. A rock containing olivine, plagioclase, and augite is a gabbro. Eruptions of liquid occur at different stages in this process, whereupon rapid cooling by contact with seawater forms basaltic glass (shiny black) that can be collected and measured.

crystallization. This is the process whereby, on cooling, crystals grow from a liquid and become chemically and physically separated from the evolving liquid. Accumulating the resulting crystals results in plutonic igneous rocks called cumulates, such as the gabbro of the lower oceanic crust or the layered basic intrusions. Sampling the remaining liquid by eruption at various stages of fractionation results in a series or suite of liquids with compositions that can be related by subtraction of the crystal phases that are being fractionated. An example is shown in Figure 3; the compositions of a suite of basalt glasses dredged from a small region of the Mid-Atlantic Ridge show evidence of removal first of olivine, followed by olivine and plagioclase, and finally by olivine, plagioclase, and augite. The lower crust and upper mantle beneath mid-ocean ridges contain the rocks that result from removal of these combinations of minerals: dunite, troctolite, and gabbro, respectively. Thus, the fractional crystallization process in this case explains the mineralogy both of plutonic rocks formed from the fractionated crystals and of erupted volcanic rocks formed from the complementary evolving liquids. From the same primary liquid, a diversity of evolved rocks can result, whether by continued evolution along a particular liquid line of descent or along different liquid lines of descent determined by the conditions of fractionation. The important variables determining which fractionation trend a given liquid follows include the pressure, water activity, and oxidation state in the magma chamber.

Assimilation and Magma Mixing

Some suites of igneous rocks display compositional relationships, textures, or other field relations showing that they did not evolve by simple removal of fractionating crystals from a single primary liquid. Many other processes can affect the evolution of igneous rocks in the lithosphere. These include assimilation (the melting and dissolution of surrounding rocks with lower melting points) and magma mixing (the combination of various batches of primary liquid that may have been created by melting of different sources) to different extents and/or at different conditions. Generally, magmas do not arrive in the crust with very much superheat; i.e., they are on their liquidus and will begin crystallizing if energy is taken out of the magma. On the other hand, assimilation of solid wallrocks requires transfer of energy from the magma to the assimilant in order to melt it. Hence, assimilation and fractional crystallization are generally coupled, occurring together and with the balance between mass of wallrock melted and mass of magma crystallized determined by the energy balance of the overall system. Assimilation can readily change the chemical, isotopic, and volatile composition of an evolving magma, because the wallrocks encountered are likely to be much older, more chemically more evolved, and more hydrothermally altered than the primary intrusion is. Evidence that processes such as assimilation and magma mixing have occurred may be drawn from chemical analysis of igneous rocks, but may also be seen directly through the occurrence of incompletely melted enclaves or stoped blocks within a magmatic body.

Glossary

adiabatic gradient The pressure–temperature path followed by rocks moving upwards or downwards without exchanging heat with their environment; given in K GPa^{-1}.

assimilation The melting and dissolution into a magma of surrounding rocks, encountered during melt transport or differentiation.

asthenosphere That part of Earth's mantle where temperatures are high enough and the rocks weak enough to allow heat transport by convection.

diapir A buoyant upwelling, especially a discrete mass or episodic upwelling, rather than a continuous flow.

dyke A (near-vertical) crack filled with crystallized magma; a former conduit for the transport of magma through the lithosphere.

fractional crystallization The formation of crystals from a magmatic liquid under conditions in which the crystals are physically or chemically isolated from further interaction with the remaining liquid.

igneous rock A rock formed by cooling from the molten state.

lava Molten or partially molten rock as it erupts or flows on the surface of a planetary body.

liquidus, liquidus curve The maximum temperature (as a function of pressure) at which a rock of a particular composition will be completely molten; also the temperature at which a magma or lava of a given composition will begin crystallizing.

lithosphere The outer shell of Earth, where heat transport is dominated by conduction.

magma Molten or partially molten rock in the interior of a planetary body.

porous flow The motion of a fluid driven by a pressure gradient through interstitial space in a solid, such as grain boundaries or pore space.

potential temperature The temperature that a parcel of asthenosphere would reach if expanded to 1 atmosphere pressure without melting or exchanging heat with its environment; given in °C.

solidus, solidus curve The minimum temperature (as a function of pressure) at which a rock of a

particular composition will begin to melt; also the temperature at which a magma or lava of given composition will be completely crystallized.

See Also

Earth: Mantle. **Igneous Rocks:** Granite; Komatiite. **Large Igneous Provinces**. **Mantle Plumes and Hot Spots**. **Mining Geology:** Magmatic Ores. **Pyroclastics**. **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges; Mid-Ocean Ridges. **Volcanoes**.

Further Reading

Asimow PD, Hirschmann MM, and Stolper EM (1997) An analysis of variations in isentropic melt productivity. *Philosophical Transactions of the Royal Society of London, Series A* 355: 255–281.

Basaltic Volcanism Study Project (1981) *Basaltic Volcanism on the Terrestrial Planets*. New York: Pergamon.

DePaolo DJ (1981) Trace-element and isotopic effects of combined wallrock assimilation and fractional crystallization. *Earth and Planetary Science Letters* 53(2): 189–202.

Grove TL, Kinzler RJ, and Bryan WB (1992) Fractionation of mid-ocean ridge basalt (MORB). In: Morgan JP, Blackman DK, and Sinton JM (eds.) *Mantle Flow and Melt Generation at Mid-Ocean Ridges*, pp. 281–310. American Geophysical Monograph 71. Washington DC: American Geophysical Union.

Hirschmann MM (2000) Mantle solidus: experimental constraints and the effects of peridotite composition. *Geochemistry Geophysics Geosystems* 1: 70.

Kelemen PB, Hirth G, Shimizu N, Spiegelman M, and Dick HJB (1997) A review of melt migration processes in the adiabatically upwelling mantle beneath oceanic spreading ridges. *Philosophical Transactions of the Royal Society of London, Series A* 355: 283–318.

Langmuir CH, Klein EM, and Plank T (1992) Petrological systematics of mid-ocean ridge basalts: constraints on melt generation beneath ocean ridges. In: Morgan JP, Blackman DK, and Sinton JM (eds.) *Mantle Flow and Melt Generation at Mid-Ocean Ridges*, pp. 183–280. American Geophysical Monograph 71. Washington DC: American Geophysical Union.

Sigurdsson H, Houghton B, McNutt SR, Rymer H, and Stix J (eds.) (2000) *Encyclopedia of Volcanoes*. San Diego, CA: Academic Press.

Wickham SM (1987) The segregation and emplacement of granitic magmas. *Journal of the Geological Society* 144: 281–297.

Yoder HS (1979) The evolution of the igneous rocks: fiftieth anniversary perspectives. Princeton, NJ: Princeton University Press.

IGNEOUS ROCKS

Contents

Carbonatites

K Bell, Carleton University, Ottawa, ON, Canada

Introduction

Few other igneous rocks are as intriguing and as fascinating as carbonatites. Made up of at least 50% carbonate minerals, carbonatites have distinctive trace element chemistries, many contain unusual, accessory minerals and some are even associated with economic mineral deposits. Because carbonated melts can be produced by very low degrees of partial melting (<0.01%) of a volatile-rich source, they may be sensitive indicators of thermal instabilities in the mantle (*see* **Earth:** Mantle) and may mark asthenospheric upwellings, deep mantle plumes, and crustal delamination.

The main physical property that separates carbonatitic melts from most others is their viscosity. Because these melts are ionic liquids, there is little or no polymerization and hence these magmas have very low viscosities. Carbonatitic melts can therefore sample

large mantle volumes, and infiltration experiments demonstrate that they can percolate rapidly in polycrystalline olivine by chemical exchange between melt and matrix. Estimates of ascent rates based on fluid-flow calculations suggest that carbonatitic melts can migrate to the surface at speeds of 20 to 65 ms^{-1}.

Once considered rare, carbonatites are now fairly commonplace, with more than 500 occurrences reported to date. Although carbonatites are found on all continents, including Antarctica, almost half are found in Africa where they are intimately associated with the East African Rift Valley system. Brazil, Canada, and north-western Russia are other regions where carbonatites are abundant. So far only two oceanic island occurrences have been found: the Cape Verde and the Canary Islands. About 50% of carbonatites are associated with extensional environments, such as rift valley systems, while others are associated with major faults and large-scale domal swells. Carbonatites are rarely, if ever, associated with subduction-related environments. In constructing any model for the origin of carbonatites, the restriction of most carbonatites to continental areas suggest that thickened lithosphere plays an important role in the production of CO_2-rich melts. Figure 1 shows the general distribution of carbonatites on a worldwide basis. Also shown for reference, are areas underlain by Archaean cratons.

Even though carbonatites are volumetrically insignificant, they provide insights into terrestrial differentiation unrivalled by few other igneous rocks. First of all, they provide information about the geochemistry of the Earth's mantle, secondly they can be used to monitor the evolution of the mantle during the last 3 Ga of Earth history, and lastly they tell us something about the migration of low viscosity melts at both mantle and crustal levels.

Field Relationships

Just like any other igneous rock, carbonatites can take the form of plutons, dykes, sills (although somewhat rare), cone sheets, lava flows, and pyroclastic deposits. Table 1 gives localities for some of these

Table 1 Forms that carbonatites can take

Form	*Example*
Lava flows	Oldoinyo Lengai, Tanzania
Lava flows	Fort Portal, Uganda
Lava flows	Kontozero, Russia
Tuffs	Kerimasi, Tanzania
Tuffs	Kaiserstuhl, Germany
Tuffs	Cape Verde Islands
Linear dykes	Gross Brukkaros, Namibia
Linear dykes	Alnö, Sweden
Linear dykes	Turiy, Russia
Sills	Kaluwe, Zambia
Cone sheets and ring dykes	Oka, Canada
Cone sheets and ring dykes	Dicker Willem, Namibia
Plutons	Jacupiranga, Brazil
Plutons	Sokli, Finland
Plutons	Firesand, Canada

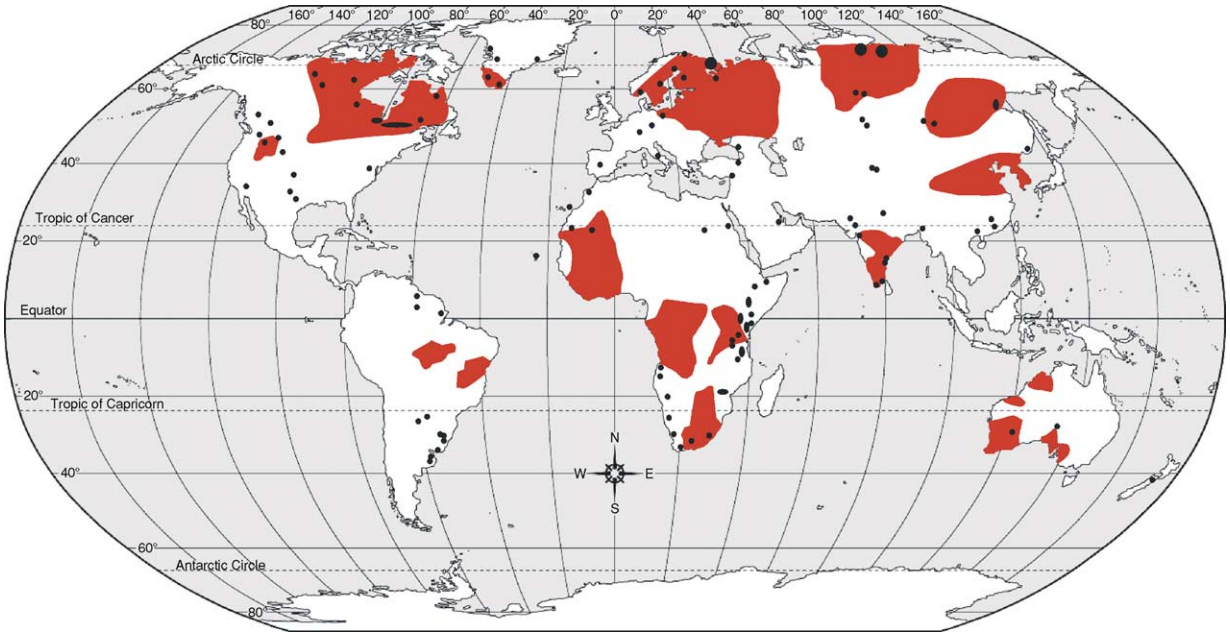


Figure 1 Distribution of carbonatites and presently-exposed Archaean cratons. Note the widespread distribution of carbonatites, and that only two carbonatite occurrences are found in oceanic islands. © Hougthon Mifflin Company.

different forms. Although it is difficult to make generalizations about carbonatitic complexes, they usually consist of annular or cylindrical bodies, many less than 5 km across, of silicate rocks and/or carbonatite material cut by younger carbonatite. Invariably, the carbonatite is much younger than the associated silicate rocks, and commonly much less in volume. Many carbonatites occur in parallel dyke swarms, and many are associated with small diatreme breccias and sub-volcanic pipes and plugs. Figure 2A and B show idealised sections through a typical carbonatite complex.

When first proposed, the magmatic origin for carbonatites was vigorously challenged, but the evidence is now so overwhelming that even the most critical of sceptics accept carbonatites as true igneous

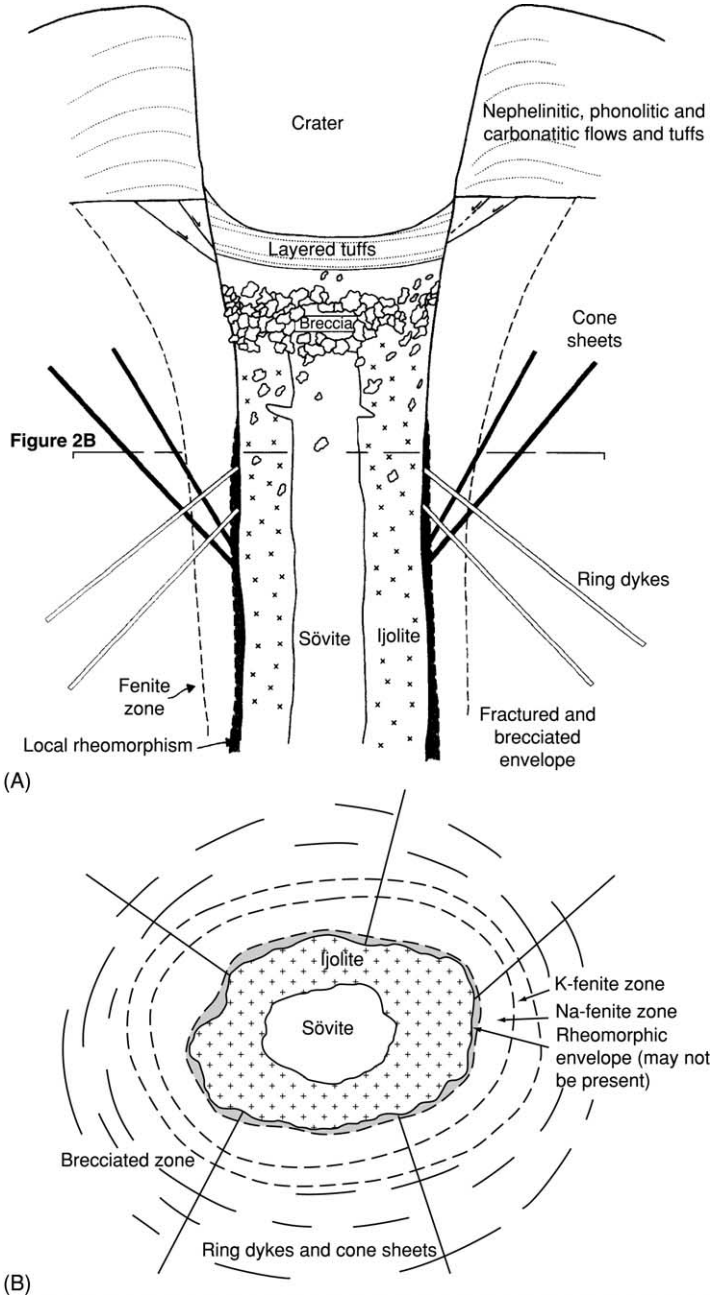


Figure 2 Idealized sections through a carbonatite complex. The early silicate rocks, such as ijolite and urtite, are cut by carbonatite. Sövite is a carbonatite rock type. Ferite is a contact metasomatic rock type. Dykes, cone sheets, and ring dykes, as well as veins, are among the last events to occur. In many complexes, the youngest events are represented by veins of dolomite carbonatites, ferrocarbonatites, and REE-rich carbonatites. Adapted from Sage and Watkinson (1991).

Table 2 Rock associations

	Age	*Rock type*
Carbonatite-melilitolite		
Oka complex, Quebec, Canada	120 Ma	Melilitolite (*okaite*), pyroxenite, ijolite, urtite, alnoite
Turiy complex, Russia	380 Ma	Melilitolite, melteigite, ijolite, urtite, phoscorite
Carbonatite-phonolite-nephelinite		
Oldoinyo Lengai	Active	Ijolite, nephelinite, phonolite
Cape Verdes	<12 Ma	Alkali basalt, basanite, nephelinite, phonolite, tephrite, syenite
Napak, Uganda	<40 Ma	Ijolite, nephelinite, phonolite
Carbonatite-pyroxenite		
Vuoriyavi, Russia	380 Ma	Pyroxenite, ijolite, melteigite, syenite
Phalaborwa, South Africa	ca 2050 Ma	Pyroxenite, syenite
Carbonatite-syenite		
Salitre, Brazil	80 Ma	Syenite, pyroxenite
Khibina, Russia	380 Ma	Syenite, ijolite, jacupirangite, foyaite, trachyte
Carbonatite-lamprophyre		
Kandalaksha dykes, Russia	380 Ma	Ultramafic lamprophyres
Gardar, Greenland	ca 1200 Ma	Ultramafic lamprophyres
Carbonatite-kamafugite		
Western Uganda	Holocene	Kalsilitite, melilitite, leucitite
Italy	Holocene	Melilitite

rocks. Evidence includes the presence of chilled margins, Peleé's tears, vesicles, lava flows, dykes, veins, and flow textures. Although most carbonatites occur as plutonic rocks, there are well over forty localities with extrusive carbonatites, at nine of which there are lavas. It should also be noted that carbonatites can take the form of hydrothermal and replacement bodies.

Carbonatites rarely occur on their own. Most are spatially associated with a variety of silicate rocks of similar age. Of the classic carbonatite complexes scattered throughout East Africa, most are associated with nephelinites and phonolites, or their plutonic equivalents. Silicate rocks, when present, form a much greater volume than the carbonatites and are usually characterized by great structural and textural complexity. The close proximity between silicate rocks and carbonatites within the same complex means that the origin of one cannot be divorced from the origin of the other. Table 2 groups possible carbonatite-silicate rock associations.

Ages

Carbonatitic magmatism extends back to at least 3 Ga, although carbonatites of Archaean age are comparatively rare. The oldest known carbonatite to date occurs in Greenland (Tupertalik, 3.0 Ga). Other Archaean carbonatites include Silliinjarvi in Finland, and Dolodau and Lac Shortt in Canada. The youngest carbonatite is represented by the lava flows from Oldoinyo Lengai in Tanzania, which is the only known active carbonatite volcano on Earth strategically situated in the floor of the East African Rift Valley system. Although carbonatitic magmatism on a worldwide scale has been continuous throughout geological time, distinct groupings occur at about 2.8, 2.1, 1.8, 1.0, .60, and .35 Ga, ages that roughly correspond to periods of major orogenic activity. As a generalization, however, most carbonatites occur in extensional, rather than compressional, environments.

In terms of age, carbonatites appear to become more abundant towards more recent times. This might simply reflect the difficulties encountered in preserving carbonatites or alternatively it might simply reflect the much higher geothermal gradient that was in existence during the Archaean.

Mineralogy

Calcite and dolomite are the major rock-forming minerals in carbonatites. Other carbonates include the Fe-rich varieties and some of the much rarer REE (rare-earth element) carbonates. Table 3 lists some of the minerals commonly found in carbonatites. The division of most carbonatites into three groups is based on whether calcite, dolomite, or Fe-rich carbonates (*see* **Minerals:** Carbonates) are the predominant minerals. The plutonic variety of calcite carbonatite is sometimes called sövite and when medium to fine grained is referred to as an alvikite. Both of these names have been carried over from the very early days of carbonatite research. One unusual carbonatite, found only at Oldoinyo Lengai volcano in Tanzania, consists of sodium, potassium,

Table 3 Some minerals found in carbonatites

Carbonates	*Phosphates*
Calcite	Apatite
Dolomite	Monazite
Ankerite	*Halides*
Siderite	Fluorite
Bastnasite (Ce,La)FCO_3	Halite
Burbankite	Sylvite
Parisite	*Oxides*
Nyerereite $(Na,K)_2Ca(CO3)_2$	Baddeleyite
Gregoryite $(Na,K)_2CO_3$	Hematite
Silicates	Ilmenite
Amphibole	Magnetite
Aegerine-augite	Perovskite
Diopside	Pyrochlore
Olivine	
Phlogopite	
Sulphides	
Pyrite	
Pyrrhotite	
Galena	
Sphalerite	

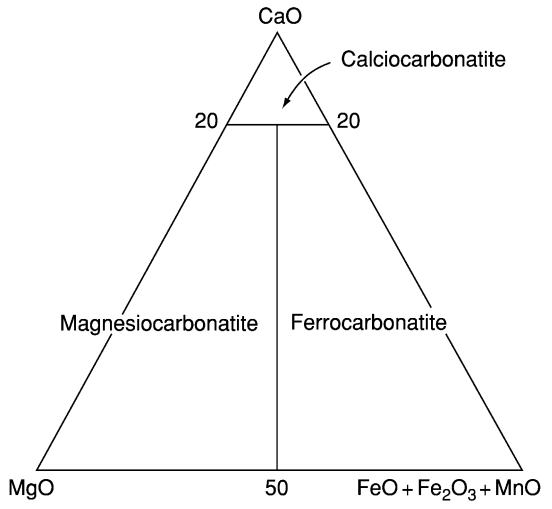


Figure 3 Classification scheme for carbonatites suggested by the IUGS based on chemical composition. Oxides in wt percent.

and calcium carbonates, a variety known as natrocarbonatite. Other minerals that are commonly found in carbonatites include apatite, amphibole, magnetite, mica, olivine, and pyroxene.

As an alternative to using a classification scheme based on mineralogy, carbonatites can also be classified on the basis of their chemical compositions (*see* **Rocks and Their Classification**). Figure 3 shows one such classification scheme that divides carbonatites into calciocarbonatites, magnesiocarbonatites, and ferrocarbonatites. The term silicocarbonatite is sometimes used for those rare carbonatites that contain greater than 20 wt% SiO_2.

Economic Deposits

Carbonatites are a major source of Nb, phosphates (*see* **Sedimentary Rocks:** Phosphates), F, and the REEs. Minerals occur in magmatic, hydrothermal, and supergene deposits. A list of some of the more important deposits is given in Table 4. Carbonatites still constitute the world's greatest source of Nb. Pyrochlore, the most important Nb-bearing mineral, is principally mined in Brazil (e.g., Araxá, Tapira) and some is mined in Canada (St Honoré). However, not all carbonatites contain pyrochlore.

Because carbonate minerals are easily dissolved during chemical weathering, the more resistant minerals, such as pyrochlore and monazite, can be concentrated and easily mined as residual material. Rare earth elements in carbonatites are common and can be found as magmatic minerals or in minerals directly precipitated from hydrothermal solutions. Among such minerals are included monazite, bastnaesite, britholite, burbankite, parisite, and synchesite.

Although most of the world's phosphate production comes from marine phosphorites, the presence of primary apatite in all carbonatites makes them a potential source for fertilisers. More money is generated from this commodity than any other

Table 4 Some examples of mineral deposits associated with carbonatites

Complex	*Commodity*	*Details*
Araxá, Brazil	Nb	Pyrochlore in laterite
St Honoré, Canada	Nb	Pyrochlore in carbonatite
Jacupiranga, Brazil	Phosphate	Primary apatite in carbonatite
Sukulu, Uganda	Phosphate	Residual apatite in weathered carbonatite
Mountain Pass, USA	REE	Bastnaesite in carbonatite
Gakara-Karonge, Burundi	REE	Hydrothermal veins of bastnaesite
Tapira, Brazil	Ti	Anatase in laterite
Powderhorn, USA	Ti	Perovskite in pyroxenite
Phalaborwa, South Africa	Cu + Vermiculite	Vermiculite in pyroxene pegmatoid
Ipanema, Brazil	Vermiculite	Vermiculite in glimmerite
Amba Dongar, India	Fluorite	Hydrothermal overprint on fenite
Okurusu, Namibia	Fluorite	Hydrothermal overprint on fenite

extracted from carbonatites. Apatite, for example, is mined at Khibina in the Kola region of Russia (*see* **Fluid Inclusions**).

The presence of accessory fluorite in many carbonatites shows the importance of F in carbonatitic liquids. Most fluorite deposits, however, are formed by late-stage hydrothermal activity. Among the most important deposits are those at Amba Dongar, India, and Okorusu, Namibia. Fluid inclusion studies from the fluorite at Amba Dongar reveal low crystallization temperatures between 100 and 150°C.

Th and U are also associated with many carbonatite complexes, and are contained within minerals such as thorite and monazite.

Another intriguing economic aspect of carbonatites is the possibility that they may contain diamonds. Diamond-bearing carbonatites have been reported from Uzbekistan and although abundant, the diamonds are very small. This discovery supports any model that proposes that diamonds are formed from strongly compressed carbonatitic melts.

Geochemistry

Carbonatites can be easily separated from other carbonate-rich rocks, especially marbles, by their unusual and distinctive trace element contents. Representative carbonatite analyses are given in Table 5. Among the most characteristic features are high REE (>500 ppm; all are light-REE enriched), high Sr (>700 ppm), Ba (>250 ppm) and V (>20 ppm) contents. Table 6 gives some trace element ratios. Figure 4A, a diagram of Ba + Sr (ppm) versus REE + Y (ppm), clearly discriminates between carbonatites and other rock types. Enrichment in the light REE can be seen from Figure 4B. Normalized trace element data from Oldoinyo Lengai, along with values for average oceanic island basalt, are given in Figure 5.

The Nd, Pb, and Sr isotopic systematics from young carbonatites (<200 Ma) tell a fascinating story, and show that most carbonatites share many isotopic features seen in oceanic island basalts, in spite of the fact that carbonatites are mainly found on continents. Clearly shown is the mantle origin for carbonatites. The two diagrams shown in Figures 6 and 7 illustrate these isotopic similarities, as well as the isotopic heterogeneity shown by carbonatites. Isotopic data from East African carbonatites (Figure 8) provide an interesting case, where the near-linear array reflects the mixing of two end-members, both of which occur in oceanic island basalts. Even within individual centres the spread of isotopic data can be large, particularly if the silicate rocks are included. The idea of a closed system, magma chamber undergoing differentiation cannot be reconciled with the isotopic data. Melts can only be generated by a series of discrete partial melting events that take place within an isotopically heterogeneous mantle.

Because carbonatites date back to at least 3 Ga, they can be used to monitor the chemical evolution of sub-continental upper mantle over a considerable part of the Earth's history. Data from Canada, Greenland, and the Fennoscandian Shield, are shown in Figure 9. Although there is considerable scatter of the data, the low data points define a source for the carbonatites that was clearly established early in the Earth's history, and which has remained isotopically undisturbed since at least 3 Ga. This is also supported by Hf isotope data.

The stable isotopic systematics tell another story about magma differentiation, and magma sources. Unfortunately, carbonatites are particularly prone to isotopic exchange processes involving weathering, alteration under atmospheric conditions, and hydration, so care must be taken to ensure that only fresh samples are analysed.

C and O isotopic data from freshly erupted samples from the Oldoinyo Lengai natrocarbonatite overlap the mantle field based on data from oceanic basalts (Figure 10) and support the conclusion that carbonatites are ultimately derived from a mantle-derived parental melt. For the most part, carbonatites have much lighter C and O than limestones. A comparison is made in Figure 10 among the stable isotope composition of carbonatites from several different continents and limestones of both Phanerozoic and Precambrian age. A convergence towards mantle values is shown by much of the data from carbonatites from Greenland, Europe, North America, and South America.

Few S isotopic data exist for carbonatites. The range of $\delta^{34}S$ values from sulphides from carbonatites are quite variable and show marked differences to mantle values. A single sulphide sample from the Oldoinyo Lengai natrocarbonatite has a δ value of +2.8‰, thus overlapping with the range of sulphide values found in high temperature carbonatites. Sulphides from several Proterozoic calciocarbonatites from Canada, as well as carbonatites from Russia, indicate a large range of values (−7 to +3‰) which coincide with most of the values obtained from other carbonatites. There is some indication that each complex may have its own distinctive average S isotopic composition. In view of these observations, some of the differences in S isotope compositions between carbonatite complexes have been attributed to isotope heterogeneity within the mantle. In contrast to the sulphides, the sulphates are invariably higher in their S isotope compositions (range +4 to +14) reflecting the oxidation state of the S.

Table 5 Chemical compositions of carbonatites

wt%	*Cacite carbonatite*	*Dolomite carbonatite*	*Ferrocarbonatite*	*Natrocarbonatite*
SiO_2	2.72	3.63	4.7	0.16
TiO_2	0.15	0.33	0.42	0.02
Al_2O_3	1.06	0.99	1.46	0.01
Fe_2O_3	2.25	2.41	7.44	0.05
FeO	1.01	3.93	5.28	0.23
MnO	0.52	0.96	1.65	0.38
MgO	1.80	15.06	6.05	0.38
CaO	49.1	30.1	32.8	14.0
Na_2O	0.29	0.29	0.39	32.2
K_2O	0.26	0.28	0.39	8.38
P_2O_5	2.10	1.90	1.97	0.85
H_2O^+	0.76	1.20	1.25	0.56
CO_2	36.6	36.8	30.7	31.6
BaO	0.34	0.64	3.25	1.66
SrO	0.86	0.69	0.88	1.42
F	0.29	0.31	0.45	2.50
Cl	0.08	0.07	0.02	3.40
S	0.41	0.35	0.96	nd
SO_3	0.88	1.08	4.14	3.72
ppm				
Li	0.1	nd	10	270
Be	2	<5	12	nd
Sc	7	14	10	nd
V	80	89	191	116
Cr	13	55	62	<3.0
Co	11	17	26	1.8
Ni	18	33	26	1
Cu	24	27	16	nd
Zn	188	251	606	88
Ga	<5	5	12	<20
Rb	14	31	nd	178
Y	119	61	204	7
Zr	189	165	127	2
Nb	1204	569	1292	28
Mo	nd	12	71	125
Ag	nd	3	3	bd
Cs	20	1	1	6
Hf	nd	3	nd	<0.4
Ta	5	21	1	0.3
W	nd	10	20	49
Au	nd	nd	12	18
Pb	56	89	217	22
Th	52	93	276	4
U	9	13	7	11
La	608	764	2666	545
Ce	1687	2183	5125	645
Pr	219	560	550	19
Nd	883	634	1618	102
Sm	130	45	128	8
Eu	39	12	34	2
Gd	105	nd	130	2
Tb	9	5	16	0.1
Dy	34	nd	52	2
Ho	6	nd	6	0.1
Er	4	nd	17	0.3
Tm	1	nd	2	0.3
Yb	5	10	16	bd
Lu	1	0.1	nd	0.01

bd = below detection; nd = not determined.

Data from Woolley and Kempe (1989), Dawson *et al.* (1995), Keller and Spettel (1995) and Simonetti *et al.* (1997).

Table 6 Trace element ratios

	Calcite carbonatite	*Dolomite carbonatite*	*Ferrocarbonatite*	*Natrocarbonatite*[a]
Ba/Sr	0.41	0.97	3.85	1.44
Ba/La	4.93	7.46	10.8	50.3
Th/U	5.8	7.2	39	1.0
Ce/Pb	30	25	24	12.6
Zr/Nb	0.16	0.29	0.1	0.1
Ba/Nb	2.49	10	22.4	65.6
La/Nb	0.50	1.3	2.1	12
Th/La	0.10	0.75	9.7	0.1
Nb/U	133	44	185	14.4

[a]Data from natrocarbonatites of the June, 1993 eruption of Oldoinyo Lengai with <0.3% silica. From Simonetti *et al.* (1997).

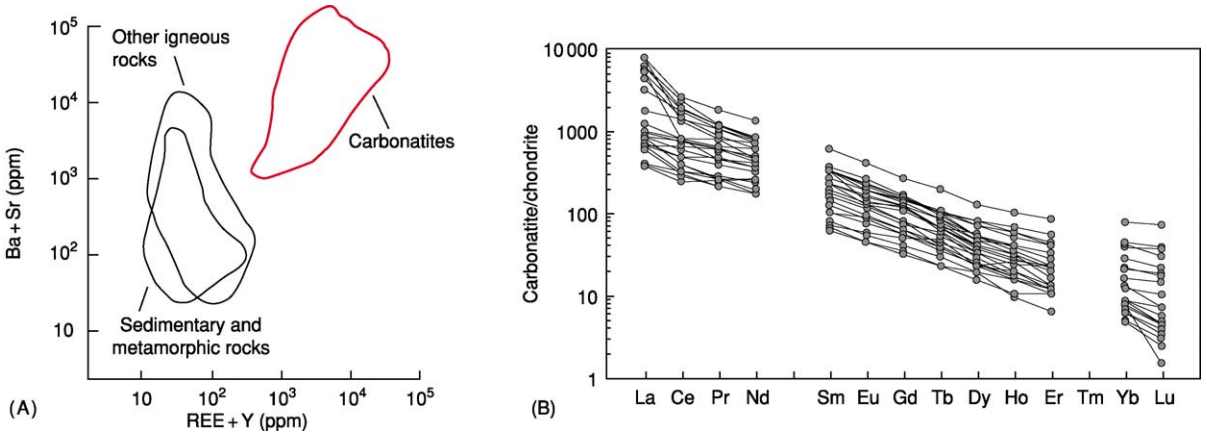


Figure 4 (A) Ba + Sr (ppm) vs REE + Y (ppm). Note the separation of the carbonatites from all other rocks types. From Samoilov (1991). (B) Chondrite-normalised plot for some whole-rock sövite from Europe and North America. Note the similarity in patterns and especially the enrichment in the LREEs.

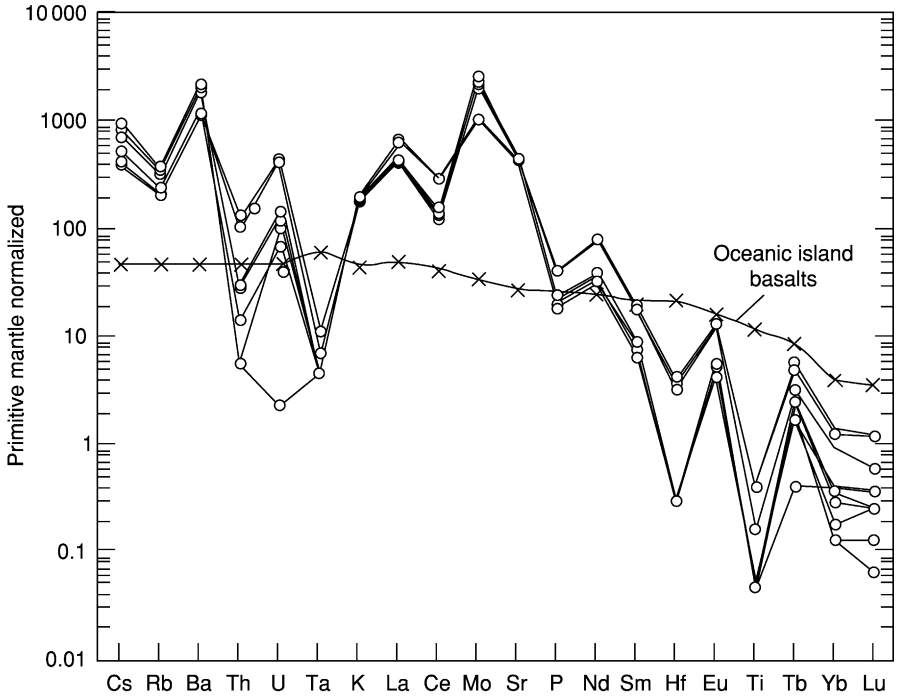


Figure 5 Normalized trace element diagram of the June 1999 lavas from Oldoinyo Lengai, normalized to primitive mantle. Note the peaks for Ba and Sr, and the depletions for Ta, Hf, and Ti. Also an average value is given for oceanic island basalts. From Simonetti *et al.* (1997).

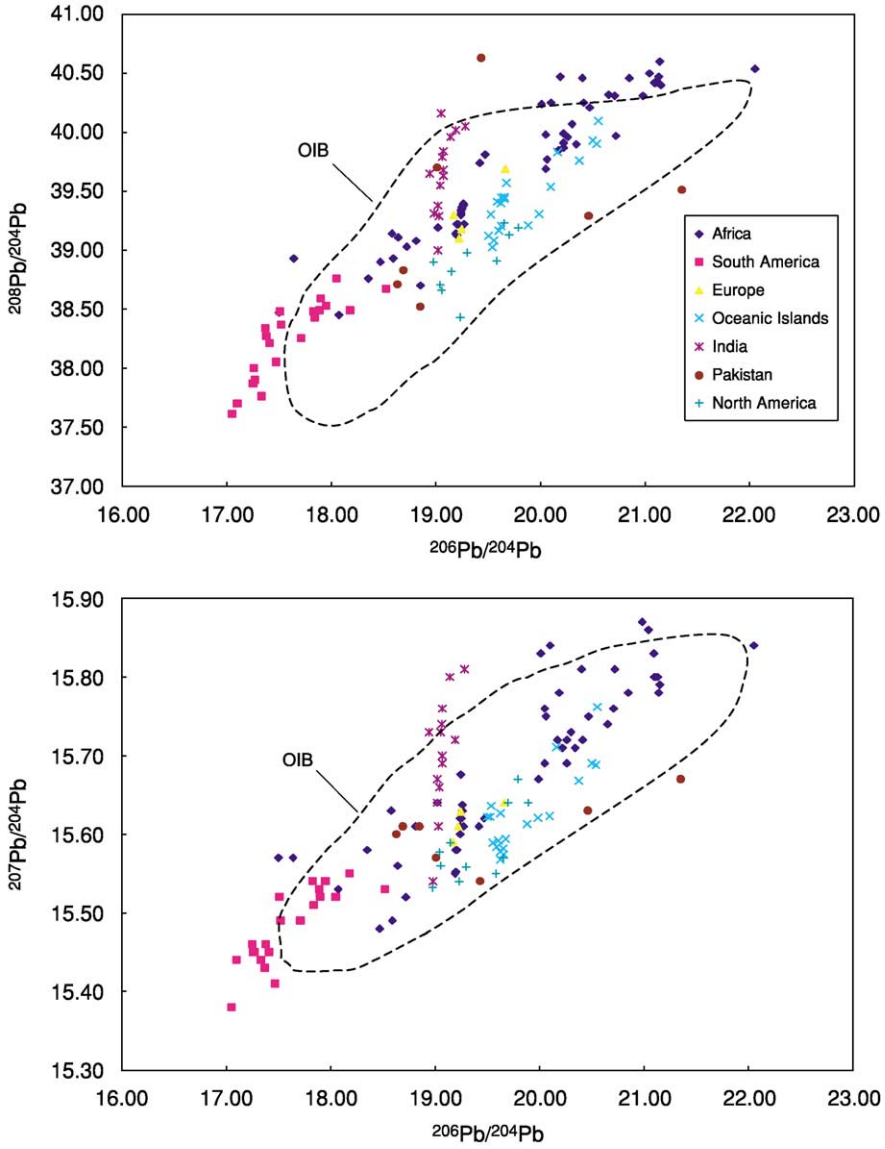


Figure Pb-Pb isotope ratio diagrams of young (<200 Ma) carbonatites. OIB = field for oceanic island basalts. Because these samples are young, the ratios plotted are assumed to be those that existed in the magmas at the time of their formation. Note the similarity of many of the carbonatite data points to those from oceanic island basalts.

Chlorine isotopes have also been measured in carbonatites. Unaltered carbonatites from Africa and Germany have δ^{37}Cl of between −0.8 and +0.1‰. Because the whole mantle δ^{37}Cl is approximately +4.7, the anomalously low chlorine values in carbonatites suggests anomalous mantle involving a crustal component, or processes that fractionate the isotopes, perhaps during magma degassing.

Oldoinyo Lengai

Because Oldoinyo Lengai is the only active carbonatite volcano it deserves special mention. Rising to nearly 3000 m above Tanzania's Eastern Rift Valley, the stratovolcano has a basal diameter of about 12 km and an approximate volume of 60 km^3. During the last few hundred thousand years, the volcano has erupted a series of carbonatitic, phonolitic, and nephelinitic lavas as well as pyroclastic material. Major ash eruptions documented during historic time occurred in 1917, 1926, 1940, and 1966–1967.

The natrocarbonatite flows from Oldoinyo Lengai (Figures 11 and 12) are unusual, both in terms of their mineralogy and their chemical composition. They are made up mainly of two alkali-rich minerals, nyrereite and gregoryite, both of which are

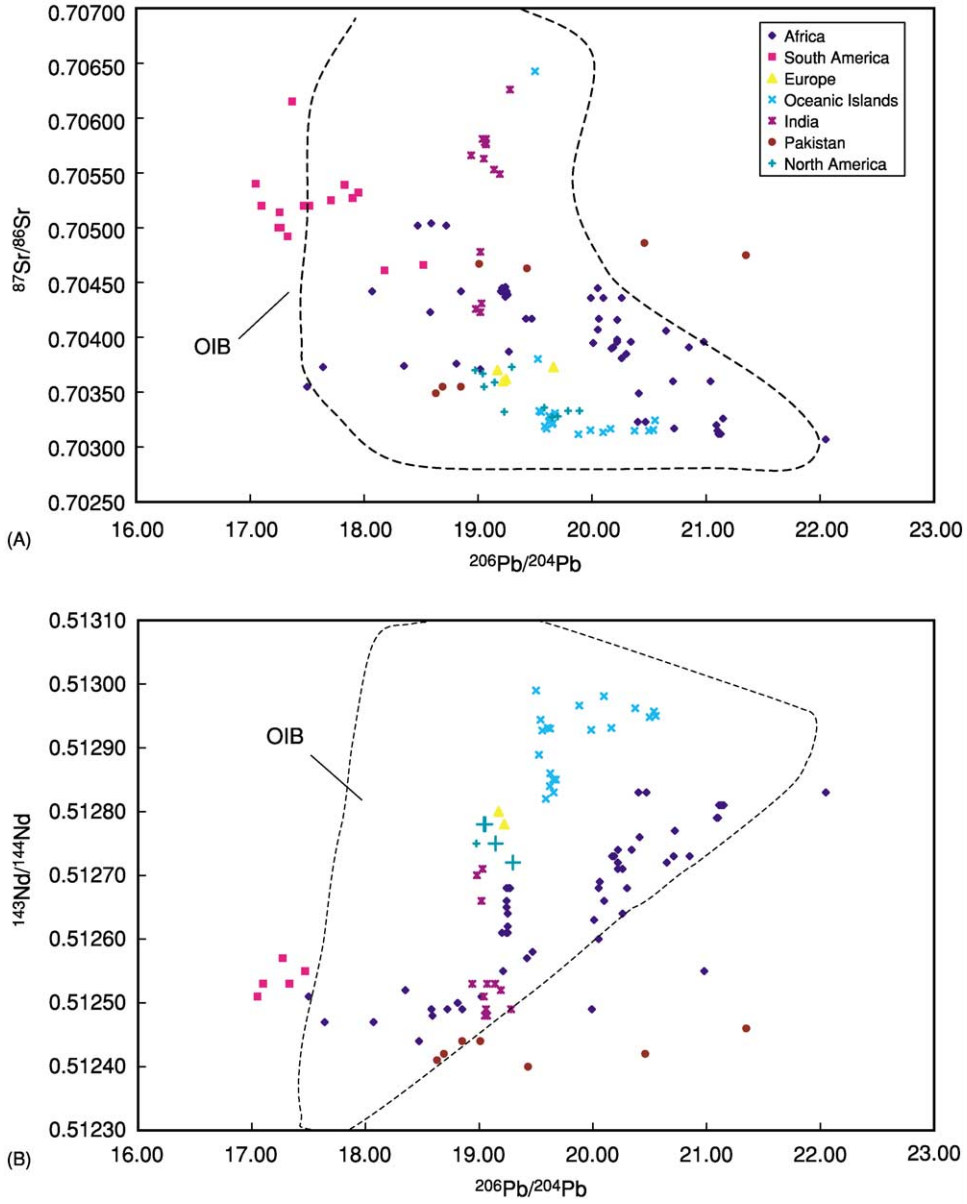


Figure 7 (A) $^{87}Sr/^{86}Sr$ vs $^{206}Pb/^{204}Pb$ for young carbonatites. (B) $^{143}Nd/^{144}Nd$ vs $^{206}Pb/^{204}Pb$ for young carbonatites (<200 Ma). Note the similarities of the carbonatite data with those from oceanic island basalts. A handful of data points for carbonatites with values >0.710 have been omitted.

alkali carbonates with Na well in excess of K. These two carbonates form phenocryst phases, and are present in the groundmass along with fluorite, sylvite, and other minerals. The lavas are black on eruption but because of their hygroscopic nature quickly turn white as the lava reacts with water from the atmosphere.

Natrocarbonatites also have the lowest measured eruption temperatures of any terrestrial magma (490 to 590°C), and they also have a viscosity, thermal diffusivity, specific heat capacity, and latent heat of fusion considerably lower than basaltic lavas. Important features of the trace element geochemistry of natrocarbonatites are enrichment in Ba, Ba/Sr >1, low Zr (<5 ppm), Th/U ~1, LREE normalized values >1000 × chondrites, and $(La/Sm)_N$ >40.

The existence at Oldoinyo Lengai of highly-alkaline carbonatitic liquids (Na-rich) and the spatial association of sodic-rich silicate rocks led to a model involving liquid immiscibility in which a carbonated alkali-rich, silicate parent magma exsolved to form two immiscible liquids. Laboratory studies have shown that immiscibility can indeed take place with liquids of appropriate compositions.

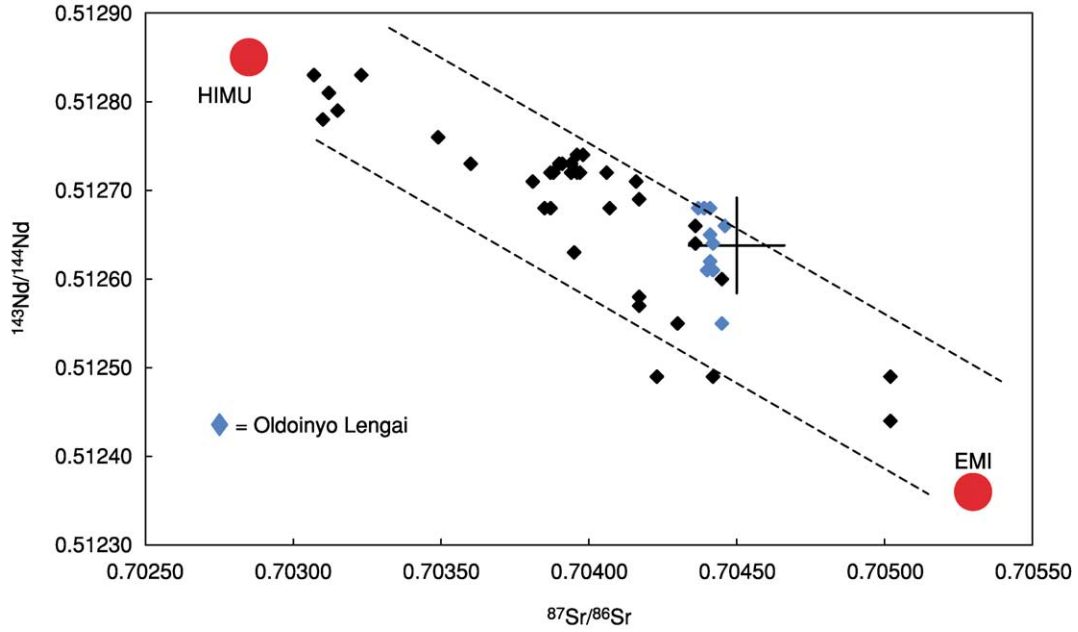


Figure 8 Nd and Sr isotopic data from young East African carbonatites. Note the near-linear array, that suggests mixing between two distinct mantle sources, shown by the red dots, similar to those found in oceanic island basalts (HIMU and EMI). The data from Oldoinyo Lengai cluster close to the cross that marks the position of values for the undifferentiated silicate part of the Earth. Although the natrocarbonatites from Oldoinyo Lengai are quite different in chemical composition to other carbonatites, their isotopic data follow the same trend as other carbonatites from East Africa.

The absence of the high alkali contents in all other types of carbonatite raises doubts about the seminal role that Oldoinyo Lengai plays in understanding carbonatite genesis. It was once thought that all carbonatitic melts were alkali rich, and that alkalies were lost during fluid migration from the melt during fenitisation. Fascinating as the findings from Oldoinyo Lengai may be, in an introduction to a collection of papers dedicated solely to Oldoinyo Lengai, the consensus was that "natrocarbonatites should be considered the result of an extreme process rarely encountered in nature".

Mode of Origin

There has been a great deal of debate about the origin of carbonatites, and this is still a matter of ongoing concern. On the basis of evidence from phase equilibrium studies, field relationships, and geochemistry, there are three possible ways of producing carbonatitic melts. Firstly, they can be direct partial melts of a mantle containing carbonate phases. Secondly, they can be produced by liquid immiscibility from a carbonated alkali-rich melt. And finally, they can be generated by extreme crystal fractionation of a silicate magma. Studies involving geochemistry, mineralogy, petrography, and field relationships show that even within one igneous province, such as the Kola Alkaline Province, or even within an individual complex, it is likely that carbonatites are produced in more than one way.

As to how individual carbonatites form is a difficult question to answer, since we still lack good criteria to separate one mechanism from another. If carbonatitic melts are primary liquids, they are probably dolomitic in composition, have a Ca/Ca + Mg ratio of between 0.5 and 0.7 and a silica content of 4–5 wt%. However, many dolomite carbonatites can be found as late stage veins, dykes, and stringers cutting older calcite carbonatites.

The Source(s)

One of the more important questions concerning carbonatites is the site of generation of their parental liquids. At first it was thought that the carbonated melts were generated by melting metasomatized lithosphere, but this is not supported by the radiogenic isotopic data, which would require the involvement of more than one mantle source similar to those reflected in oceanic island basalts (Figure 8).

The mantle source for carbonatites must contain carbonate phases. The stability fields of Mg-rich calcite, dolomite ($CaMg(CO_3)_2$), and magnesite

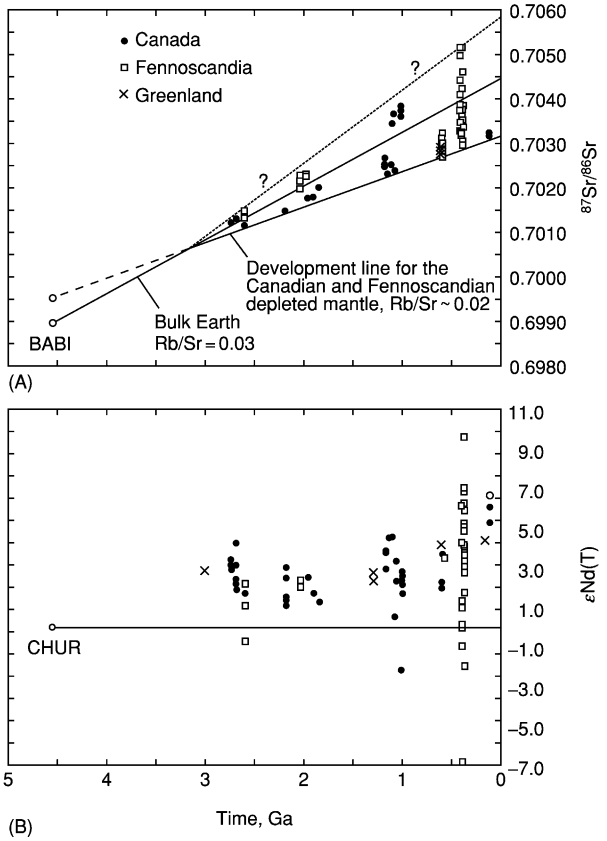


Figure 9 Development lines for Sr and Nd. (A) $^{87}Sr/^{86}Sr$ ratio versus time is shown for Canadian and Fennoscandian carbonatites. The development line indicates an ancient mantle reservoir at least 3.0 Ga old. The Earth is considered to have started with a $^{87}Sr/^{86}Sr$ ratio similar to that found in basaltic achondritic meteorites. The initial value is shown by the point marked BABI. (B) $\varepsilon Nd(T)$ versus time where $\varepsilon Nd(T) = (^{143}Nd/^{144}Nd_{sample} - ^{143}Nd/^{144}Nd_{CHUR})/^{143}Nd/^{144}Nd_{CHUR}$ where $^{143}Nd/^{144}Nd_{sample}$ is the initial ratio in the sample and $^{143}Nd/^{144}Nd_{CHUR}$ is the initial ratio of CHUR at the same point in time. Any of the data points above the horizontal line reflects an ancient reservoir that has undergone depletion relative to achondritic meterorites. Note that the depletion event must have occurred before 3.0 Ga. From Bell and Tilton (2002).

($MgCO_3$) make these minerals ideal candidates for trapping CO_2 in the mantle. Whether the carbonate is primary, in other words present in the mantle since the Earth formed, or whether it is recycled (*see* **Carbon Cycle**) during subduction into the mantle in the form of limestone or carbonated sea floor material, is an interesting question. From the results of recent experiments, however, it now seems that carbonates from a carbonated eclogite (the metamorphosed equivalent of basalt plus calcite) will probably be removed from the downgoing slab before reaching a depth of 300 km. Thus, the possibility of carbonates being recycled into the transition zone and into the deep mantle seems unlikely.

Figure 13 shows an example of a volatile-rich mantle containing H_2O and CO_2. At depths of about 75 km the solidus forms a shoulder or ledge under conditions of near constant pressure, which coincides with solubility of the $MgCO_3$ component in the magma when dolomite becomes stable under mantle conditions. Although there is still some question about the exact depth of this ledge and its shape, there is general agreement that it does exist and that it plays a key role in suppressing volatile-rich melts as they attempt to reach the surface. At pressures below this ledge, very low degrees of partial melting will generate a dolomitic carbonatite magma. As the temperature of melting progressively increases, and hence the degree of partial melting, the melt compositions become progressively more silica-rich, yielding melts of lamprophyric, melilitic, and nephelinitic compositions.

As we have seen, most carbonatites have mantle isotopic signatures similar to those found in oceanic island basalts. Far from being constant in their isotopic composition, the heterogeneities found in oceanic island basalts have been generally attributed to material subducted into the deeper parts of the mantle, that have been allowed to age over geologic time. The isotopic data restrict the mantle source well below the source region of mid-ocean ridge basalts.

The task of evaluating the depths of the mantle sources on the basis of Sr, Nd, and Pb is difficult, if not impossible. Fortunately, noble gas data have come to the rescue and provide some indications about the relative depths of the source material. Noble gases (e.g., Ar, Ne, Kr, Xe) trapped in minerals, such as apatite and magnetite, point to a relatively primitive mantle, which is less degassed than the source that generated mid-ocean ridge basalts. Of particular interest are the ^{129}Xe anomalies, low $^{40}Ar/^{36}Ar$ and $^{4}He/^{3}He$ ratios, and $^{20}Ne/^{22}Ne$ vs $^{21}Ne/^{22}Ne$ regressions similar to those from Loihi, Hawaii, that contain a known plume component. Wherever the source is in the mantle, it appears that it has not had the opportunity to mix with gases from our present atmosphere, nor did it let the ^{129}Xe escape.

Plumes and Carbonatites

Attention has recently focused on the relationship between carbonatites and mantle plumes, and although perhaps not all are plume-related, there is some evidence to support this idea. Evidence for plume-related magmatism includes: (i) the similarity of isotope signatures between oceanic island basalts and carbonatites; (ii) the association of carbonatites

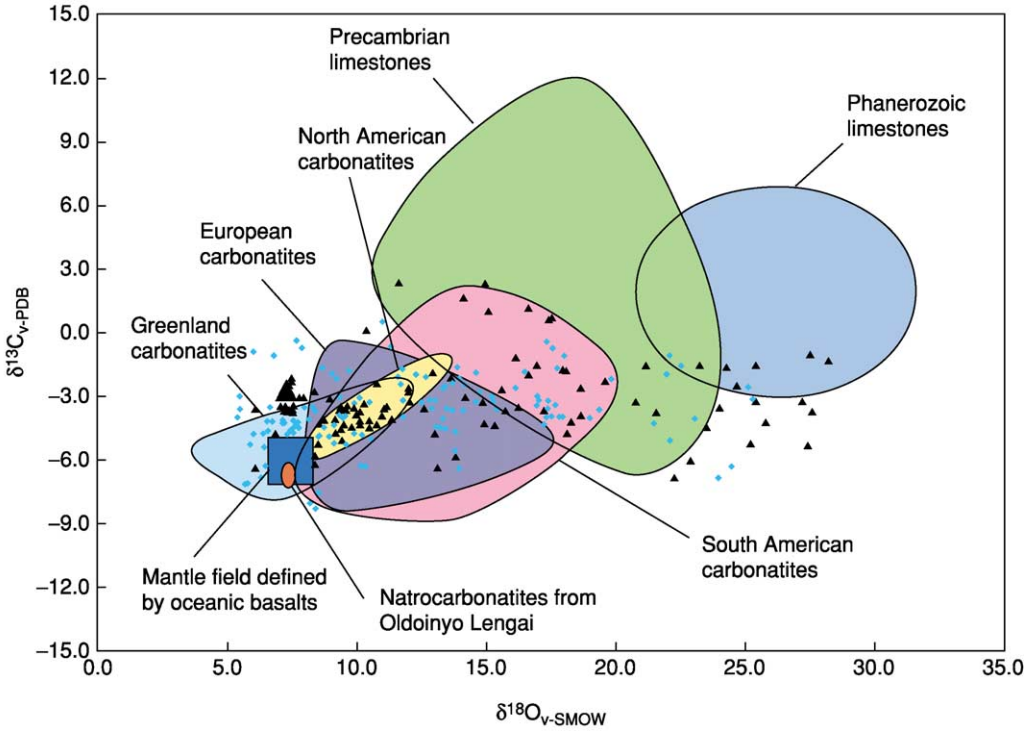


Figure 10 Stable isotope data. Comparison between limestone and carbonatite isotopic signatures. Note that the isotopic data from carbonatites end at the mantle box based on oceanic basalts. Each of the fields enclose >90% of the data points for each particular group. The data from Africa carbonatites are marked by diamonds, those from India by triangles.

with two known oceanic plumes, the Canary and the Cape Verde; (iii) the temporal and spatial relationships of some carbonatites to flood basalt provinces (for example the Deccan, Siberian, and Keeweenawan events); and (iv) the primitive natures of some of the noble gas data. Plume-related magmatism has also been proposed for the young carbonatites in East Africa, including Oldoinyo Lengai.

One issue in understanding carbonatite magmatism is the role that the continental lithosphere plays in their origin. Some plume models involve mixing between material from the plume and the overlying lithosphere, whereas others regard the sole role of the lithosphere as being one of mechanically constraining volatiles within a plume head. Plume heads that become attached to the lithosphere might even form sites for the generation of carbonated melts.

Based on models that have been proposed for plume-related magmatism, carbonatites could well be generated from the cooler, volatile-rich parts of a plume head. With the presence of the shoulder on the volatile-rich, mantle solidus at depths of about 75 km, the release of volatiles may induce crack propagation and explosive activity and promote the movement of low viscosity silicate and carbonatitic melts. A schematic representation of this is shown in Figure 14.

Metasomatism

Migration of volatiles and fluids (probably aqueous), at the margins of many carbonatites, infiltrate and metasomatise the surrounding country rocks. These rocks become highly fractured, and new minerals are developed (such as feldspar, sodic pyroxenes, and alkali amphiboles) as the volatiles and fluids chemically interact with the country rocks. The process of metasomatism at crustal levels is called fenitisation and the resulting rocks fenites, named after the Fen complex in Norway, where they were first recognized. Potassic metasomatism is considered to characterize high level, lower temperature metasomatism, whereas the sodic type is characteristic of unspecified deeper levels and higher temperatures. It has even been suggested that some syenitic rocks may represent the ultimate products of fenitisation.

The low viscosity that characterizes carbonatitic melts along with their low dihedral wetting angles make them ideal agents for metasomatism within the mantle. With their high reactivity, such melts can

Figure 11 (A) Oldoinyo Lengai, Tanzania during one of its phases of explosive activity (Photo: JB Dawson, taken August, 1966). Ash clouds were reported by aircraft at flight levels of 10 700 m. Oldoinyo Legai rises about 2886 m above the floor of the East African Rift Valley System, and was first shown on a map compiled by two missionaries in 1855. (B) Aerial view from the north-west of ash cone that formed in the crater in 1966 (Photo: JB Dawson, taken August, 1966).

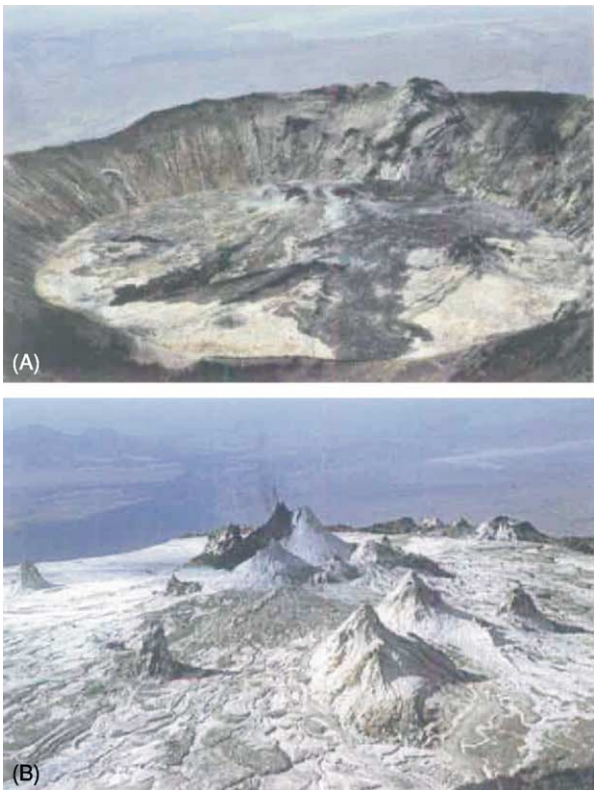


Figure 12 (A) Active north crater of Oldoinyo Lengai looking north from the summit (Photo: J Keller, taken June, 1989). Lake Natron can be seen in the background. The crater is about 250 m across. On far side of crater are two active vents erupting black, very fluid natrocarbonatite lava flows. Inactive parasitic vents with lava flows occur on the rim of the crater. Note the white colour of the weathered carbonatite flows. (B) Looking into the north crater from the summit (Photo: J Keller, taken October, 2003). Upper left the rift escarpment, upper right Lake Natron (2300 m below!) The crater is full, and compared with **Figure 12A** hardly any of the crater wall is visible. The active vents have formed spatter cones. Note the very fluid lava flows in the foreground.

change the mineralogy of the mantle, and melts may be changed themselves, undergoing 'chemical death' without reaching the surface. Reactions that can take place within the mantle are:

$$3\text{dolomite} + \text{diopside} = 2\text{forsterite} + 4\text{calcite} + 2CO_2$$

$$\text{dolomite} + 4\text{orthopyroxene} = \text{clinopyroxene} + \text{olivine} + CO_2$$

Incursions of carbonatitic melts can have a profound effect, not only on the mineralogy of the mantle, but also on the chemical composition of the mantle. If carbonatitic melts reach the lithosphere, minerals can precipitate to generate a veined mantle with unusual mineralogy. Melting of these veins, along with the wall rocks that surround them, can generate unusual magmas that may lead to

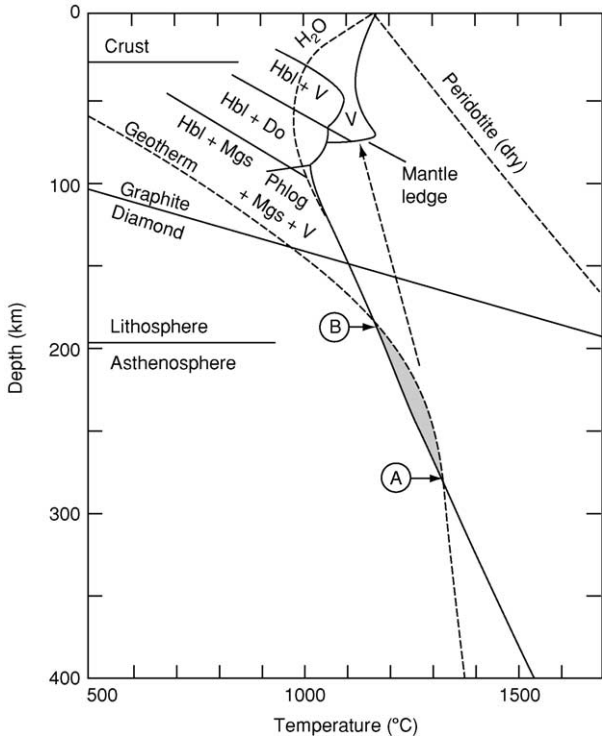


Figure 13 Generalized diagram of the system peridotite +C+H+O. Note the position of the mantle ledge at a depth of about 75 km. Melts are generated between A and B where the geotherm intersects the mantle solidus. Melts at relatively low temperatures will hit the mantle ledge producing metasomatism and possibly explosive activity. System is one in which CO_2 is well in excess of H_2O. Phlog = phlogopite, Do = dolomite, Hbl = hornblende, Mgs = magnesite and V = vapour. Adapted from Wyllie *et al.*(1990).

melts of kimberlitic, lamproitic, and lamprophyric affinity (*see* **Igneous Rocks:** Kimberlite).

Phase Equilibrium Studies

Numerous experiments have been carried out in various laboratories using both natural and synthetic mixtures to help understand the melting conditions needed to generate carbonatitic melts, and to monitor the evolution of carbonated melts within a magma chamber. Using the system CO_2-H_2O-CaO-($\pm Al_2O_3$)-MgO-SiO_2 as a reasonable approximation to a volatile-rich mantle peridotite, we can place some constraints on the origin and evolution of carbonatitic melts. Such studies show that it is indeed possible to produce carbonatitic liquids by: (i) small degrees of partial melting of a carbonated peridotite, by liquid immiscibility; and (ii) by fractional crystallization of a carbonated silicate melt.

Primary carbonatitic melts can be generated at depths of >200 km and these have compositions that lie close to dolomite. As the melt migrates to the surface, at a depth of about 75 km, the magma 'freezes' and volatiles are given off that can generate explosive diatremes. If the melt reacts to produce wehrlite then the melt can become increasingly enriched in $CaCO_3$ and SiO_2 and hence siliceous carbonatite might escape to the surface.

If the parental melt is a carbonated silicate melt, such as an olivine melanephelinite, such a liquid can take various paths. One path can lead to the silicate-carbonatite liquidus field boundary and the liquid can precipitate out cumulate carbonatites. Another path at relatively low pressures can lead to a miscibility gap, and can exsolve an immiscible carbonate-rich melt. These phase relations are shown in Figure 15. Many systems indicate that a carbonatite produced in this way can only contain a maximum of 80% $CaCO_3$. With CO_2 saturation, immiscible separation can take place at much deeper levels. If the high temperature, immiscibly-derived liquid differentiates then it must trend towards enrichment in $(Na,K)_2CO_3$. Magnesiocarbonatite magmas are also capable of precipitating calciocarbonatite rocks. In addition, it should not be overlooked that many calciocarbonatites may represent crystal cumulates.

Concluding Remarks

Carbonatites must originate from a volatile-rich mantle (*see* **Earth:** Mantle), and the obvious choice of minerals containing the CO_2 necessary for their origin are calcite, dolomite, and magnesite. Thinking in much broader terms about carbonatites, one key question involves the carbon cycle (*see* **Carbon Cycle**). Does carbonatitic magmatism reflect recycling of C into the mantle by the subduction of carbonate-bearing material such as limestone or carbonated eclogite, or does it simply reflect an attempt by the Earth to purge itself of primitive volatiles? On the basis of isotopic data, especially the noble gas data, crustal recycling seems to be unlikely and we are more inclined to turn to those models that involve outgassing of primitive material.

It is difficult to imagine a primitive Earth without a CO_2-rich atmosphere. It seems reasonable to assume that some CO_2 was trapped during the Earth's accretionary history and remnants of this are still retained within the lower parts of the mantle. Our two neighbouring planets, Venus and Mars, have CO_2 rich atmospheres, and there is no reason to believe that the Earth was any different from them in its early history (*see* **Solar System:** Mars; Venus). Some of the volcanic features observed from Magellan imagery (Figure 16), especially the channels (canali), have been considered to be the result of carbonatitic magmatism.

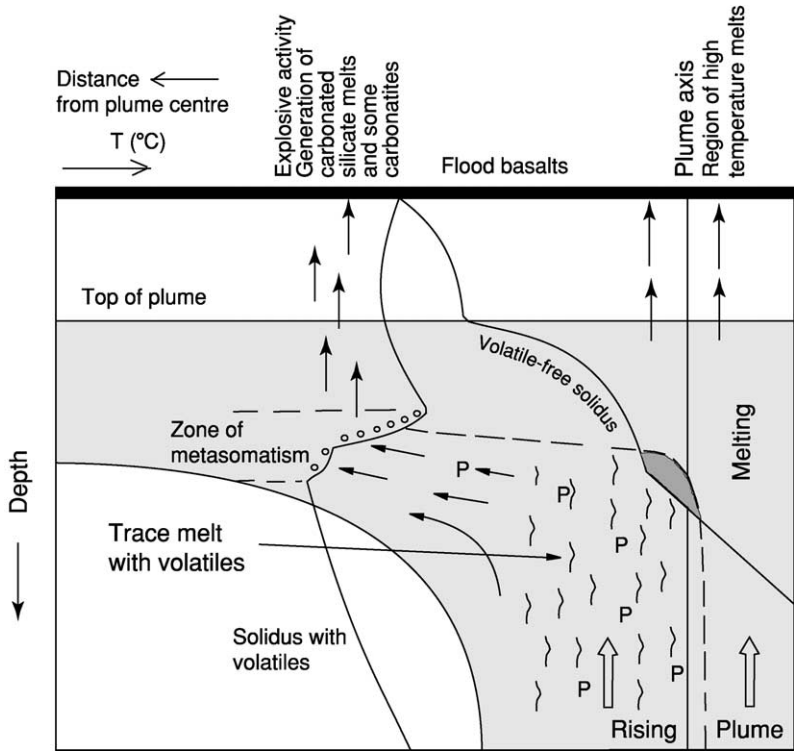


Figure 14 Plume head showing the position of mantle solidii, one vapour-free and the other vapour-rich. Migration of volatile-rich melts from the cooler, outer parts of the plume hit the mantle ledge producing metasomatism, explosive activity, and the possible release of low viscosity melts. P represents material from the deeper parts of the mantle. Adapted from Wyllie (1988).

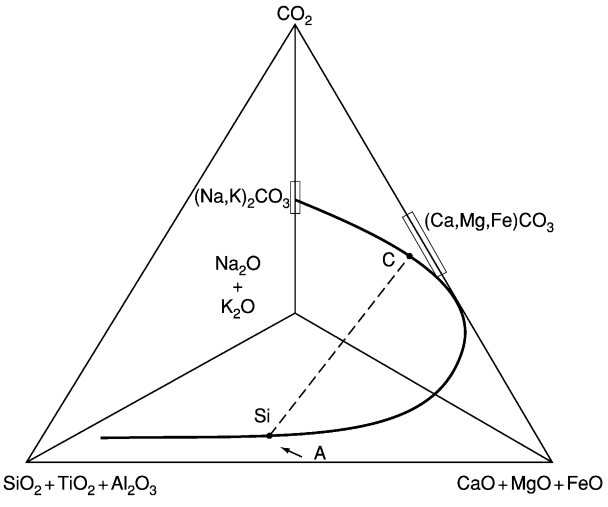


Figure 15 Quaternary diagram showing liquid trends in a silicate-carbonate system. Solid curved line is a hypothetical solvus. Point A represents a carbonated silicate melt which on cooling and fractionation moves into the two liquid field where carbonatite of composition C is exsolved. This coexists with a silicate liquid of composition Si. The dashed line connects the conjugate silicate-carbonate pair. Adapted from Kjarsgaard and Hamilton (1989).

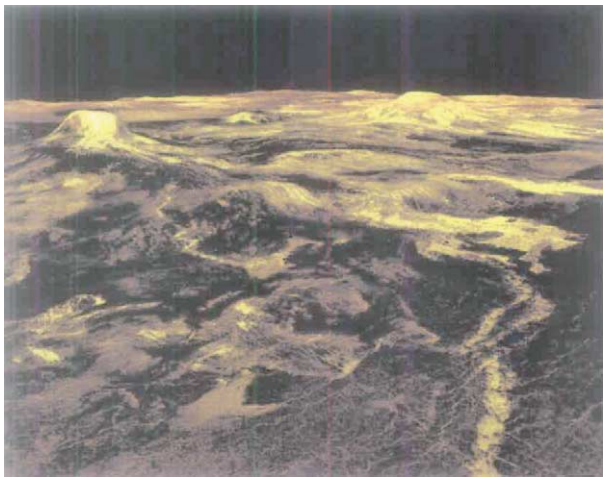


Figure 16 Lavas on Venus. This 3D, false colour, computer generated, radar image shows Venusian lava flows (yellow in colour) extending for at least 700 km from Gula Mons (upper left), a volcano about 3 km high. The other volcano is Sif Mons that lies about 730 km from Gula Mons. Note the intense fracturing in the right foreground. (JPL image PIA 00102; Magellan mission to Venus.)

See Also

Carbon Cycle. **Earth:** Mantle. **Igneous Rocks:** Kimberlite. **Lava**. **Minerals:** Carbonates. **Rocks and Their Classification**. **Sedimentary Rocks:** Phosphates. **Solar System:** Venus; Mars. **Volcanoes**.

Further Reading

Bailey DK (1993) Carbonatite magmas. *Journal of the Geological Society of London* 150: 637–651.

Bell K (ed.) (1989) *Carbonatites: Genesis and Evolution*. Unwin Hyman.

Bell K and Keller J (eds.) (1995) *Carbonatite Volcanism*. Springer-Verlag.

Bell K and Tilton GR (2002) Probing the mantle: the story from carbonatites. *Eos* 83.

Dawson JB, Pinkerton H, Norton GE, Pyle DM, Browning P, Jackson D, and Fallick AE (1995) Petrology and geochemistry of Oldoinyo Lengai lavas extruded in November 1988: magma source, ascent and crystallization. In: Bell K and Keller J (eds.) *Carbonatite Volcanism,* pp. 47–69. Springer-Verlag.

Harmer RE and Gittins J (1998) The case for primary, mantle-derived carbonatite magma. *Journal of Petrology* 39: 1895–1903.

Keller J and Spettel B (1995) The trace element composition and petrogenesis of natrocarbonatites. In: Bell K and Keller J (eds.) *Carbonatite Volcanism,* pp. 70–86. Springer-Verlag.

Kjarsgaard BA and Hamilton DL (1989) The genesis of carbonatites by immiscibility. In: Bell K (ed.) *Carbonatites: Genesis and Evolution,* pp. 388–404. Unwin Hyman.

Kogarko LN, Kononova VA, Orlova MP, and Woolley AR (1995) *Alkaline Rocks and Carbonatites of the World. Part 2: Former USSR*. Chapman and Hall.

Le Bas MJ (1977) *Carbonatite-Nephelinite Volcanism*. John Wiley and Sons.

Le Bas MJ (1984) Oceanic carbonatites. In: Kornprost J (ed.) *Kimberlites 1: Kimberlites and Related Rocks, Proceedings of the Third International Kimberlite Conference*, pp. 169–178. New York: Elsevier.

Sage RP and Watkinson DH (1991) Alkalic rock-carbonatite complexes of the Superior Structural Province, northern Ontario, Canada. *Chronique de la Recherche Minière* 504: 5–19.

Samoilov VS (1991) The main geochemical features of carbonatites. *Journal of Geochemical Exploration* 40: 251–262.

Simonetti A, Bell K, and Shrady C (1997) Trace- and rare-earth-element geochemistry of the June 1993 natrocarbonatite lavas, Oldoinyo Lengai (Tanzania): implications for the origin of carbonatite magmas. *Journal of Volcanology and Geothermal Research* 75: 89–106.

Wall F and Zaitsev AN (eds.) (2004) Phoscorites and carbonatites from mantle to mine: the key example of the Kola Alkaline Province. *The Mineralogical Society Series,* 10.

Woolley AR (1987) *Alkaline Rocks and Carbonatites of the World. Part 1: North and South America*. University of Texas Press.

Woolley AR (2001) *Alkaline Rocks and Carbonatites of the World. Part 3: Africa*. London: The Geological Society.

Woolley AR and Kempe DRC (1989) Carbonatites: nomenclature, average chemical compositions, and element distribution. In: Bell K (ed.) *Carbonatites: Genesis and Evolution,* pp. 1–14. Unwin Hyman.

Wyllie PH (1988) Solidus curves, mantle plumes, and magma generation beneath Hawaii. *Journal of Geophysical Research* 93: 4171–4181.

Wyllie PJ, Baker MB, and White BS (1990) Experimental boundaries for the origin and evolution of carbonatites. *Lithos* 26: 3–19.

Granite

A I S Kemp, University of Bristol, Bristol, UK

Introduction

The transformation of the Earth's earliest mafic to ultra-mafic crust into the stable high-standing continental landmasses was achieved by granitic plutonism. This involves the generation and ascent of incompatible element-rich silicic magmas, which leave behind dense dehydrated residues that either accumulate in the deep crust or sink into the mantle (**Figure 1**). Dating of the highly resistant mineral zircon, which crystallizes from silicic magmas, hints that this differentiation process was underway by approximately 4.4 Ga, shortly after the Earth's formation, and it continues to the present day. Unlike their mafic counterparts basalts and gabbros, granitic rocks are composed of light silica-rich minerals and are difficult to destroy by subduction or tectonic delamination. Granitic rocks therefore preserve a continuous record of the changing thermal and possibly atmospheric conditions throughout Earth's history.

In broad terms, granite is a crystalline plutonic igneous rock consisting essentially of quartz and feldspar, in which the relative proportion of the former is between 20% and 60%. Yet, this simple definition belies the remarkable variety of colour, grain size, texture, mineralogy, and composition that is intrinsic to the granitic family. Furthermore, granite bodies are fossil magma chambers and as such provide a 'snapshot' of part of a dynamic magmatic system; however,

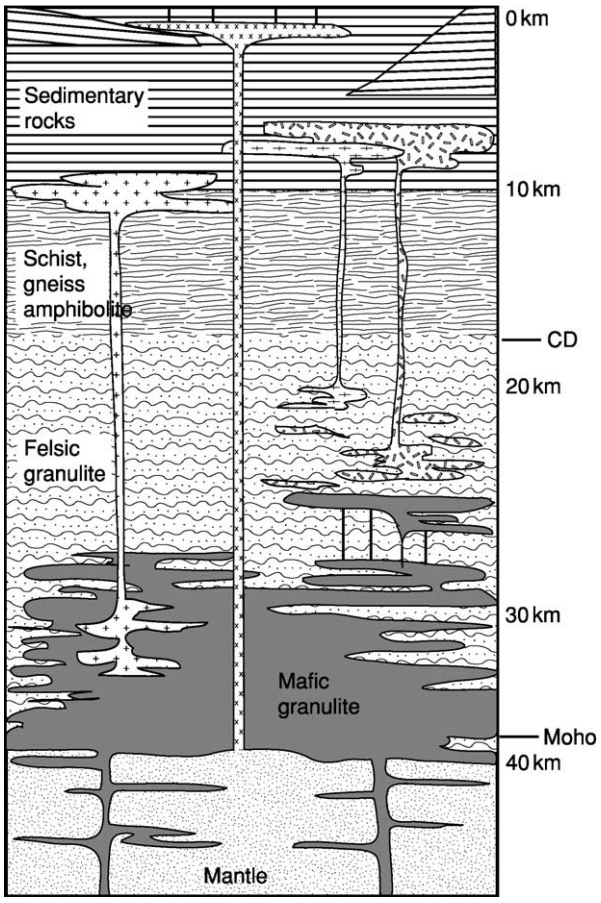


Figure 1 Simplified view of crustal differentiation during partial melting and granite formation. Granitic rocks have crossed patterns. CD, Conrad Discontinuity. (Adapted from Johannes W and Holtz F (1996) *Petrogenesis and Experimental Petrology of Granitic Rocks*. Berlin: Springer-Verlag.)

as with animal fossils, the picture is often incomplete and potentially unrepresentative. Valuable information lies encoded in textures and in mineral and bulk-rock compositions, though its interpretation is rarely straightforward.

For these reasons, models explaining the nature and origin of granitic rocks have always been contentious. Debate culminated in the high-profile and sometimes acrimonious 'granite controversy' of the early twentieth century, waged between the transformists and the magmatists. This focused on the perceived 'room problem' for the emplacement of granitic magmas, the role as fluids as 'granitizing' agents, the link between granitic and volcanic rocks, and whether the large amounts of granite observed could be differentiated from basaltic liquid. Concerning this last aspect, a prominent disputant, H H Read, remarked that it 'degrades granitic magma to the status of the dregs of the primary basaltic', epitomizing the polarity of views expressed at this time.

Nonetheless, the debate stimulated conceptual advances in granite geology and in the Earth sciences in general. For example, the experimental studies of N L Bowen and co-workers that firmly resolved the debate in the magmatists favour introduced some fundamental concepts, such as fractional crystallization and partial melting, that remain the cornerstone of modern igneous geochemistry and the key to understanding crustal differentiation. With the advent of plate tectonics, it became possible to place granitic magmatism into a wider geodynamic framework, whereby the physical processes driving the extraction, transport, and emplacement of granitic magmas became quantifiable. This framework has been coupled with thermal models, to constrain the nature of the heat sources required to generate granitic magmas and the most favourable tectonic scenarios for this to occur. Modern studies seek to integrate these physical and geochemical parameters and ultimately to build a picture of the source rocks of the granite plutons, the way in which these melted to generate the silicic magma, and how this traversed the crust and acquired its present mineralogy and chemical composition.

Classification Schemes

Rocks of the granite family exhibit a continuum of compositions, reflecting derivation from diverse and multiple crustal and mantle sources, all of which are typically a matter of deduction. It is therefore prudent to classify granitic rocks according to parameters that can be directly observed and measured, such as modal mineralogy and chemical composition.

The simplest subdivision is made on the basis of the ratio of alkali feldspar to plagioclase, giving rise to the alkali feldspar granite, granite (*sensu stricto*), granodiorite, and tonalite categories. The abundance of mafic minerals typically increases with the proportion and anorthite content of the plagioclase.

A host of chemical classification schemes also exist. Of these, the alumina saturation index (ASI), based on the molecular ratio of Al_2O_3 to $(CaO + Na_2O + K_2O)$, is particularly useful, since it can be tied readily to mineralogy (Table 1). Further subdivision can be made on the basis of alkali–lime relationships, which has the advantage of highlighting the differentiation trends shown by different granitic suites and the links to potential parental magmas (Figure 2).

Finally, granites may be classified according to their oxidation state, which is manifested in the ratio of ferrous to ferric iron (Fe^{2+}/Fe^{3+}) or more simply in the assemblage of iron oxides and mafic silicates present relative to a reference reaction or 'buffer'. The most useful oxygen buffer for granitic rocks is the fayalite–magnetite–quartz (FMQ) reaction (Figure 3).

Table 1 Classification of granitic rocks according to the alumina saturation index (ASI), listing the diagnostic mineralogy and common accessory phases occurring in rocks of each category (accessory zircon and apatite are ubiquitous). Note that metaluminous magmas will become weakly peraluminous by fractionation of calcic phases (e.g. amphibole) or assimilation of aluminous sediments. In applying this classification, care must be taken to ensure that the ASI is determined from pristine samples, as subsolidus alteration and weathering can leach soluble CaO and Na_2O, thereby enriching the sample in immobile Al_2O_3 and elevating the ASI beyond its primary magmatic value

Category	*ASI*	*Diagnostic mineralogy*	*Biotite colour*	*Accessory phases*
Strongly peraluminous	>1.1	Cordierite, garnet, muscovite, andalusite sillimanite, topaz	Rusty red-brown	Monazite, xenotime, uraninite
Weakly peraluminous	1.0–1.1	Biotite, aluminous amphibole (cummingtonite or gedrite), calcic amphibole	Red-brown to dark brown	Monazite, titanite, allanite
Metaluminous	<1.0	Clinopyroxene, hornblende, epidote, calcic plagioclase (to An_{90})	Dark chocolate brown to greenish brown	Titanite, allanite
Peralkaline	<1.0	Alkali amphibole, aegirine, albitic plagioclase	Very dark brown to black	Fluorite, titanite, allanite, thorite

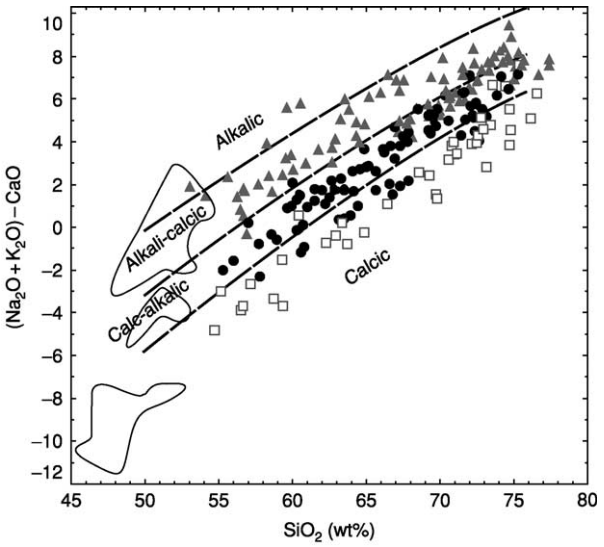


Figure 2 Subdivision of granitic rocks according to the modified alkali lime index of Frost BR, Barnes CG, Collins WJ, *et al.* (2001), as given by a plot of $(Na_2O + K_2O) - CaO$ versus SiO_2. Note that the trends defined by the three granitic suites – the subduction-related Peninsula Ranges (mostly calcic; open squares) and Sierra Nevada (mostly calc-alkalic; filled circles) batholiths (western USA) and the syn- to post-collisional Caledonian plutons of the UK (mostly alkali-calcic; grey triangles) – project at low silica towards coeval mantle-derived rocks in the three areas (circled fields).

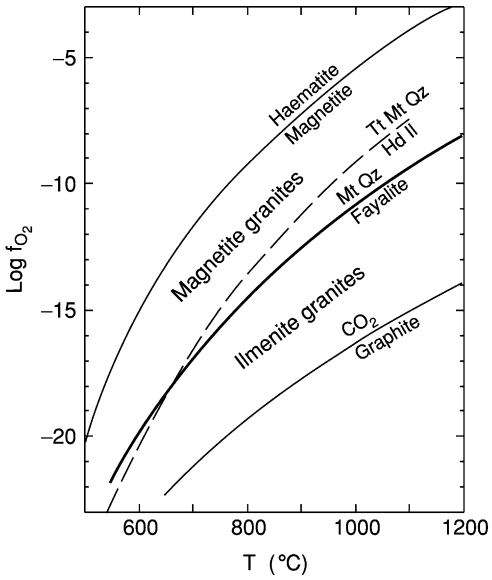


Figure 3 Plot of log f_{O_2} (oxygen fugacity) against temperature, showing the locations of commonly used buffer curves and the crystallization fields of ilmenite- and magnetite-bearing granitic rocks. Hd, hedenbergite; Il, ilmenite; Mt, magnetite; Qz, quartz, Tt, titanite.

Granitic magmas located below FMQ are considered to be reduced and crystallize ilmenite ($Fe^{2+}TiO_3$), whereas magnetite ($Fe^{2+}Fe_2^{3+}O_4$) appears in those relatively oxidized magmas that have a higher oxygen fugacity (f_{O_2}) than FMQ. Ilmenite is destabilized at even higher f_{O_2}, whereupon magnetite is accompanied by titanite as the major titanium-bearing phase. Mafic silicates (e.g. hornblende, pyroxene) tend to be enriched in magnesium in strongly oxidized granites, since the iron is sequestered by magnetite.

Determining the oxidation state of granitic rocks is important for three reasons. First, it has been suggested that the oxygen fugacity of a granitic magma reflects its particular source rocks, and so this parameter can provide valuable petrogenic information. Second, the differing magnetic susceptibilities of oxidized and reduced granites result in contrasting aeromagnetic signatures, and this can be used as a mapping and exploration tool in areas of limited outcrop. Last, the redox state of granitic magmas has a bearing on the types of ore deposits associated with granitic plutons (*see* **Mining Geology:** Magmatic Ores). For example, copper, gold, and molybdenum mineralization tend to be associated with magnetite-bearing granites,

whereas economic deposits of tin, tungsten, and tantalum are related to ilmenite-bearing intrusives.

Where are Granites Found?

Granitic rocks, their extrusive equivalents (rhyolites and dacites), and sedimentary and metamorphic rocks of granitic composition are easily the most abundant constituents of the upper continental crust, and they are widespread on all continents. An early observation was that large composite granitic bodies (batholiths) were commonly, but not exclusively, found as linear belts defining mountain ranges and, therefore, that their emplacement was linked in some way to orogenesis or at least to large-scale interactions between lithospheric plates (*see* **Tectonics:** Mountain Building and Orogeny). The paucity of granitic bodies in oceanic environments is testimony to the importance of continental materials, and possibly lithospheric thickness, in the genesis of these magmas, implying that granitic generation is associated with the remobilization of older crust.

However, granitic bodies are encountered in all tectonic settings, although their volumes, compositions, and the nature of the associated igneous rocks vary accordingly (Table 2). The relationship between granitic composition and tectonic setting can be used to reconstruct the tectonic evolution of ancient orogenic belts.

Ascent and Emplacement Mechanisms

Granitic bodies range in volume from less than 1 km^3, for single intrusions, to in excess of 10^6 km^3, for example in batholithic structures such as the Coast Plutonic Complex in British Columbia. They are classically depicted as circular-to-elliptical plutons on geological maps, though the shapes, field relationships, and internal structures of these bodies may be far more complex. The outcrop pattern of a granite pluton depends on the level and topography of exposure and on the three-dimensional shape of the body. To a large extent, the latter is dictated by the emplacement mechanism, which is in turn related to the depth of emplacement – and thus to the rheology of the host rocks – and to the geodynamic regime into which the pluton was intruded. Emplacement occurs at a neutral buoyancy level or mechanical discontinuity, such as the brittle–ductile transition, whereupon magma flow switches from dominantly vertical to dominantly horizontal.

The way in which large plutonic masses are emplaced in the rigid upper crust has long been contentious. The so-called 'room problem' has even formed the basis of arguments that granite has a non-magmatic origin. However, our concepts of how large volumes of granitic magma are extracted from the source region, transported tens of kilometres through the crust, and intruded at higher structural levels have recently been revolutionized in two ways. First, the traditional view that partial melting in the deep crust proceeds until a rheological threshold is reached, whereupon the source region slowly upwells *en masse* as a gravity-driven hot-Stokes diapir has been largely supplanted by a more dynamic scenario, in which the entire magma segregation, extraction, and transport process is controlled by deviatoric stress attending tectonic activity, perhaps augmented by compaction at high melt percentages. The exact way in which this occurs is probably specific to the source rock type and the nature and rate of deformation, but valuable insight is provided by deep-crustal migmatite terranes. These areas are characterized by deformation-enhanced segregation of partial melts from their residues and subsequent migration along planar anisotropies (foliations or lithological layering) to dilational structural sites, such as boudin-necks or fold axial surfaces (Figure 4A). An additional driving force is the volume increase associated with some partial-melting reactions, which can increase permeability by promoting microfracturing. In areas of local melt accumulation, high magma pressures may develop in response to compressional stress and/or fluctuation in magma supply. Subsequent magma (and thus heat) transfer to higher crustal levels occurs through interconnected networks of narrow structurally controlled channels or active shear zones during transpression, and as a result partial melts can be efficiently expelled from the source region. Excellent examples of these relationships occur in the Arunta Inlier of central Australia and in the northern Appalachians of eastern North America (*see* **North America:** Northern Appalachians), where plutons are localized in regional shear-zone systems and rooted in migmatites.

In addition to its pervasive flow in distributed pathways, granitic magma may ascend from source regions in a more focused fashion by exploiting pre-existing faults, or as self-propagating dykes, essentially driven by buoyancy. Thermal models have established that such magmatic conduits penetrating cold crust must maintain a critical thickness (typically 3–12 m) and magma-flow rate to avoid freezing and so are most viable where the generation and extraction of melt from the source is rapid. Magmatic diapirism operating over short length scales may be of limited importance in the ductile deep crust, where thermal and rheological contrasts with the host rocks are minimal, and where the viscosity of granitic

Table 2 Empirical relationships between the common granitic rock types, their typical compositions, and the geodynamic environments into which they are emplaced. The various granite types may be juxtaposed in plutonic belts whose tectonic setting has evolved during an orogenic cycle

Tectonic setting	*Occurrence*	*Dominant granite type*	*Composition*	*Associated igneous rocks*	*Other features*	*Examples*
Mid-ocean ridges, ophiolites	Dykelets and net-vein complexes to small plutons	Tonalite	Metaluminous; calcic with very low K, Rb	Pillow basalt, sheeted dykes, gabbro	May be associated with amphibolites in shear zones	Western Ophiolite Belt, Albania; Lizard Complex, south-west England; Oman Ophiolite
Oceanic arc	Strongly zoned plutons	Tonalite, granodiorite	Metaluminous; calcic to calc-alkaline	Gabbro, quartz diorite, diorite, monzodiorite	Low LILE contents, isotopically primitive Cu–Au deposits	New Britain; Papua New Guinea; Aleutian Islands; Alaska
Continental arc	Vast linear batholiths elongate parallel to the trench; zoned plutons are common	Tonalite, granodiorite, granite	Mostly metaluminous to weakly peraluminous; vary from calcic, calc-alkaline to alkali-calcic with time and distance from the trench	Gabbro, quartz diorite; andesite to dacite volcanics	High Na/K, Sr/Y; primitive to mixed isotopic signatures; Cu–Mo deposits	Peninsula Ranges Batholith, USA; Sierra Nevada Batholith, USA; Coastal Batholith, Peru; Patagonian Batholith, Chile
Syn-collisional	Batholiths orientated parallel to the structural grain, sheets in shear zones	Granodiorite, granite	Metaluminous to strongly peraluminous; calc-alkaline and alkali-calcic	Gabbro, diorite (uncommon)	Moderate Na/K, mixed to crust-like isotopic signatures; high-T, low-P metamorphism; Sn–W deposits	Lachlan Fold Belt, Australia; Massif Central, France; Appalachians, USA; Himalayas
Post-collisional	Individual plutons or groups of plutons oblique to structural trends	Granodiorite, granite	Metaluminous to strongly peraluminous; alkali-calcic	Appinite, shoshonite, lamprophyre	Low Na/K; mixed to crust-like isotopic signatures	Caledonian Fold Belt, UK; western Lachlan Fold Belt; Hercynian Belt, western Europe.
Continental rift or intraplate	Discrete plutons, ring complexes	Granite	Metaluminous to peralkaline; alkalic	Gabbro, basalt, silicic volcanics	Isotopically primitive highly differentiated, high Nb, Zr, Zn, Ga, Nb, Th, U, REE deposits	East African Rift; Eastern Red Sea Margin; Corsica; Tertiary Igneous Province, UK

LILE, large-ion lithophile elements.
REE, rare earth elements.

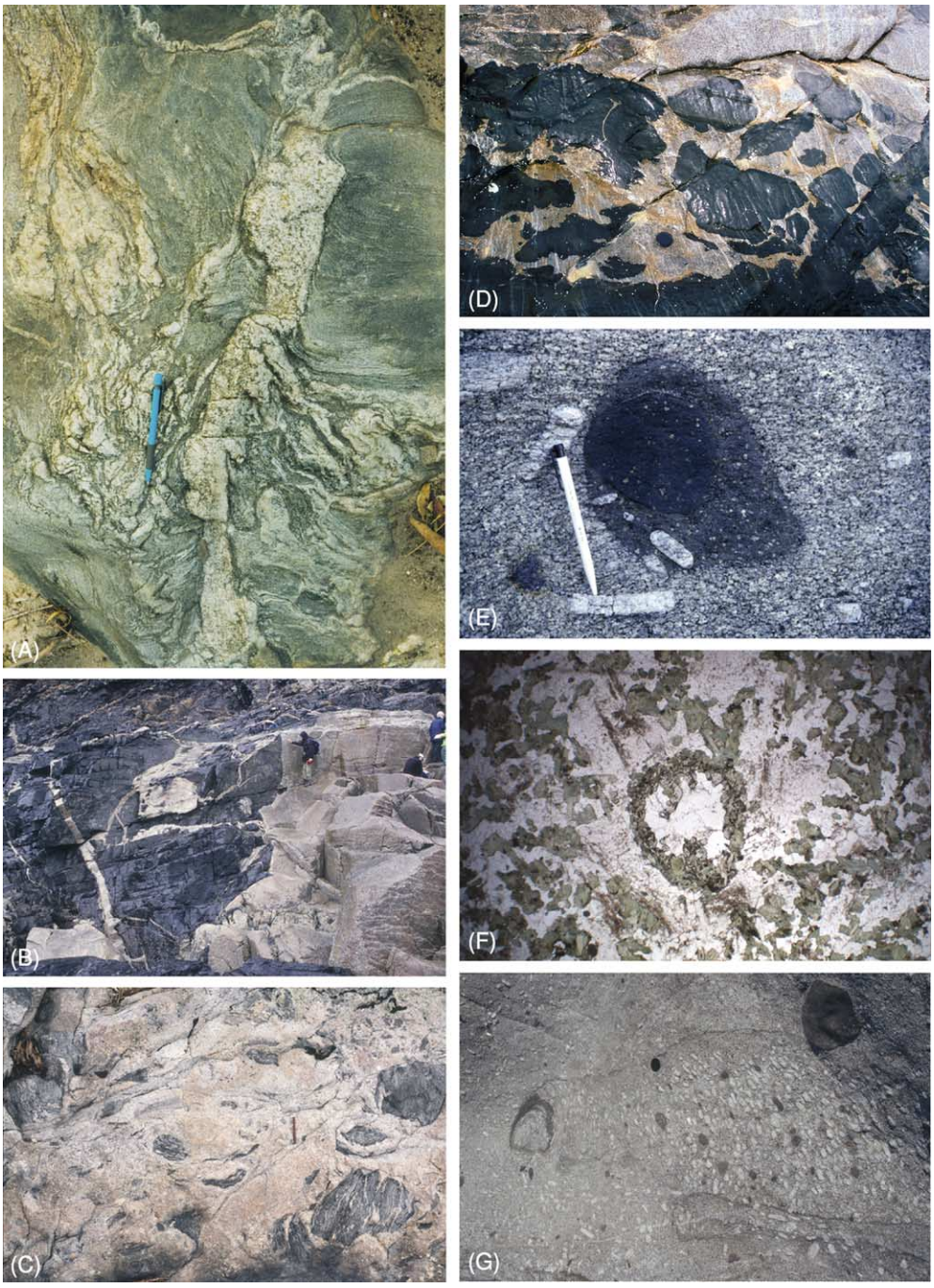


Figure 4 Granitic rocks in the field. (A) Migmatite outcrop from south-eastern Australia, showing the deformation-enhanced segregation and accumulation of locally derived partial melts into a dilatant shear zone. In this case, the leucogranite in the shear zone is compositionally equivalent to the kilometre-scale leucogranitic plutons of the area. (B) Strongly peraluminous leucogranite invading and stoping its metasedimentary roof rocks along bedding planes and cracks (Cornwall, UK). (C) Enclaves of gneiss and migmatite in a deep-crustal (*ca.* 20 km) granitic pluton in south-eastern Australia that has been retained close to the generative region. The enclaves correspond to less fertile horizons of the granite's source rocks. (D) Mingling between a syn-plutonic basaltic dyke and a metaluminous tonalite, Moruya Batholith, eastern Lachlan Fold Belt (Australia). (E) Composite enclave, in which a mafic igneous enclave (basaltic andesite) is enclosed by a microgranular enclave. Note the physical incorporation of potassium-feldspar phenocrysts from the enclosing granite (Elba Island, Italy). (F) Thin section of a microgranular enclave, showing a large quartz grain rimmed by pyroxene, which is overgrown in turn by hornblende (3 mm across). The quartz grain is interpreted as a xenocryst incorporated from the host granite, which reacted with, and was partially resorbed by, the hotter enclave magma. (G) Unusual vortex-like structure in granitic rock, which is thought to represent a cross-section through a magma flow conduit that has subsequently been 'infilled' by crystal-rich magma during crystallization. Features such as this hint at the dynamic nature of upper-crustal magma chambers and demonstrate how considerable compositional heterogeneity can be caused by simple magmatic flow processes in granitic crystal mushes (Elba Island, Italy).

magmas is elevated by the carriage of unmelted source material (restite).

Furthermore, it is becoming increasingly apparent from geophysical imaging (seismic, gravity), analysis of magnetic directional fabrics, and detailed field mapping that the three-dimensional geometries of many granitic plutons are tabular or funnel-shaped, with inclined feeder zones, and that plutons have an internally sheeted structure on a metre to kilometre scale. The implication is that, rather than being intruded in a single episode, granitic plutons (and, on a larger scale, batholiths) were constructed sequentially and *in situ* by the coalescence of multiple sill-like magma pulses. The most compelling demonstrations of this are in the Himalayas and Greenland Caledonides (*see* **Europe:** Scandinavian Caledonides (with Greenland)), where great topographical relief and superb exposure reveal that leucogranitic bodies have laccolithic geometry, where the constituent lenticular intrusions are separated by thin screens of country rock. The same situation applies in some magmatic arcs. Gravity surveys show that the vast Coastal Batholith of Peru is floored at surprisingly shallow depths (less than 7 km) and comprises a collage of numerous sheet-like plutons.

Tabular plutons in the upper crust are fed by an underlying magma conduit, and, as highlighted above, this can occur in several ways. Networks of feeder dykes are clearly exposed beneath the sheeted leucogranites of the Himalayas. The importance of felsic-magma transport by dyking in the brittle uppermost crust receives further support from the geometry of sub-volcanic ring complexes, in which magma emplacement occurs as 'ring dykes' in conical fractures. Some of the best examples of these are the alkaline ring complexes of the British Tertiary Igneous Province.

One important consequence of incremental and dynamic models of pluton assembly is that the 'room' problem is largely circumvented. The incoming granitic magmas are progressively accommodated by a combination of tectonism – involving movement on pre-existing tensional faults and fractures and ponding of magma in dilatant jogs or pull-apart structures – and roof lifting during pluton inflation or ballooning, with subsidence of the floor of the magma chamber, possibly in response to magma extraction at depth. The proximity of many plutons to major fault systems and their parallelism to regional strain patterns, as is especially evident in the Circum-Pacific subduction-related batholiths, support the notion of structurally controlled emplacement. However, it is worth noting that plutons emplaced late in the orogenic history overprint deformational fabrics and are randomly aligned relative to the structural grain. There is sometimes evidence for the localized stoping and assimilation of the roof rocks (Figure 4B), though this is probably a late-stage phenomenon rather than a major ascent mechanism.

Enclaves

Pieces of rock enclosed by granitic bodies that exhibit a colour or textural contrast with the surrounding granite are termed enclaves. Although comprising a small volumetric portion of most plutons, enclaves are eye-catching and exhibit a bewildering array of colours, shapes, sizes, textures, and compositions. Enclaves are a source of information about the sorts of processes that may have operated during the evolution of granitic magmas, and so their interpretation is an important aspect of granitic studies.

The main enclave types encountered in granitic rocks are listed in Table 3. Some are xenoliths spalled from the country rock, but the origin and significance of other enclave varieties remain debated. High-grade metamorphic enclaves have been entrained from depth, though whether they are restite from the granitic source or simply mid-crustal xenoliths is difficult to determine (Figure 4C). A restitic origin seems reasonable for those enclaves with melt-depleted compositions, such as the cordierite–spinel lumps found in some peraluminous granites. Mafic igneous enclaves have basaltic to dioritic compositions and are derived from the injection and fountaining of syn-plutonic mafic magma into the granitic pluton whilst it was partially crystalline (Figure 4D). As a result of the thermal and rheological contrasts between the hot fluid basaltic magma and the cool viscous granite, the mafic enclave commonly exhibits quenched margins and complex involute shapes resulting from contraction.

In appearance, mafic igneous enclaves are transitional to microgranular enclaves. These may also have a finer-grained margin, though, rather than chilling, in some instances this is related to concentration of biotite resulting from a reaction between the iron- and magnesium-rich enclave and the potassium- and water-rich granite. Microgranular enclaves have unusual igneous-looking textures, and their mineralogy and composition commonly, though not invariably, correlate with those of the host. They are variously considered to represent an intermingled globule of relatively mafic magma that has undergone hybridization with the host granite, a reincorporated chilled margin derived from a less-evolved facies of the host, or partially melted restite entrained from the source. Each of these origins may be correct in specific instances. A mingling origin seems apposite where the enclave exhibits features that confirm its incorporation in the molten state, such as magmatic

Table 3 Summary of the various enclave types that are encountered in granitic rocks. Note that, for deep-seated granitic bodies enclosed by their migmatitic source rocks, country-rock enclaves may correspond to fragments of the protolith or to relatively refractory source horizons

Enclave type	*Occurrence*	*Morphology*	*Description*	*Possible origin*
Country rock	All granites; may be concentrated towards the periphery	Tabular, angular	Matches exposed country rock; may have a reaction rim (especially calc-silicates) or be in the arrested stages of assimilation	Accidental entrainment
High-grade metamorphic	Most common in strongly peraluminous granites; virtually absent from peralkaline granites	Tabular, lenticular	Higher metamorphic grade and typically more complexly deformed than host rocks; micaceous objects are common	Mid-crustal xenolith; restite
Mafic igneous	All granites; occur in swarms or associated with syn-plutonic mafic dykes	Variable – globular, elliptical, lenticular, pillow-like, or amoeboid; cuspate or crenulate margins	Fine grained, commonly with chilled margins; may show evidence of magmatic deformation, usually stretching or contortion	Intermingled globules of relatively mafic magma
Microgranular	All granites, though rare in peralkaline; dispersed more or less evenly throughout the pluton	Rounded, elliptical to ovoid, rarely amoeboid	Generally darker and finer-grained than the host, though the mineralogy may be the same; typically biotite-rich; may entrain crystals of the host granite	Hybridized mafic globule; fragment of less evolved granite; restite
Coarse igneous	All granites, but uncommon	Rounded or angular	Fragments of relatively coarse-grained plutonic rock; may be mafic or felsic	Cumulate; xenolith of unexposed pluton

flow textures or the engulfing of phenocrysts from the surrounding granite (Figure 4E). The occurrence of disequilibrium or hybridization features also attests to an intermingled origin (Figure 4F).

Mineralogy and Textures of Granitic Rocks

Besides quartz and feldspar, a variety of minor and accessory minerals may occur in granitic rocks, as dictated by the bulk composition of the magma (including its oxidation state and water content) and the pressure and temperature at which igneous crystallization took place (Table 1). An additional complication is that, instead of precipitating directly from the granitic magma, some of the minerals may be of extraneous origin (xenocrysts derived by assimilation of the country rocks) or be residual or restitic crystals entrained from the crustal source of the granitic magma. In some cases this may be resolved by textural examination, but this is by no means diagnostic; plutonic rocks cool slowly, and thus there is ample opportunity for conversion of early (restitic or magmatic) phases to minerals stable at the pressure and temperature conditions of emplacement. Primary high-temperature textures can be obscured, or even obliterated, by these reactions.

Strongly peraluminous granites contain an aluminium-saturating phase, sometimes accompanied by orthopyroxene or pseudomorphs after this mineral. Orthopyroxene is especially common in deep-seated plutons of granulite terranes, such as the tonalites of the Hidaka Metamorpic Belt, Japan. Of the peraluminous minerals, cordierite is texturally the most complex (Figure 5). It can occur as (a) euhedral crystals, (b) large poikilitic grains intergrown with quartz and alkali feldspar, (c) anhedral masses sieved with orientated inclusions of biotite, fibrous sillimanite, and/or spinel, commonly enveloped by euhedral inclusion-free cordierite, and (d) overgrowths on garnet. These different textural varieties may even be encountered in the same sample, and they highlight the type of information available from understanding mineral textures in granitic rocks. The first and possibly the second of these textures are magmatic, whereas the third is either restitic or xenocrystic. The cordierite rims on garnet reflect decompression, which could be related to magma ascent. The garnets themselves (mostly almandine-rich) are also potentially either magmatic or restitic/xenocrystic in origin, though the small manganese-rich garnets found in leucogranites are probably magmatic. In contrast to the felty aggregates of sillimanite, andalusite may form euhedral melt-precipitated crystals in peraluminous

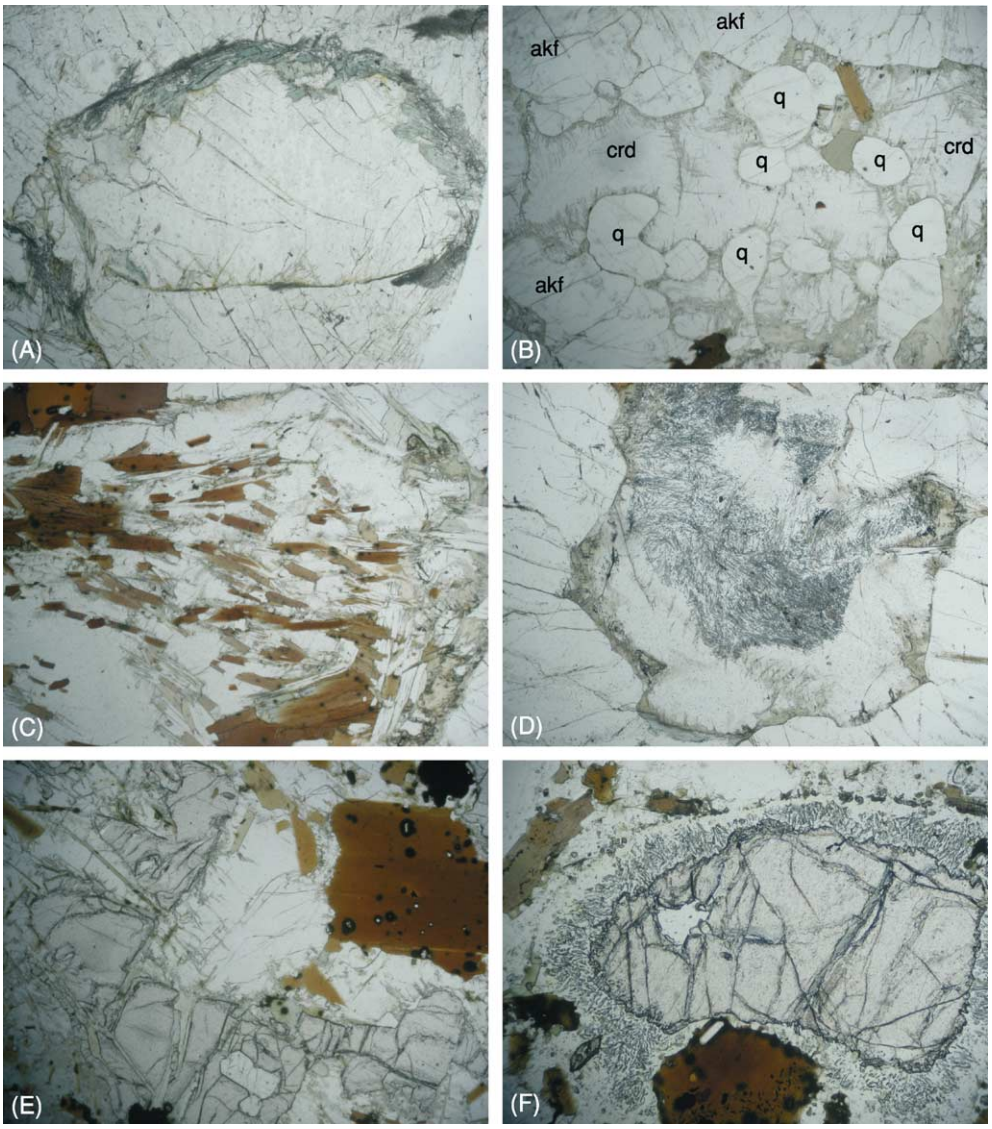


Figure 5 The different cordierite morphologies encountered in strongly peraluminous granitic rocks of the Lachlan Fold Belt, south-eastern Australia. All photographs were taken in plane-polarized light and are 3.5 mm across. (A) Well-shaped magmatic cordierite crystals with a thin rim of greenish biotite and fine andalusite. (B) Irregularly shaped cordierite grain intergrown with quartz (q) and alkali feldspar (akf), crd = cordierite. Note the fine micaceous alteration along cracks and cleavages. (C) Large cordierite mass enclosing numerous biotite inclusions, which preserve an earlier directional fabric. The granite is undeformed, so the cordierite is either restitic or (more likely) xenocrystic. (D) Fibrous aggregates of sillimanite and fine spinel grains in the core region of a cordierite grain. The clear inclusion-free rim is euhedral and melt-precipitated. (E) Reaction of garnet (high relief, brownish) to cordierite (centre, clear). (F) Garnet with a reaction corona of cordierite and wormy intergrowths of orthopyroxene and plagioclase. The small circular yellow patches inside the clear cordierite are haloes around minute zircon inclusions and reflect radiation damage caused by the decay of uranium and thorium in these inclusions.

granites. These are most commonly observed in highly differentiated plutons (co-existing with topaz), where the solidus temperature has been depressed into the andalusite stability field by high concentrations of fluorine and/or boron. Magmatic andalusite of this type occurs in some granites of the South Mountain Batholith, Nova Scotia.

Muscovite in strongly peraluminous granites also has two potential modes of occurrence. Large euhedral plates, sometimes with oscillatory zoning and 'dovetail' intergrowths with biotite, are thought to be magmatically precipitated. However, poorly shaped muscovite flakes enclosing and/or replacing other aluminous phases are probably of subsolidus origin, especially where the rock has been deformed or hydrothermally altered. Secondary muscovite is a common replacement product of feldspar and thus may occur in altered metaluminous rocks.

Metaluminous granites have a simpler mineralogy. The diagnostic minerals are hornblende and clinopyroxene (augite), which may be accompanied by orthopyroxene, accessory titanite and/or allanite. Generally, clinopyroxene is enveloped and partially replaced by calcic amphibole, and, as this process releases silica, amphiboles formed by this reaction may be perforated with quartz. Epidote is a common alteration product, though euhedral prisms in metaluminous granites emplaced at moderate pressures are magmatically precipitated.

The mineralogy and texture of peralkaline granites are most distinctive, and it is clear that these rocks are produced by the crystallization of high-temperature completely liquid magmas at shallow crustal levels. Many are brick-red and have unusually turbid feldspars containing minute haematite inclusions, reflecting pervasive hydrothermal alteration. This typically occurs by circulation of meteoric waters contained in near-surface aquifers. Alkali feldspars show irregular perthitic features ('patch' perthite) and exhibit striking granophyric intergrowth with quartz, particularly near miarolitic cavities, which are formed by exsolution of the magmatic vapour phase at low confining pressures. The plagioclase is sodic, and so most peralkaline plutons are true granites. An alkali amphibole is usually present, sometimes accompanied by sodium-pyroxene and rarely fayalite. Biotite is iron-rich and interstitial to other phases, consistent with late crystallization and moderate-to-low magmatic water contents (2% or less).

Petrogenic Studies

Silicic magmas can be produced in two fundamental end-member ways: first, by differentiation of high-temperature mantle-derived mafic magmas, and, second, by partial melting of pre-existing crustal rocks. The end-point of the first of these and the incipient stages of the second both produce hydrous melts that cluster around the thermal minimum/eutectic in the quartz–albite–orthoclase–water system.

However, the geochemical diversity of granitic rocks is controlled by a complex interplay between the composition of the source rocks – which may be a function of the tectonic setting – the conditions of partial melting, and the operation of processes such as magma mixing, restite unmixing, assimilation, fractional crystallization, and other crystal–liquid processes (e.g. convective crystallization, filter-pressing), which may operate in tandem or sequentially from the source region to the emplacement site (Figure 6). Deconvolving these effects is a major challenge facing granite petrologists (e.g. Figure 7).

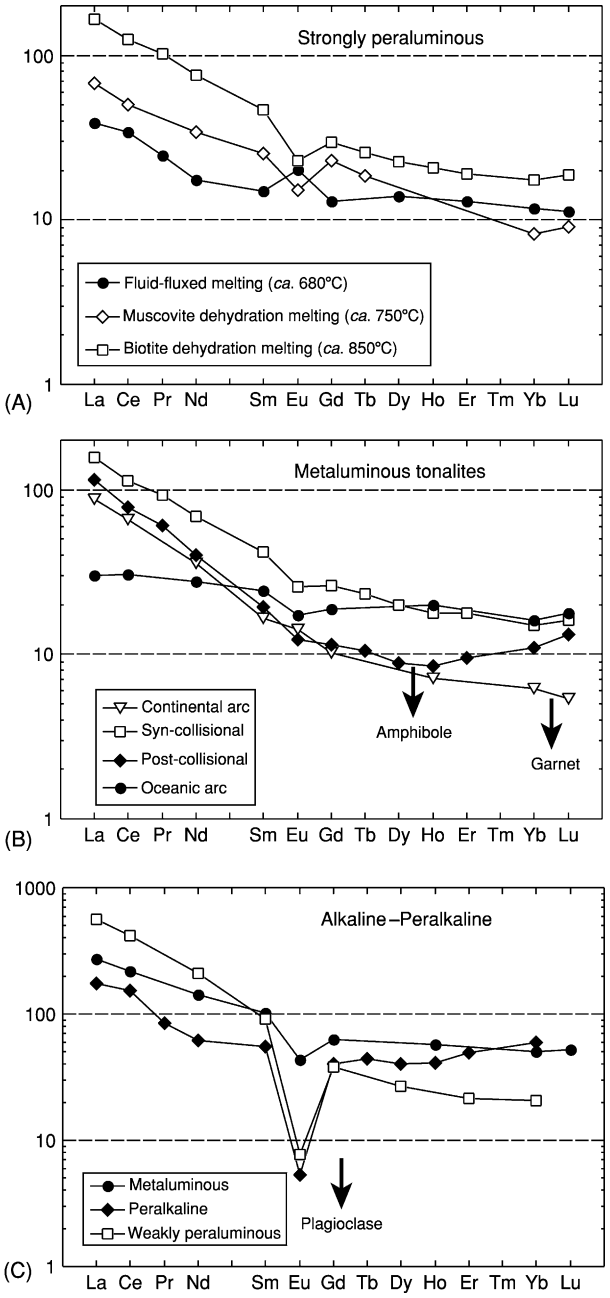


Figure 6 Chondrite-normalized rare earth element (REE) plots comparing the different granite compositions produced by melting (A) metasedimentary sources and (B) meta-igneous sources, under different conditions, with (C) alkaline to peralkaline plutons. (A) The increasing REE content with temperature in strongly peraluminous granites reflects greater dissolution of REE-rich accessory phases during melting. (B, C) Arrows indicate compositional control by garnet (heavy REE), amphibole (middle REE), and plagioclase (europium), either by fractional crystallization or by retention of these phases in the source region during partial melting. (B) Specifically, the continental-arc tonalite is thought to have been derived by high-pressure melting of garnet amphibolite at the base of the crust or in the mantle, whereas (C) the alkali granites have experienced plagioclase fractionation at shallow depths.

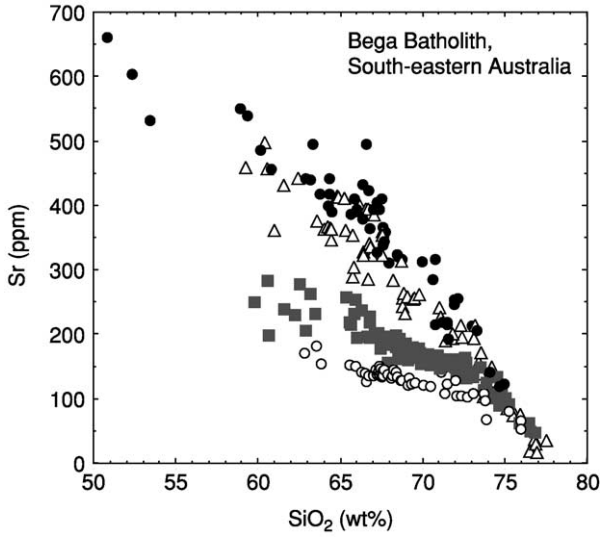


Figure 7 Plot of strontium versus silica contents for metaluminous granites of the Bega batholith, south-eastern Australia. The broadly linear variation defined by the different groups of granites has been variously ascribed to restite unmixing, two-component magma mixing, and fractional crystallization. In this context, the compositional convergence at high silica could reflect evolution towards a 'minimum melt' composition or that a similar felsic end-member magma was involved in mixing. Examples such as this highlight the potential problems of interpreting the whole-rock geochemistry of plutonic rocks in general. The various symbols represent granite suites. These are groups of rocks with similar textural and geochemical features and are thought to be related through a common source or process.

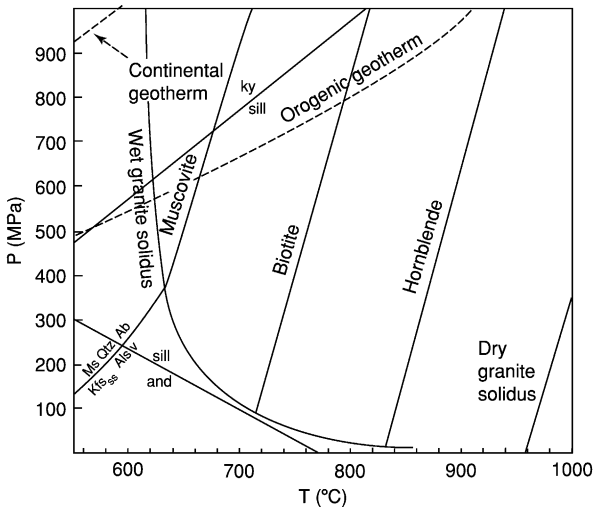


Figure 8 Pressure–temperature diagram showing the beginning of melting curves for granitic compositions and the locations of the muscovite, biotite, and hornblende dehydration-melting reactions. The important role of water in crustal melting is borne out the large temperature difference between the water-saturated ('wet') and water-absent ('dry') granite solidi. Note that melting reactions in crustal P–T space (except hornblende dehydration) are intersected only by unusually hot geotherms, such as that of young orogenic crust, and not by the geotherm of stable continental areas. Abbreviations: ky, kyanite; sill, sillimanite; and, andalusite; ms, muscovite; qz, quartz; ab, albite; Kfs_{ss}, solid solution of potassium feldspar; Als, aluminosilicate. (Modified from Johannes W and Holtz F (1996) *Petrogenesis and Experimental Petrology of Granitic Rocks*. Berlin: Springer-Verlag.)

Furthermore, there has been increasing recognition that, rather than representing liquid compositions, many granites were emplaced as mushes of precipitated or entrained restitic crystals, or even represent crystal cumulates in which the proportion of trapped melt is below 30%.

The origin of granitic rocks is clearest in oceanic areas remote from continental materials. Here, silicic magmas are generated either by extreme fractional crystallization or by the remelting of mantle-derived materials. Granitic rocks are exposed as the small-volume plagiogranites of mid-ocean ridges (now part of ophiolite complexes) and the alkali granites of some oceanic plateaux. The quartz diorite to tonalite plutons of intraoceanic arcs have a more complex genesis, though primitive isotopic ratios suggest that the source materials are essentially the same.

In contrast, the potential protoliths of the voluminous granitic plutons of continental areas include a diverse range of common crustal rock types of various ages, which may melt at a variety of pressures and temperatures. This is borne out by the spectrum of strontium, neodymium, and oxygen isotopic compositions shown by continental granitic rocks, which range from depleted mantle-like to highly evolved and similar to the compositions in the average continental crust. To some extent the source-rock composition is governed by tectonic setting. For example, the immature volcanogenic greywackes of convergent plate margins differ from the more weathered flysch of passive margins, and the compositions of the derivative granitic magma differ accordingly. For this reason, chemical features of silicic magmas may be diagnostic of the tectonic setting of their sources, rather than that under which magmatism occurred.

Water is a crucial ingredient for partial melting of pre-existing crust (Figure 8). It may be externally derived, for example by dewatering of underthrust sediments or crystallization of hydrous magmas, or it may be liberated during melting by decomposition of hydrous minerals ('dehydration' melting). The role of fluids in the lowermost crust, especially above subduction zones, requires further investigation. However, melting promoted by fluid influx into mid-crustal amphibolite-facies terranes tends to generate near-solidus peraluminous leucogranitic magmas that have limited vertical mobility. For this reason, most granitic magmas are thought to have been generated by higher temperature dehydration-melting reactions involving micas and amphiboles. These reactions have

negative d*P*/d*T* slopes, permitting melts generated in this fashion to ascend to shallow crustal levels before crystallizing at the water-saturated solidus (see Figure 8).

During dehydration melting, the nature and amount of the resultant granitic magma are controlled by the specific hydrous mineral participating in melting and its proportion in the rock. At the lowest temperatures (less than 750°C), muscovite dehydration melting in metapelites yields strongly peraluminous felsic liquids; this mechanism is thought to be responsible for the formation of the Himalayan leucogranites and the leucogranites of the Appalachians, Cornubian batholith (south-west England), and Iberian massif (Spain). The evolved isotopic ratios of these rocks (i.e. high $^{87}Sr/^{86}Sr$ and low $^{143}Nd/^{144}Nd$ at the time of formation) are consistent with derivation from metasedimentary protoliths. Modelling demonstrates that temperatures sufficient for partial melting can be achieved by internal radioactive heating and thermal relaxation associated with crustal thickening. Melting may be triggered or enhanced by decompression during orogenic collapse and exhumation, where additional heat is supplied by frictional heating along active fault systems. However, the melt fraction is limited by the modal abundance of muscovite in the rock and is thus generally less than 20%.

The large relatively biotite-rich granitic bodies that dominate most magmatic belts require higher-temperature (more than 850°C) biotite and hornblende dehydration-melting reactions from protoliths such as metagreywackes, intermediate igneous rocks (i.e. meta-andesites, metatonalites), or even amphibolites. Modelling and experimental studies predict that these reactions are capable of generating large proportions (up to 50 vol.%) of water-undersaturated granitic melt from protoliths with the appropriate (fertile) composition.

In nature the situation is more complex. In the absence of fluids, the elevated temperatures required for biotite and hornblende dehydration melting can be achieved only by significant heat input into the crust from the mantle, for example by the direct incursion of hot (more than 1000°C) basaltic liquids. Under these circumstances, it seems inevitable that the basaltic liquids will also physically interact with the crustal melts and therefore make some material contribution to the resultant granitic magmas. Several lines of evidence support this. First, the large-scale mingling and in some cases mixing of granitic and mafic (basaltic) magmas is commonly observed in the field, both in upper-crustal magma chambers and composite dykes and (more rarely) at deep crustal levels. The widespread occurrence of mafic or microgranular enclaves and mafic mineral clots derived therefrom demonstrates that dispersal of the mafic ingredient is pervasive throughout such plutons. Second, experimental melting studies of metasedimentary protoliths have consistently failed to reproduce the composition of the cordierite-rich granites found in thickened convergent orogens. However, these compositions are satisfactorily reproduced when a basaltic component is added to the experimental mixture. Third, some granitic rocks, especially metaluminous varieties, share compositional features with spatially associated mantle-derived magmas, strongly suggesting a genetic link. Fourth, the isotope ratios of many granitic bodies are systematically shifted to more primitive values than their inferred crustal sources. In some areas, granitic rocks form covariant plausibly mixing arrays on strontium–neodymium–oxygen isotope diagrams that converge towards depleted mantle-like values (Figure 9).

The way in which the mantle components are incorporated is not well understood. Simple mixing models are challenged by the physical and thermal difficulties of efficiently blending hot highly fluid basaltic magma with cooler viscous granitic magma,

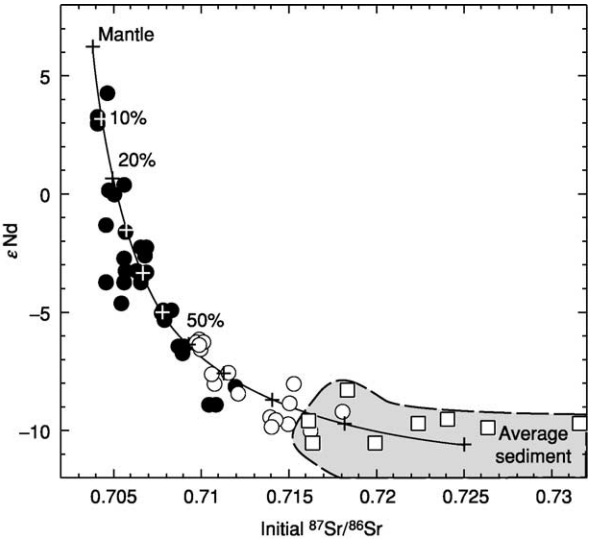


Figure 9 Diagram of epsilon neodymium versus $^{87}Sr/^{86}Sr$ for granitic rocks of the Lachlan Fold Belt (south-eastern Australia) calculated at their time of formation (the epsilon notation signifies deviation from a reference chondritic reservoir in parts per 10 000). The field defined by the country rocks – Ordovician turbidites (open squares) – is also shaded. The single array defined by the metaluminous (hornblende granite; filled octagons) and strongly peraluminous (cordierite granite; open octagons) granites has been taken to imply that the two groups formed by different degrees of mixing between mantle-derived magma and a continental crust-like end-member, approximated by the host sediments. Mixing proportions are indicated as crosses on the calculated mixing curve. However, the mismatch between the measured bulk-rock composition of the granite and that predicted by the isotopic mixing proportions suggests the operation of other processes, which are difficult to resolve with bulk-rock data.

although the viscosity contrast may be less than originally thought (10^4–10^6 Pa s for hydrous crystal-poor granite versus 10^2 Pa s for a typical basalt). The thermal constraints are satisfied by models in which granitic magmas are generated by the wholesale assimilation of crustal materials by thick convecting basaltic sills in thermally perturbed lower crust, or by the combination of felsic liquids derived from the crystallization of such sills with partial melts of the overlying crust. Compelling field evidence for such scenarios can be directly observed in deep-crustal exposures upthrust in continental-collisional zones, such as the Ivrea Zone of the Italian Alps and in the Norwegian Caledonides.

A promising new way of unravelling the various source components of granitic magmas involves decoding the isotopic information preserved in the fine-scale growth zoning of certain minerals (Figure 10). This is done by *in situ* microanalysis, using laser-ablation sampling or secondary ion mass spectrometry, both of which are capable of measuring precise isotope ratios and trace-element contents at sub-ppm levels in very small areas of the crystal (20 μm or less). Zircon is an ideal mineral for this approach since the growth zones can be directly dated by uranium–lead isotopes. Zircon is also robust and may survive as a refractory residue in granitic magmas (*see* **Minerals:** Zircons), thereby preserving age and compositional information about the granite's source rocks and the otherwise inaccessible deep crust. Hafnium is a particularly abundant trace element in zircon, and fluctuations in hafnium isotope composition within zircon crystals record the progress of processes such as magma mixing and crustal assimilation. In this way, the crystallization history of silicic magmas can be reconstructed as they ascend from their deep crustal source to their final emplacement site.

The Time-Scales of Granitic Magmatism

The time-scales of geological processes provide important clues about the formation of granitic magmas and the various factors that influence their compositional evolution. For silicic magmas, this area of study is still in its relative infancy. However, there is growing evidence that the segregation, ascent, and emplacement of granitic magmas occur on time-scales of hundreds or thousands, rather than millions, of years (Figure 11). When coupled with the crystallization rates, this places constraints on the sites and nature of magmatic differentiation and on partial-melting processes in the source region.

Information on the time-scales of melt segregation and extraction in the granite's source region comes from the accessory minerals zircon and monazite,

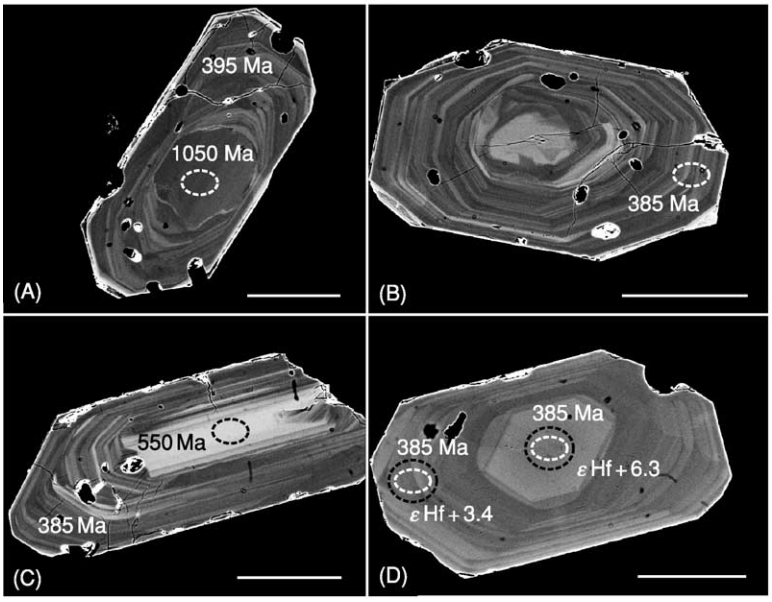


Figure 10 Cross-sections of zircon crystals imaged by back-scattered electron microscopy (scale bar is 100 μm) to show internal growth structures. Circles in (A)–(C) represent ion probe pits, and the circles in (D) represent laser ablation pits. (A) Zircon with a 390 Ma igneous crystallization age showing corroded, partially resorbed older core dated at 1050 Ma. (B) Small, irregularly-shaped zircon nucleus (too small to analyse) enclosed by strongly zoned magmatic zircon. (C) Elongate 550 Ma core surrounded by 385 Ma igneous zircon. Note the cracks radiating from the older zircon core. (D) Magmatic zircon analysed by laser ablation, showing strong core to rim zoning in hafnium isotopic composition, expressed in the standard epsilon notation (see caption to Figure 9). This reflects the operation of an open system process, such as crustal assimilation, during crystallization.

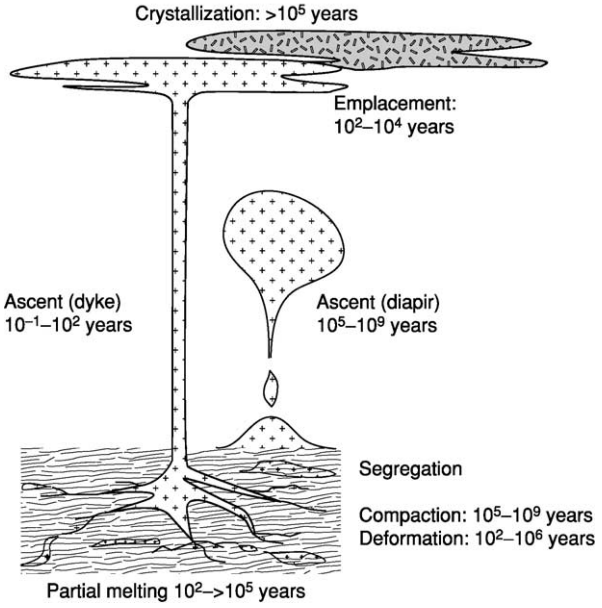


Figure 11 Summary of the time-scales estimated for the various stages in the formation of a granitic pluton.

which (together with some lithophile elements) dominate the trace-element budget of peraluminous granites. Many peraluminous leucogranites contain lower concentrations of zirconium and rare earth elements than are predicted by equations describing the solubility of the host minerals (i.e. zircon and monazite) at the temperature of partial melting. Assuming that the granitic magmas are unmodified by fractional crystallization, this is attributed to the rate of melt segregation exceeding that at which equilibrium can be maintained with dissolving zircon and/or monazite crystals in the source. As zircon dissolution kinetics can be numerically modelled, it is possible to calculate the residence time of the partial melt in the source region from the zirconium content of the melt or the degree of zirconium undersaturation. Remarkably, such calculations reveal that the Himalayan leucogranites, some of which exceed 3000 km^3, were extracted from their sources in less than 150 years. It remains to be established whether these extraction rates are applicable to granitic magmas formed from different source rocks in other tectonic settings.

Evidence concerning the rates of ascent and emplacement of granitic magmas stems from the commonly observed transport of granitic magma in dykes. As indicated above, to avoid freezing, dykes must maintain a critical width and flow rate, both of which are determined by the temperature and viscosity of the magma. In view of this, fast melt transport rates are predicted, leading to rapid filling of large high-level granitic plutons. For example, in the Himalayas, leucogranitic magmas are thought to have been transported vertically through about 10 km of crust. Using an experimentally determined melt viscosity, a critical dyke width of 3 m permits an average melt ascent velocity of 9.5 cm s^{-1}, a factor of up to 10^8 faster than magma transport modelled by diapiric ascent. At these magma delivery rates the largest laccoliths in the Himalayas could have been emplaced in about 10 years, though this is an extreme case. Rapid rates of melt transport and emplacement are also inferred for granitic plutons in subduction-related settings. For the large Cordillera Blanca batholith of the Andes, a melt ascent rate of 1 cm s^{-1} and a dyke width of 6 m lead to an emplacement time of 350 years. Peak pluton-filling times of 1000 years are modelled for the plutons of the Sierra Nevada batholith.

The time-scales of crystallization of felsic magmas are constrained by the study of radiogenic isotopes in recent high-silica volcanic rocks. These are the eruptive analogues of granitic plutons, and, in some cases, such as the voluminous (*ca.* 5000 km^3) Fish Canyon Tuff (Colorado), are thought to form by the rejuvenation of a partially crystalline granitic magma chamber. For the peralkaline rhyolites of the East African Rift, precise rubidium–strontium and uranium–thorium dating reveals that the crystals and their host fractionated liquids formed about 15 000 years before eruption. Similarly, rubidium–strontium mineral dating establishes that most feldspar in the Long Valley rhyolites (California) crystallized in less than 15 000 years. In contrast, a much longer crystallization time (more than 200 000 years) for zircons in the same rocks was inferred using uranium–thorium isotope disequilibria. Zircons in pumices from the Taupo Volcanic Zone (New Zealand) also crystallized over a period of 300 000 years prior to eruption, suggesting a substantial residence time in the magma chamber.

Although it is unclear to what extent these studies can be extrapolated to plutonic rocks, they serve to highlight the fact that the rates of crystallization of silicic magmas are potentially far out-paced by the rates of magma transport and emplacement. The implication for granitic magmas is that crystal–liquid differentiation, and thus crustal assimilation, is likely to operate in high-level magma chambers rather than in the deep crust.

A final point is that, if dyke transport is a valid description of silicic magma ascent in general, the rate-determining step in the construction of granitic batholiths is probably the melt generation in the source. The sheeted structure and internal isotopic heterogeneity characteristic of many plutons suggest an episodic or pulsed, rather than continuous, transport of magma to the emplacement site. If the

magma supply is produced by crustal fusion, then this in turn is limited by the prograde heating rate. Where felsic melt is produced from mantle-derived precursors, the rate of melt production depends on the intrusion rate of these into the crust and on the timescales of differentiation. Constraining the rate of silicic magma generation by both of these processes in nature has thus far proved elusive and is likely to be a focus of future studies.

See Also

Earth: Crust. **Europe:** Scandinavian Caledonides (with Greenland). **Igneous Processes**. **Minerals:** Feldspars; Quartz; Zircons. **Mining Geology:** Magmatic Ores. **North America:** Northern Appalachians. **Tectonics:** Mountain Building and Orogeny.

Further Reading

Bouchez JL, Hutton DHW, and Stephens WE (eds.) (1997) *Granite: From Segregation of Melt to Emplacement Fabrics*. Dordrecht: Kluwer.

Brown M (1994) The generation, segregation, ascent and emplacement of granite magma: the migmatite-to-crustally-derived granite connection in thickened orogens. *Earth Science Reviews* 36: 83–130.

Frost BR, Barnes CG, Collins WJ, *et al.* (2001) A geochemical classification for granitic rocks. *Journal of Petrology* 42: 2003–2048.

Johannes W and Holtz F (1996) *Petrogenesis and Experimental Petrology of Granitic Rocks*. Berlin: Springer-Verlag.

Kemp AIS and Hawkesworth CJ (2003) Granitic perspectives on the generation and secular evolution of the continental crust. In: Rudnick R (ed.) *Treatise of Geochemistry, Volume 12. The Crust,* pp. 349–410. Oxford: Elsevier.

Patiño Douce AE (1999) What do experiments tell us about the relative contribution of crust and mantle to the origin of granite magmas? In: Castro A, Fernandez C, and Vigneresse JL (eds.) *Understanding Granites: Integrating New and Classical Techniques,* pp. 55–76. Special Publication 168. London: Geological Society.

Petford N, Cruden AR, McCaffrey KJW, and Vigneresse J-L (2002) Granite magma formation, transport and emplacement in the Earth's crust. *Nature* 408: 669–673.

Pitcher WS (1998) *The Nature and Origin of Granite,* 2nd edn. Glasgow: Blackie.

Thompson AB (1999) Some time–space relationships for crustal melting and granitic intrusion at various depths. In: Castro A, Fernandez C, and Vigneresse JL (eds.) *Understanding Granites: Integrating New and Classical Techniques,* pp. 7–25. Special Publication 168. London: Geological Society.

Kimberlite

G J H McCall, Cirencester, Gloucester, UK

Introduction

The name kimberlite is derived from the name of the town in South Africa where diamonds were first mined in quantity after their discovery there in 1870. Coincidentally, the region of Western Australia where economic deposits of diamonds were discovered in the 1970s was named the 'Kimberleys' after the same Colonial Secretary in England.

Kimberlites are rare and volumetrically insignificant rocks, which are nonetheless studied exhaustively because they are the main source of diamonds (directly or from placer deposits derived from them).

Definition

The early definitions of kimberlite were so broad that they permitted extension to a wide variety of brecciated igneous rocks found in diatremes (explosion pipes) and carrying mantle-derived xenoliths and high-pressure macrocrysts, including mafic minettes (lamprophyres), olivine lamproites, nephelinites, and melilitites. Stricter definitions applied more recently have placed greater emphasis on groundmass composition and rock chemistry.

Kimberlites are petrographically complex in that they contain not only phases crystallized from the liquid magma but also polycrystalline fragments or crystals derived from various types of fragmental xenolith. The texture is inequigranular because of the xenoliths and macrocrysts, and the latter include both xenocrysts and high-pressure phenocrysts.

There have been innumerable definitions advanced in the literature. The AGI *Glossary of Geology* offers the following.

> Kimberlite: an ultramafic igneous rock containing at least 35% olivine with one or more of the following in the groundmass: monticellite, phlogopite, carbonate, serpentine, diopside. No leucite is allowed. Two types have been distinguished:
>
> Basaltic or type 1 kimberlite: the classic diatreme-filling diamond-bearing rocks of South Africa, much more widespread than type 2 micaceous kimberlite.

Micaceous or type 2 kimberlite: phlogopite is the main component and the smaller micas are phlogopite or tetraferriphlogopite: other groundmass minerals are diopside (mantled by titanian aegirine), spinels (chromite to magnetite), apatite (Sr-rich), perovskite, carbonates and serpentine. There is no monticellite and olivine is commonly subordinate to phlogopite in abundance. These type 2 kimberlites were called orangeites by Wagner in 1928 and the name has recently been revived, though whether it is necessary is debatable. It is restricted to South Africa.

JB Dawson has emphasized that the presence of diamond in kimberlites is not essential, and indeed kimberlites that contain diamonds are in the minority.

In 1995 RH Mitchell used a definition that improved the textural and mineralogical description. He defined kimberlites as volatile-rich (dominantly carbon dioxide) potassic ultrabasic rocks, commonly exhibiting a distinctive inequigranular texture resulting from macrocrysts (or even megacrysts) set in a fine matrix and consisting of anhedral crystals of olivine, magnesian ilmenite, chromium-poor titaniam diopside, pyrope garnet (commonly subcalcic), phlogopite, enstatite, and titanium-poor chromite. Olivine macrocrysts are a characteristic constituent of all but fractionated kimberlites. The matrix contains a second generation of primary euhedral-to-subhedral olivine, which occurs together with one or more of the following primary minerals: monticellite, phlogopite, perovskite, spinel (magnesian-ulvo-/magnesiochromite-ulvo-, and magnetite solid solutions), apatite, and serpentine. Many kimberlites contain late-stage poikilitic micas belonging to the barian phlogopite–kinoshitalite series. Nickeliferous sulphides and rutile are common accessory minerals.

Early-formed olivine, phlogopite, monticellite, and apatite are commonly replaced by deuteric serpentine and carbonate.

Chemistry

Table 1 gives some typical major-element analyses of kimberlites. Rare earth elements are enriched in kimberlites relative to other ultrabasic rocks.

Types of Intrusion

Kimberlites occur as diatremes (pipes), dykes, and sills, and one ring-dyke occurrence is known in Sierra Leone. Small surface extrusions do occur, for example in Kasma, Mali (pipe surrounded by laterized tuffs 4 m thick) and Igwise, Tanzania (three small tuff rings and a lava flow). Thermal metamorphism of the country rocks may be present, as at the Premier Mine near Pretoria, South Africa, where shale and dolomite have been metamorphosed.

Diatremes

Diatremes tend to occur as clusters or scatters along elongate zones. They have small surface areas, varying from a few square metres to several hectares (some surface areas of diatremes are listed in Table 2). The original pipe at Kimberley covered only 4 ha, whereas the Mwadui occurrence in Tanzania covers 141 ha. Double, paired, or multiple intrusions are common in Yakutia (Figure 1), and these may coalesce at depth or represent quite different eruptive phases. The pipes have the form of simple cones, tapering downwards, but they vary in surface plan

Table 1 Some chemical analyses (major elements) of kimberlites. (Reproduced from Nixon PH (1980) The morphology and mineralogy of diamond pipes. In: Glover JE and Groves DI (eds.) *Kimberlites and Diamonds*, pp. 32–47. Nedlands: University of Western Australia.)

	Lesotho kimberlite	*South African kimberlite*	*Yakutian kimberlite*	*Indian kimberlite*	*Riley Co. USA*	*Ultrabasic (peridotite)*
SiO_2	33.21	35.2	31.1	36.09	24.15	40.6
TiO_2	1.97	2.32	2.03	5.10	1.50	0.05
Al_2O_3	4.45	4.4	4.9	3.81	2.03	0.85
Fe_2O_3	6.78	9.8	10.5	4.77	6.51	12.6
FeO	3.43	9.8	10.5	4.45	1.85	12.6
MnO	0.17	0.11	0.10	0.15	0 (not detected)	0.19
MgO	22.78	27.9	23.9	24.71	24.45	42.9
CaO	9.36	7.6	10.6	3.50	15.87	1.0
Na_2O	0.19	0.32	0.31	0.20	0.30	0.77
K_2O	0.79	0.98	2.1	1.14	0.30	0.04
P_2O_5	0.65	0.72	0.66	2.31	0.62	0.04
CO_2	4.58	3.3	7.1	0.72	12.04	0.04
H_2O^+	8.04	7.4	5.9	13.05	8.50	
H_2O^-	2.66					0.84
No. of samples	25	80	623	12		

Table 2 Surface areas of some larger kimberlite diatremes. (Reproduced from Dawson JB (1980) *Kimberlites and Their Xenoliths*. Berlin, Heidelberg, New York: Springer-Verlag.)

	Hectares
Southern Africa	
Premier	31.5
Kimberley	4.1
Bultfontein	9.5
Dutoitspan	12.3
AK/1 (Orapa), Botswana	112.3
Kao (Lesotho)	17.4
Zaire	
Taiala	50.0
Kambeli	26.0
M'Bo	23.5
Tanzania	
Mwadui	141.4
Russia	
Mir	13
Zarnitsa	28
Udachinaya	52

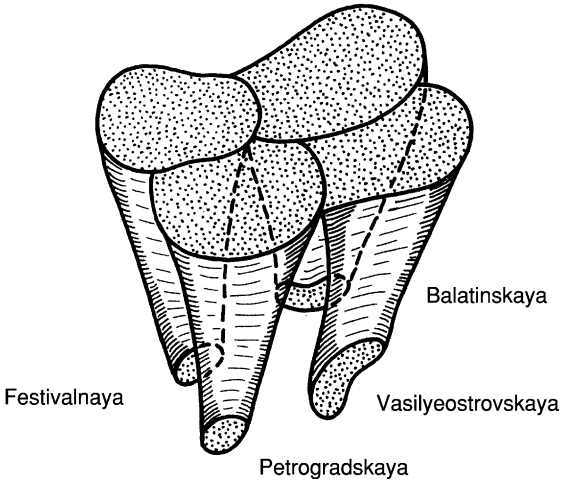


Figure 1 Structure of a group of diatremes in the Ukukinskoya Field, Yakutia (Reproduced from Mitchell RH (1995) *Kimberlites, Orangeites and Related Rocks*. New York: Plenum Press.)

(Figure 2). The generalized model of a kimberlite shown in Figure 3 takes the form of a pipe open to the surface and represented there by a tuff ring, but some pipes are blind. This model shows how different diatremes may be preserved eroded to different levels.

Dykes

Dykes occur both singly and in swarms of parallel dykes and may be multiple, representing several phases of injection of magma. They are characteristically narrow (not more than 1–2 m), but may be persistent – one over a metre wide south-west of Kimberley persists for 30 km. They may show enlargements or

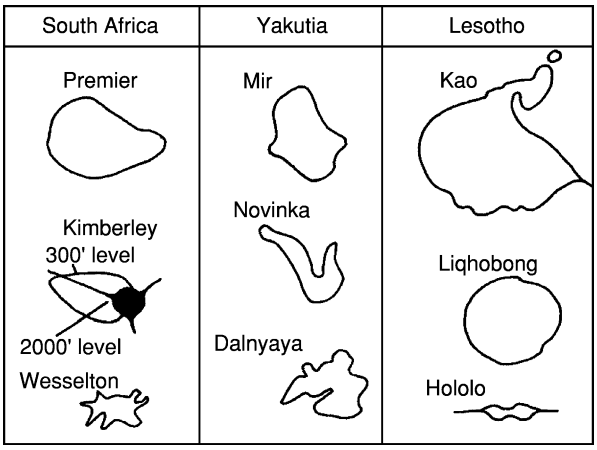


Figure 2 Some plans of kimberlite diatremes. (Reproduced from Dawson JB (1980) *Kimberlites and Their Xenoliths*. Berlin, Heidelberg, New York: Springer-Verlag.)

pods that connect with the surface, but these are not diatremes, being lenticular in section and plan: they do not contain brecciated material or fragments of overlying formations such as those typically found to have descended into a diatreme.

Sills

Sills of kimberlite are much rarer than diatremes or dykes, probably because the present erosion level must coincide closely with the injection level of these bodies, which is restricted. However, they are known to occur in South Africa, Zimbabwe, Tanzania, and Greenland.

Diatreme–Dyke–Sill Relationships

Some diatremes have been excavated sufficiently to show that they contract downwards into dykes or groups of dykes. The zones of diatremes appear to be a surface expression. The dykes occur near the roots of diatremes, which rise to considerable heights above this zone. These dykes have been referred to as antecedent dykes, but the relationship is complicated, because the diatremes flare well above the dykes, intrude the dykes, and contain different types of kimberlite. Dykes may branch into sills.

The complexity of the relationships between the various bodies is shown in Figures 4 and 5, which show a diatreme in Yakutia, and Figure 6, which depicts an idealized system.

Geotectonic Settings

Most kimberlites are found on stable cratons that have not been deformed since the Precambrian (Figure 7). Such areas have experienced epeirogenic

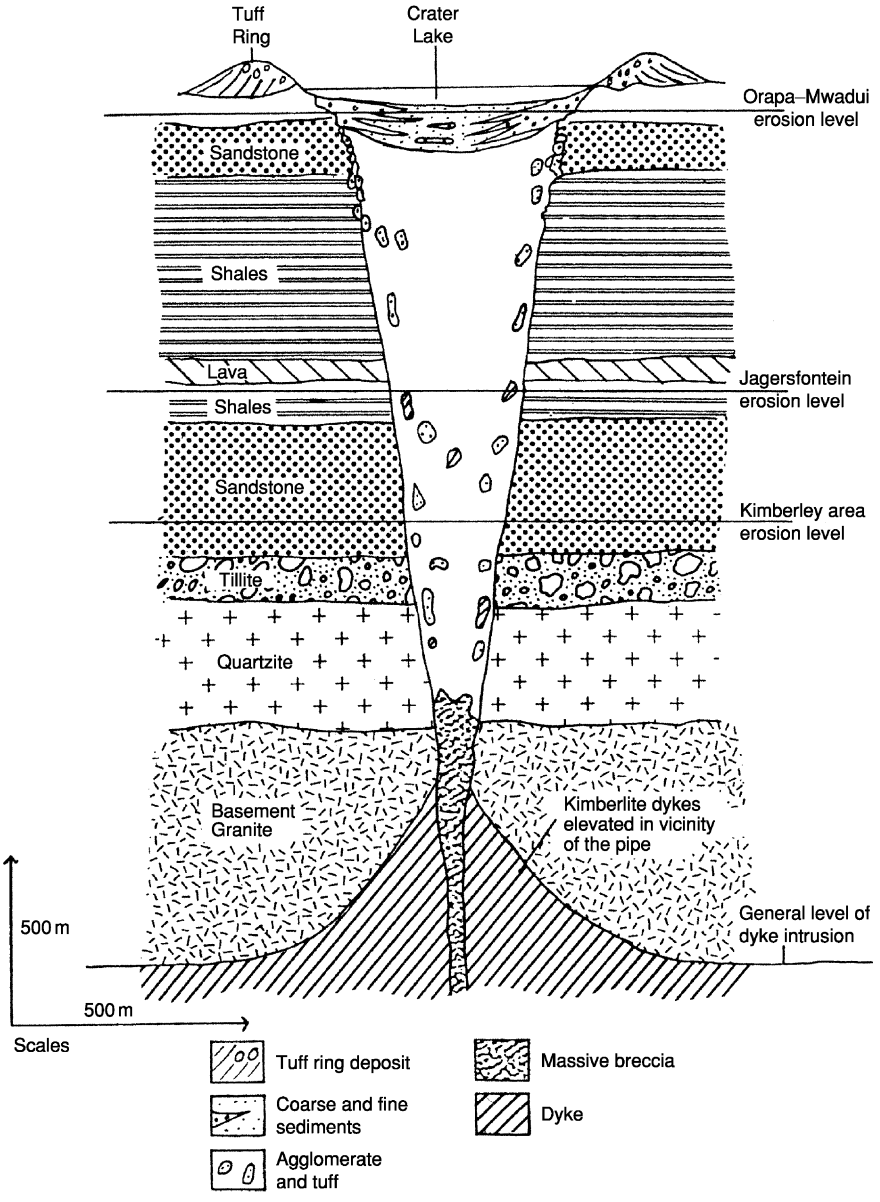


Figure 3 Generalized model of a kimberlite diatreme and its sub-diatreme dykes. (Reproduced from Dawson JB (1980) *Kimberlites and Their Xenoliths.* Berlin, Heidelberg, New York: Springer-Verlag.)

flexuring prior to the intrusion. They may be associated with local dome or swell structures. They occur within fundamental fractures that transect the deepest known basement terrains. Although they also occur in circumcratonic fold belts, they were not intruded there during orogenic cycles. They typically occur both on cratons and in the marginal fold belts where magmatic activity was otherwise very limited. Where the intrusions are underlain by circumcratonic orogenic belts with limited contemporaneous magmatic activity, they are characteristically non-diamondiferous.

Age Relationships

Kimberlite developments in the various shield areas or provinces show various ages. For example, the Premier diatreme near Pretoria is transected by a basic sill that is confidently dated at 1115 Ma, which means that the kimberlite is Precambrian and older than this intrusion into it. A date in the Permian has been obtained for the Dokolwayo occurrence in Swaziland on the eastern edge of the Kaapvaal Craton, although the majority of the South African occurrences can be dated radiometrically as

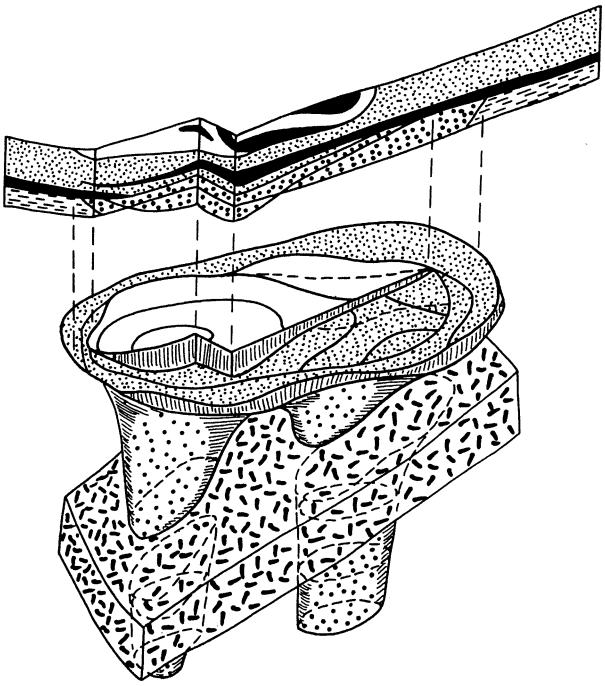


Figure 4 Isometric diagram of the double Palaeozoic-age Krasnopresneskaya pipe, Yakutia; the pipe has been intruded by a basaltic sill (dashed), and, at the top, bowl-shaped depressions are filled with younger volcaniclastic sedimentary rocks and covered by basalt flows. (Reproduced from Mitchell RH (1995) *Kimberlites, Orangeites and Related Rocks.* New York: Plenum Press.)

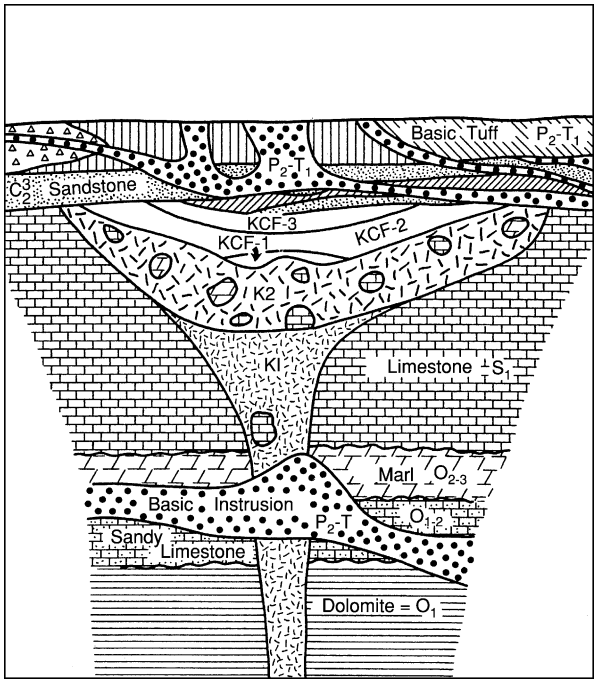


Figure 5 Cross-section of the diatreme shown in Figure 4. K1 and K2, kimberlite breccias; O, Ordovician country rocks; S, Silurian country rocks; C, later Carboniferous rocks; P-T, later Permo-Triassic rocks. (Reproduced from Mitchell RH (1995) *Kimberlites, Orangeites and Related Rocks.* New York: Plenum Press.)

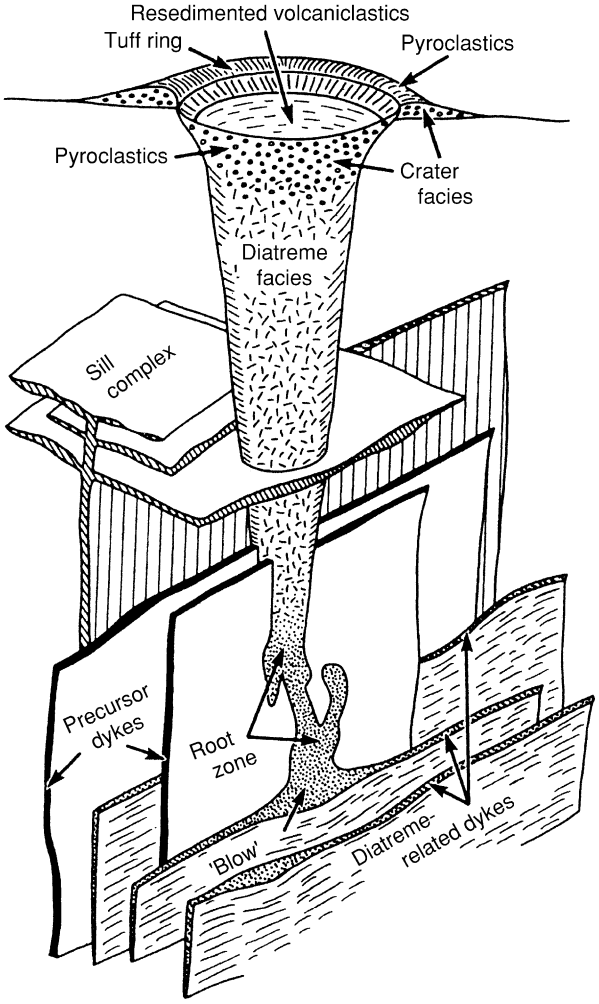


Figure 6 An idealized kimberlite magmatic system (not to scale). Hypabyssal rocks include sills, dykes, root zone, and a 'blow' (Reproduced from Mitchell RH (1995) *Kimberlites, Orangeites and Related Rocks.* New York: Plenum Press.)

post-Jurassic (Figure 8). In West Africa there are four ages of intrusions into the Eburnian Craton: 2100–2300 Ma, 1150 Ma, 700 Ma, and 60–100 Ma. The Mwadui occurrence, the largest kimberlite diatreme known, is post-Precambrian and has not been substantially eroded: the original crater-like sediments at the top are preserved. The Igwise occurrences are recent. Two distinct ages of 1140 Ma and 840 Ma have been obtained for diatremes in India; the Siberian kimberlites are of at least four ages: Precambian, Palaeozoic, and both Early and Late Mesozoic (Figure 9). Whereas single diatremes or diatreme complexes represent intrusions of a single suite over a short period, repeated intrusions far apart in time within a single shield may use the same fracture systems.

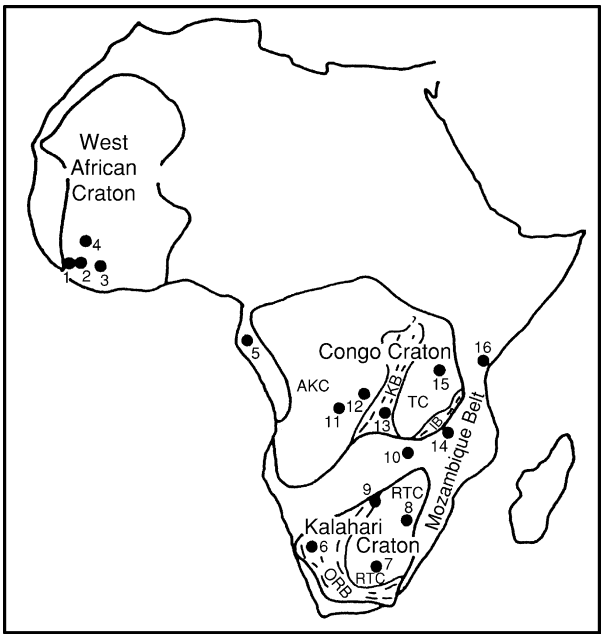


Figure 7 The relation of kimberlite occurrences in Africa to ancient cratons and their boundaries. (Reproduced from Dawson JB (1980) *Kimberlites and Their Xenoliths.* Berlin, Heidelberg, New York: Springer-Verlag.)

Kimberlite magmatism has occurred at various times in the Earth's history, one major phase being in the Cretaceous. South and West Africa have seen three or four periods of kimberlite magmatism, and there have been several epochs of such magmatism in the central USA and Yakutia. Certain geographical areas coinciding with old cratons are 'kimberlite prone'.

Xenoliths

Within kimberlites a wide variety of rock types occur as fragments. Minerals derived from these rocks also become incorporated within the host kimberlite. Thereby, the mineralogy becomes extremely complex. There are five sources of xenoliths:

- fragments of the immediate wall rocks at the level of preservation;
- fragments derived from earlier formations that existed at the time of intrusion but have since been removed by erosion;
- blocks of still-buried formations through which the intrusion passed;

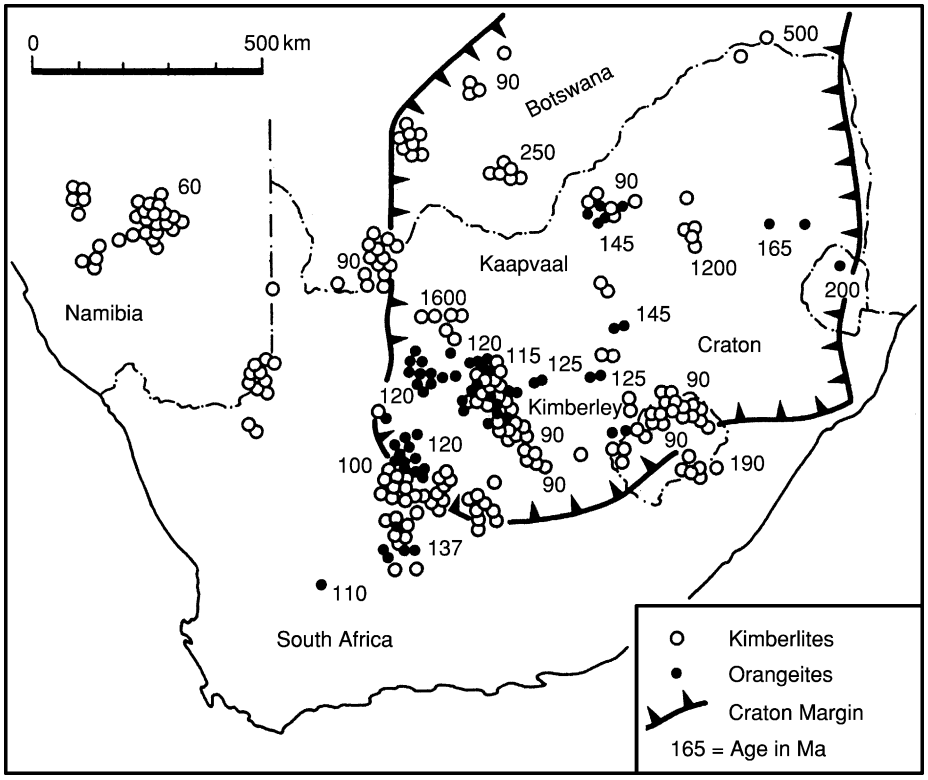


Figure 8 Kimberlite and 'orangeite' occurrences in southern Africa, with their radiometrically derived ages. Note the Precambrian age (1200 Ma) at the site of the Premier Mine and the single Permian (200 Ma) age in Swaziland. (Reproduced from Mitchell RH (1995) *Kimberlites, Orangeites and Related Rocks.* New York: Plenum Press.)

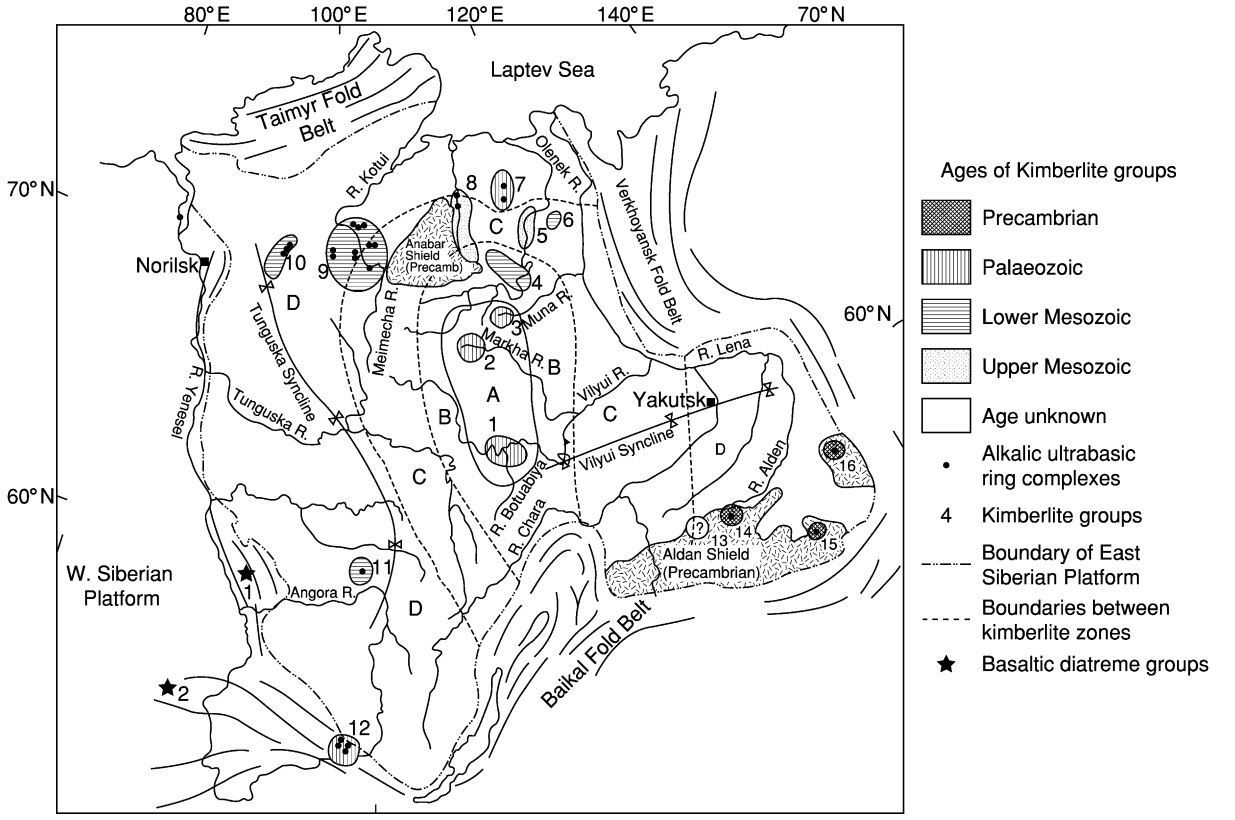


Figure 9 Distribution of kimberlites in the East Siberian Platform, showing the different ages. (Reproduced from Dawson JB (1980) *Kimberlites and Their Xenoliths.* Berlin, Heidelberg, New York: Springer-Verlag.)

- granulites derived from deep metamorphic terrains (in the crust); and
- fragments believed to come from the Earth's upper mantle (containing high-pressure mineral phases).

Examples of xenoliths derived from earlier formations are the Karoo basalt xenoliths found at Kimberley, 350 km from their nearest outcrop: they are estimated to have descended 1500 m within the eroded-off top of the intact diatreme before it was mined. In Yakutia, blocks derived from still-buried formations may include fragments of the ancient Precambrian shield (the crystalline Siberian Platform).

Xenoliths of rock types with a mineralogy and density indicating higher pressures and temperatures than those normally pertaining within the Earth's crust are believed, by inference, to emanate from the mantle. They not only provide the most important evidence of the components of the mantle but also provide an insight into the processes that generated the kimberlite diatreme, which originated at great depths. The have been exhaustively studied, and state-of-the-art methods of investigating them are continually being developed, in particular those based on isotopes. Such methods are being extended to single mineral grains and inclusions within diamonds.

The deep-seated xenoliths may be divided into five main groups: peridotite–pyroxene suite; eclogites and grospydites; metamorphosed peridotites (rich in amphibole or mica); glimmerites and the MARID suite; and miscellaneous. Together, these xenoliths usually do not amount to more than 2% of the bulk rock, but in Lesotho they may form 20–30% (Figure 10). There is much variation from diatreme to diatreme: in some the peridotite suite predominates, whereas in others eclogite is dominant.

The peridotite–pyroxene suite, of ovoid to discoidal xenoliths up to 30 cm in maximum dimension, includes various types distinguished on the basis of modal mineral composition – dunite, lherzolite, harzburgite, websterite, and pyroxenite. The Russians in Yakutia have noted that the spinel in these rocks is magnetite, not the chromite that is characteristic of peridotites not occurring as xenoliths in kimberlite. The origin of the rounding is obscure. Most xenoliths are structureless, but a few are banded. Globules of

Figure 10 Kimberlite choked with mantle-derived xenoliths; Matsoko pipe, Lesotho. (Reproduced from Nixon PH (ed.) (1973) *Lesotho Kimberlites.* Lesotho National Development Corporation.)

sulphide and oxides of ore minerals may be present. Some garnet lherzolites contain high concentrations of rare earth elements with the lighter ones being enriched, and there is some evidence that they were originally cumulates.

The eclogites are coarse-grained rocks composed of pyroxene and garnet, with many accessory minerals (Figure 11), and they are commonly banded. Again, rounded xenoliths are common. The peraluminous grospydites are eclogites containing grossular garnet, kyanite and/or corundum, and rarely sillimanite; there are also quartz eclogites, amphibole eclogites, and orthopyroxene eclogites. Diamond-bearing eclogites are known from at least nine intrusions in South Africa, the AK/1 pipe in Botswana, and the Mir pipe, Yakutia; the diamonds are mostly high-temperature octahedra (though some are rhombic dodecahedra), and they may be accompanied by graphite.

Temperature ranges of equilibration of eclogite xenoliths, peridotite and eclogite inclusions in diamonds, and the megacryst suite are given in Figure 12.

Glimmerites are xenoliths dominated by phlogopite, which are found in Yakutia. The MARID-suite xenoliths, found in South Africa, are also dominated by phlogopite, with various amounts of amphibole, clinopyroxene, ilmenite, rutile, apatite, and olivine. These are thought to have crystallized high in the mantle under oxidizing conditions, but alternatively it has been suggested that they are of wall-rock metasomatic origin.

The miscellaneous xenoliths include alkremites (found in the Udachnaya pipe in the former Soviet

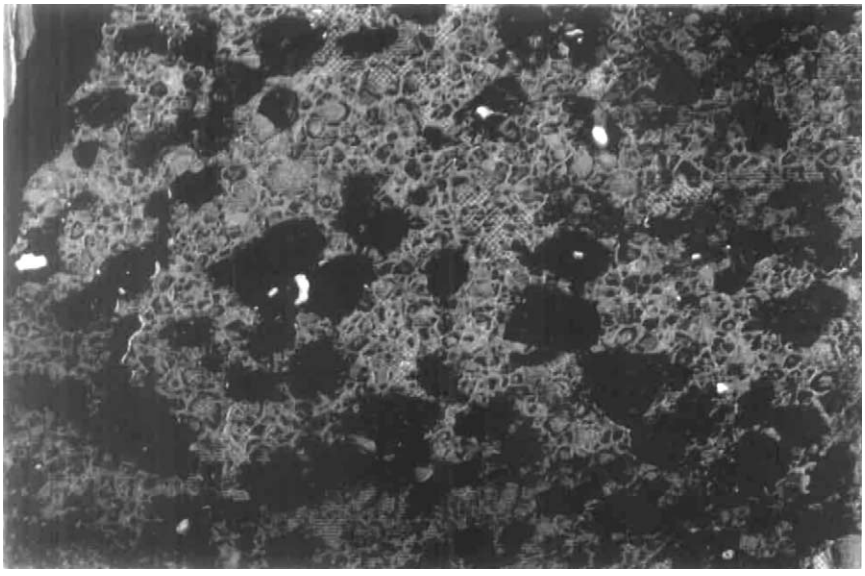


Figure 11 Eclogite nodule; Kao pipe, Lesotho. (Reproduced from Nixon PH (ed.) (1973) *Lesotho Kimberlites.* Lesotho National Development Corporation.)

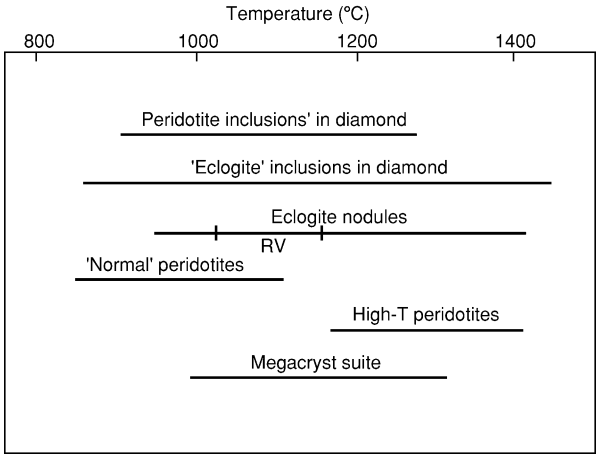


Figure 12 Temperature ranges of various inclusions and 'nodules', normal peridotites, and the megacrysts suite (Reproduced from Dawson JB (1980) *Kimberlites and Their Xenoliths.* Berlin, Heidelberg, New York: Springer-Verlag.)

Union), which are coarse assemblages of garnet and dark green spinel. They are interpreted as early dense cumulates of an aluminous mantle under pressures equivalent to a depth of 70 km.

Megacrysts

Single large crystals (more than 2 cm) are common in kimberlites; they are alternatively referred to as discrete nodules. Two series are recognized: chrome-rich and chrome-poor. Such megacrysts may contain other minerals. There are also diopside, mica, and zircon megacrysts. It has been suggested that some megacrysts are genetically linked to the kimberlite, but such a relationship has been rejected by some authors, who consider the megacrysts to be accidental mantle-derived inclusions.

The minerals found in kimberlites and their origins are summarized in Table 3.

Studies of Diamonds and Their Inclusions

Nowadays many studies are carried out on diamonds in kimberlites and the inclusions within them, in an attempt to determine the pressure and temperature (i.e. mantle depth) at which they were formed. In a study of the diamonds from Pipe DO-27, Lac de Gras, Slave Craton, North-west Territories, Canada, a high proportion were found to be cubo-octahedral stones with resorption characteristics. Syngenetic inclusions are of three suites: 25% are peridotitic (garnet, clinopyroxene, sanidine, sulphide); 50% are eclogitic (garnet, olivine, orthopyroxene, sulphide); and 25% are super-deep (ferropericlase, magnesium-perovskite, nickel). The super-deep suite of minerals represents the mantle at depths greater than 670 km.

Upper Mantle Configuration

The Earth's upper mantle is petrologically complex. Xenoliths and xenocrysts from kimberlites provide a wealth of high-temperature and high-pressure rock types therefrom, but the problem is to delineate where in the mantle they originated and how they relate to each other genetically. Two main zones – the harzburgite zone and the garnet lherzolite zone – can be recognized, but these exhibit large- and small-scale heterogeneities, represented by pyroxenites, eclogites, and the MARID suite of rocks. These differences probably represent both lateral and vertical heterogeneities in the mantle. Shearing and metasomatism are also evident; brittle and plastic deformation accompany magma rise and open up channels. Convective processes may also operate. A tentative model for the upper mantle and crust derived from studies of kimberlite xenoliths and xenocrysts is shown in Figure 13. The super-deep inclusions in diamonds, mentioned above, represent a further zone below 670 km.

This field of research is on going, and each kimberlite conference produces new evidence and models, without necessarily clarifying the picture, such are the contradictions and differing interpretations of the evidence.

Diamond Provenance

There are several models for the origin of the diamonds in kimberlites. Since they are found within eclogites of established mantle provenance, it is reasonable to suppose that they come from the mantle, where high pressures and temperatures occur. Current models of diamond formation differ mainly in the source of the carbon. Juvenile methane or other hydrocarbons, oxidized during ascent through the upper mantle or at the lower boundary of the lithosphere, is favoured by one school. This is supported by the presence of peridotite inclusions in some diamonds. Other models invoke crystallization from kimberlite liquids or from ultrabasic melts during the formation of cratonic roots. Another set of models introduces the carbon to the mantle by subduction processes, the carbon being not juvenile and possibly biogenic in origin. This model is supported by the existence of diamonds containing inclusions of the eclogite suite. It seems likely, however, that several diamond-forming processes operate. Regardless of the model favoured, it is generally accepted that most diamonds are xenocrysts in the transporting kimberlite magma. Disruption and disaggregation of diamond-bearing horizons by the passage of magma

Table 3 A summary of the disaggregated minerals found in kimberlites, and their origins x, present; xx, fairly abundant; xxx, abundant. (Reproduced from Nixon PH (1980) The morphology and mineralogy of diamond pipes. In: Glover JE and Groves DI (eds.) *Kimberlites and Diamonds*, pp. 32–47. Nedlands: University of Western Australia.)

Earth zone (Host rock)	*Disaggregated resistant mineral suite*	*Composition (approximate values)*	*Relative abundance*	*Comment*
Upper Crust	Widely variable;	–	xxx	Scapolite, hornblende, apatite, etc. also present
Lower Crust	Garnet,	Py–alm,	x	
(granulites,	Clinopyroxene,	Omphacite,	x	
etc.)	Orthopyroxene,	Hypersthene	x	
	Rutile		x	
Lithosphere	Olivine	Fo_{93}	xxx	Depleted mineralogy – high Mg/Fe, Cr/Al and
(lherzolites	Orthopyroxene	En_{94}	xxx	low Ca and Na, although not as extreme as
etc.)	Clinopyroxene	Cr diopside	xx	in the diamond inclusion suite. Differentiation
	Garnet	Cr pyrope	xx	within the mantle locally extends chemical
	Spinel	Chromite	x	ranges. Metasomatism produces additional
	Mica	Phlogopite	xx	minerals similar to kimberlite suite below.
				Diamond and graphite recorded
Mantle	Garnet	Py–alm–gross	x	Rare suite but is locally abundant. Generally
differentiates	Clinopyroxene	Omphacite	x	much higher Fe/Mg ratios than in lherzolites
(eclogite and	Rutile, ilmenite		x	
grospydite)	Kyanite		x	
	Corundum		x	
	Diamond		x	
Kimberlite suite	Olivine	$Fo_{85–95}$	xxx ?	Many secondary minerals not listed but note
(including	Mica	Phlogopite	xxx	wide variety of fine-grained spinels and
MARID	Amphibole		x	perovskite. Zircon present
'precipitates')	Rutile		x	
	Ilmenite	Mg-bearing	xx	
	Clinopyroxene	Diopside	xx	
	Garnet	Py–alm	x	
	Diamond		x	
Mantle residue	Olivine	$Fo_{93–94}$	x	Rare suite in concentrates. 'Ultra-depleted' –
at kimberlite	Orthopyroxene	$En_{92–96}$	x	extremely high Mg/Fe, Cr/Al and low Ca, Ti,
foci (diamond	Clinopyroxene	Cr diopside	x	and Na
inclusion	Garnet	Knorringite–pyrope	x	
suite)	Spinel	Chromite	x	
Asthenosphere	Olivine	Fo_{90}	x	'Fertile' composition, i.e., relatively low Mg/Fe
(lherzolites,	Orthopyroxene	Calcic bronzite (En_{87})	x	and Cr/Al but higher Ti compared with
etc.)	Clinopyroxene	Subcalcic diopside	x	typical lithosphere
	Garnet	Pyrope	x	
Discrete	Orthopyroxene	Calcic bronzite (En_{87})	x(x)	Silicates have similar composition to those in
nodules	Clinopyroxene	Subcalcic diopside	xx	fertile nodules with a tendency to higher
(megacrysts)	Garnet	Pyrope (little alm)	xxx	Fe/Mg ratios. Ilmenite pyroxene lamellar
	Ilmenite	Mg-bearing	xxx	intergrowths are part of this suite
	Zircon		x	

from great depths incorporates the diamonds as xenocrysts in the magma. Very hot magmas may completely resorb the diamonds. Only magmas rising from great depths will carry diamonds.

Major Provinces

The global distribution of major kimberlite provinces is shown in Figure 14. The recently discovered province of diamondiferous kimberlites in the Slave Craton of Canada, is shown. It should be noted that province number 9, in the north of Western Australia, is a lamproite province not a kimberlite province, although it is a rich source of diamonds. Southern Africa is overwhelmingly the site of the most numerous kimberlite developments, as shown in Figure 8.

Prospecting Methods

Because kimberlite pipes rarely have a surface expression, prospecting for diamonds is arduous and requires the application of specialized methods. This in turn requires specialized training. The methods include:

- separation of heavy minerals (coarse and fine fractions): red-brown pyrope (discrete nodule

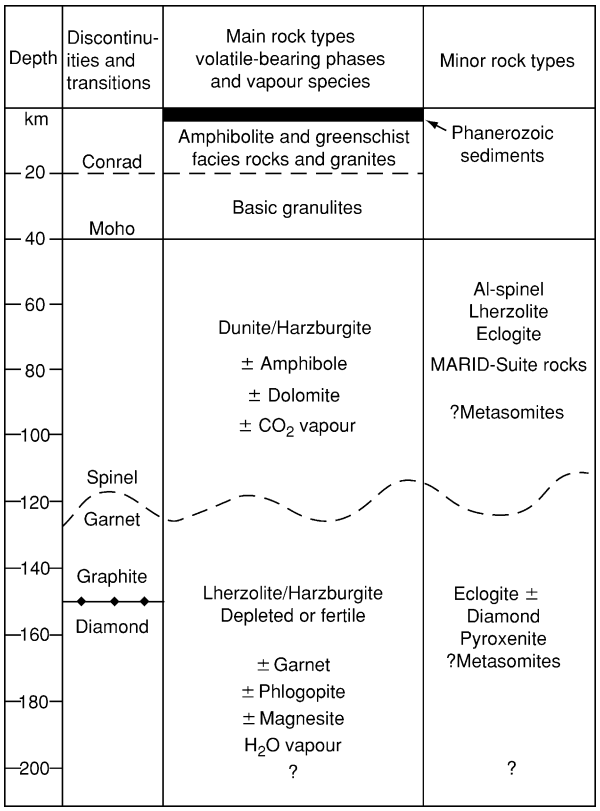


Figure 13 Tentative model for the depth distributions of the various rock types and the volatile-bearing and vapour phases in the subcratonic lithosphere, derived from kimberlite occurrences. The depth of the Conrad discontinuity is based on seismic work on the South African craton. (Reproduced from Dawson JB (1980) *Kimberlites and Their Xenoliths.* Berlin, Heidelberg, New York: Springer-Verlag.)

suite); purple-red chromium-pyrope (depleted lherzolite); ilmenite (coarse, rounded, with perovskite alteration or cream skin, distinct from the tabular variety); chromium-diopside (depleted lherzolite); coarse chromite; enstatite, and olivine grains;
- geochemistry (trace element suite);
- scintillometry;
- satellite imagery;
- air-borne photography; and
- geophysical surveys (magnetic, electromagnetic, induced polarization, resistivity, seismic, gravity).

Very few pipes carry diamonds in the kimberlite rock (Figure 15), and those that do may not contain economic deposits.

Mining

Mining of kimberlites may be opencast, as in the case of the Kimberley occurrence (Figure 16). Here, before it was consolidated, there were innumerable operators working small claims in the pit and transporting the ore up to the surface on a profusion of cable ways. Mining may also involve deep mining methods, as in the case of the Premier Mine, where a thick basic sill prevented opencast working below a certain depth. Here, the operation is large scale and standard-gauge electric trains move about underground, powered by overhead lines. This major operation in a huge pipe aims to recover diamonds that constitute a minute fraction of the rock, something like 1/100 000.

Lamproites

Lamproites are much less abundant than kimberlites, but like them most are the products of diatremes emanating from great depths. They were not considered important until the 1960s, when J B Jeppe, a South African, followed up a long-standing report of a single diamond find in the north of Western Australia and looked at the lamproites of the Fitzroy Region, earlier mapped by A Wade and petrologically described under a collection of rather unnecessary rock designations (cedricite, wolgidite, etc). Jeppe found a single diamond in one diatreme outcrop, but it was not until the early 1970s that a team from Tanganyika Holdings and London Tin, in a syndicate with four other companies, carried out extensive prospecting using well-tried methods and located the Argyle occurrences of the AK1 pipe and also alluvial deposits, both of which contained diamonds (Figure 17). Approximately one-quarter of the world's diamond production is now concentrated in this area (Figure 18). The pipes are of Miocene age (20 Ma), and there are more than 100 separate pipes, plugs, and sills, with some rare dykes.

Lamproites are potassic leucite-bearing rocks and include olivine lamproite, olivine–diopside lamproite, and leucite lamproites containing phlogopite, diopside, and postassic richterite. They range from ultrabasic (20–29% MgO) to basic (5% MgO). The lamproites are derived from depleted (garnet–clinopyroxene-poor) peridotite mantle beneath the Kimberleys region. This means that kimberlite indicator minerals are very scarce compared with regions where minerals are being shed from kimberlite sources, making exploration more difficult. Lamproites seem to occur on the margins of Archaean cratons. They have been found in Spain, Corsica, Zambia, West Africa, India, the USA (Figure 19), Gaussberg Island off Antarctica, and, recently, Brazil, where there are also kimberlites containing diamonds. There are two or three occurrences in South Africa, which have an atypical mineralogy and are possibly transitional to kimberlites. Diamonds were discovered in the Prairie Creek, Arkansas, occurrence in 1906, but it was misidentified as a kimberlite;

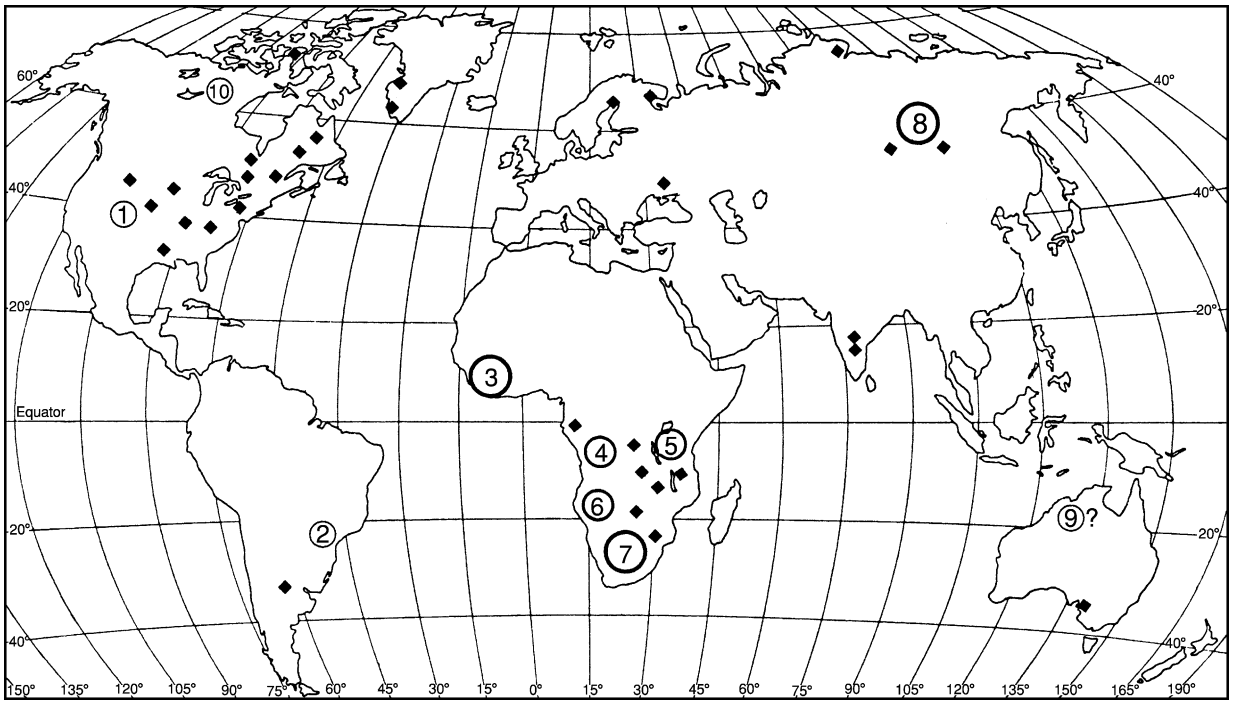


Figure 14 Global distribution of major kimberlite and diamond provinces (circles with numbers) and more isolated occurrences (black diamond symbols). Number 9 is a lamproite province, not a kimberlite province, and number 10 is a recent major development.

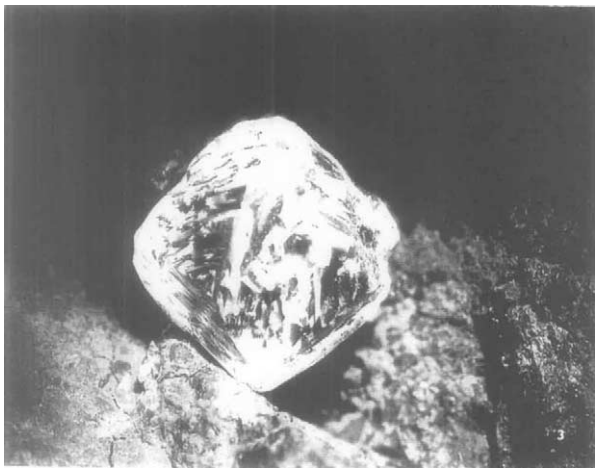


Figure 15 An uncut diamond in kimberlite, South Africa, 10×10 mm. (Reproduced from Bauer J and Bouska V (1983) *A Guide in Colour to Precious and Semi-Precious Stones*, pp. 80–83. London: Octopus.)

Figure 16 The Kimberley diatreme worked out. (Reproduced from Dawson JB (1980) *Kimberlites and Their Xenoliths.* Berlin, Heidelberg, New York: Springer-Verlag.)

the Mahjgawan, India, occurrence yielded diamonds even earlier, in 1827, but was also misidentified as a kimberlite.

Lamproites and the diamonds within them are derived from the mantle, but they do not appear to have much petrogenetic relationship to kimberlites, for

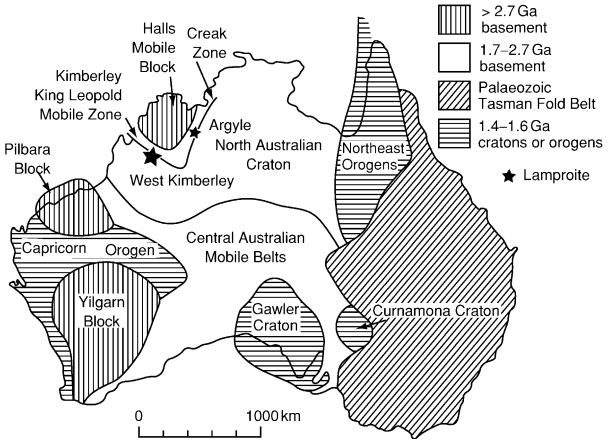


Figure 17 The Precambrian cratonic areas of Australia, showing the location of the Leopold Zone lamproite occurrences and the Argyle diamondiferous pipe. (Reproduced from Mitchell RH and Bergman SC (1991) *Petrology of Lamproites.* New York: Plenum Press.)

Figure 18 A 58 carat pink cut diamond worth $A38 000 (≃£18 000 sterling) from the Argyle Mine in the Kimberleys lamproite province, Western Australia (photograph supplied by the Western Australian Museum).

which they have been mistaken, although the South African occurrences do suggest the possibility of a connection.

Kimberlites and Carbonatites

It has been suggested, particularly in Russia, that there is a relationship between kimberlites and carbonatite intrusions and volcanic eruptives, but,

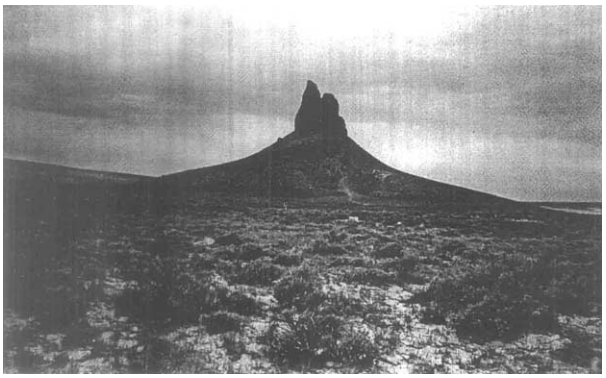


Figure 19 Boars Tusk, a lamproite neck in the Leucite Hills, Wyoming, USA. Some Western Australian lamproites form similar but less dramatic features. (Reproduced from Mitchell RH and Bergman SC (1991) *Petrology of Lamproites.* New York: Plenum Press.)

although both are of deep-seated origin, carbonatites occur in and close to rifted geotectonic zones (for example East Africa), characteristically form ring complexes in the underworks of volcanoes (as in the huge eroded Kisingiri volcano in Kenya), and are associated with sodic rather than potassic alkaline igneous suites (intrusive ijolites, melteigites, and jacupirangites carrying nepheline and aegirine, also intrusive uncomopahgrites and turjaites carrying melilite and perovskite, and surface volcanics, nephelinites, and melilitites).

See Also

Earth: Mantle. **Gemstones**. **Geochemical Exploration**. **Igneous Processes**. **Igneous Rocks:** Carbonatites. **Shields**. **Volcanoes**.

Further Reading

Bauer J and Bouska V (1983) *A Guide in Colour to Precious and Semi-Precious Stones,* pp. 80–83. London: Octopus.

Davies RH, Griffin WL, Pearson NL, *et al.* (1999) Diamonds from deep pipe DO-27, Slave Craton, Canada. In: Gurney JJ, *et al.* (eds.) *The J B Dawson Volume: the P H Nixon Volume,* Proceedings of the 7th International Kimberlite Conference April 1998, Cape Town, pp. 148–155. Red Roof Design.

Dawson JB (1980) *Kimberlites and Their Xenoliths.* Berlin, Heidelberg, New York: Springer-Verlag.

Glover JE and Groves DI (eds.) (1980) *Kimberlites and Diamonds.* Nedlands: University of Western Australia, Geology Department and Extension Service.

Glover JE and Harris PG (eds.) (1985) *Kimberlite Occurrence and Origin: A Basis for Conceptual Models in*

Exploration. Nedlands: University of Western Australia, Geology Department and Extension Service.
Gurney JJ, *et al.* (eds.) (1999) *The J B Dawson Volume: the P H Nixon Volume*. Proceedings of the 7th International Kimberlite Conference April 1998, Cape Town. Red Roof Design.
Janse AJA (1995) A history of diamond sources in Africa: part I. *Gems and Gemmology* 31: 228–255.
Janse AJA (1998) A history of diamond sources in Africa: part II. *Gems and Gemmology* 32: 2–30.
Jaques AL, Lewis JD, and Smith CB (1986) The kimberlites and lamproites from Western Australia. *Geological Survey of Western Australia Bulletin* 132: 1–268.
Kornprobst T (ed.) (1984) *Kimberlites and Related Rocks*. Proceedings of the 3rd International Kimberlite Conference. Amsterdam, Oxford, New York, Tokyo: Elsevier.
Mitchell RH (1995) *Kimberlites, Orangeites and Related Rocks*. New York: Plenum Press.
Mitchell RH and Bergman SC (1991) *Petrology of Lamproites*. New York: Plenum Press.
Nixon PH (ed.) (1973) *Lesotho Kimberlites*. Lesotho National Development Corporation.
Nixon PH (1980) The morphology and mineralogy of diamond pipes. In: Glover JE and Groves DI (eds.) *Kimberlites and Diamonds*, pp. 32–47. Nedlands: University of Western Australia, Geology Department and Extension Service.
Nixon PH (1980) Regional diamond exploration: theory and practice. In: Glover JE and Groves DI (eds.) *Kimberlites and Diamonds*, pp. 64–75. Nedlands: University of Western Australia, Geology Department and Extension Service.
Sobolev NV (1977) *Deep-Seated Inclusions in Kimberlite and the Problem of the Upper Mantle*. Washington: American Geophysical Union.

Komatiite

N T Arndt, LCEA, Grenoble, France
C M Lesher, Laurentian University, ON, Canada

Introduction

It is easy to explain roughly what a komatiite is but difficult to give a rigorous definition. The simple definition, as given by Arndt and Nisbet in 1982; is that komatiite is an ultramafic volcanic rock. A limit of 18% MgO separates komatiites from less magnesian volcanic rocks such as picrites, ankaramites, or magnesian basalts. The term komatiitic basalt is applied to volcanic rocks containing less than 18% MgO that can be linked, using petrological, textural, or geochemical arguments, to komatiites. The definition seems to include intrusive rocks; this was probably not intended.

Implicit to the definition of komatiite is the notion – difficult to prove – that komatiites crystallize from liquids that contained more than about 18% MgO. Complications arise from the existence of other MgO-rich volcanic rocks that either formed through the accumulation of olivine from less magnesian liquids, or crystallized from magmas with geochemical characteristics quite unlike those of most komatiites. An example of the first type is a phenocryst-charged basaltic liquid (a picrite according to some definitions); an example of the second is meimechite, a rare alkaline ultramafic lava (*see* **Lava**) with unusual major and trace element composition.

To distinguish komatiite from other types of highly magnesian volcanic rock, it is useful to include spinifex texture in the definition (**Figure 1A, B**). Spinifex, a texture characterized by the presence of large skeletal or dendritic crystals of olivine or pyroxene, is present in many, but not all komatiite flows. A workable definition of komatiite should include the phrase "komatiite is an ultramafic volcanic rock containing spinifex or related to lavas containing this texture". With the last part of the definition, we can make allowance for the manner in which texture varies within komatiitic units. For example, many komatiite flows have an upper spinifex-textured layer and a lower olivine-cumulate layer (**Figure 1C**); and other flows grade along strike from layered spinifex-textured portions to massive olivine-phyric units. With the inclusion of the phrase about spinifex, the lower olivine-cumulate portions of layered flows or the olivine-phyric units can also be described as komatiite. On the other hand, meimechites, picrites, and other rock types that contain no spinifex are excluded. For further discussion, see the **Further Reading** section at the end of this article.

Because of the facility with which olivine fractionates or accumulates in low-viscosity ultrabasic liquids, the compositions of komatiite lavas vary considerably. MgO contents range from 18% (the limit between komatiite and komatiitic basalt), to as high as 50% in the lower cumulate portions of layered flows (**Figure 2**). Elements that are immobile during the metamorphism and hydrothermal alteration, which affects all komatiites to a greater or lesser extent,

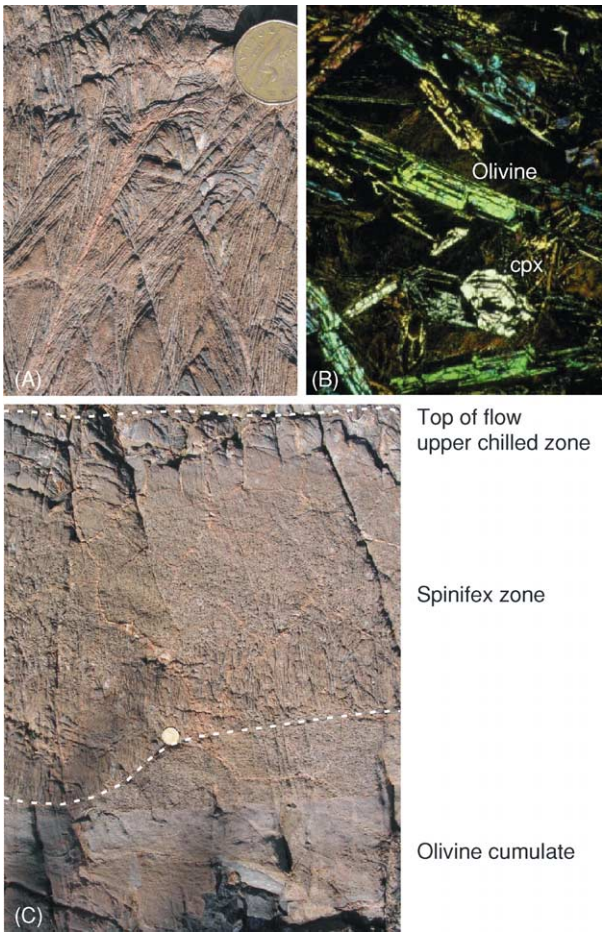


Figure 1 (A) Spinifex-textured komatiite in outcrop (sample from Munro Township, Canada). The bladed habit of large parallel olivine crystals is clearly visible. (B) Spinifex-textured komatiite in thin section. Sample from the Belingwe Belt, Zimbabwe. Skeletal olivine crystals lie in a matrix of clinopyroxene and altered glass. (C) The upper spinifex-textured portion and the upper part of the olivine cumulate of a layered komatiite flow from Munro Township. The curved contact between the spinifex and cumulate zones is unusual; in most flows this contact is horizontal.

plot on olivine-control lines in variation diagrams. The least magnesian compositions are found in the lower parts of spinifex zones, the most magnesian in the olivine cumulates. Chilled flow tops and porphyritic lavas have intermediate compositions.

The maximum MgO contents of komatiite liquids, estimated using the compositions of chilled flow margins and the forsterite contents of olivine, are between 28 and 30%. The dry 1-atmosphere liquidus temperature of these liquids, calculated from experimental data, is between 1560 and about 1600°C.

Several different geochemical types can be distinguished using Al_2O_3/TiO_2 and rare earth elements (REE). The komatiites first discovered by Viljoen and Viljoen in 1969 are of the 'Al-depleted' or Barberton type. As shown in Figure 3, these rocks have relatively low Al_2O_3/TiO_2 and Gd/Yb, the latter ratio being a measure of relative depletion of the heavy rare Earth elements (HREE). The second group of komatiites, called 'Al-undepleted' or Munro-type, has near chondritic ratios of Al_2O_3/TiO_2 and Gd/Yb. A third type, Al-enriched komatiite, is common in komatiites from Gorgona Island, Colombia, described by Echeverria in 1980, and a fourth type, Ti-rich komatiite occurs in the Baltic shield and in other parts of Ontario.

The typical habitat of komatiite is an Archaean greenstone belt. Ultramafic lavas comprise between 0 and about 20% of well-preserved volcanic successions and appear to have similar abundances in both Middle- and Late-Archaean belts. True komatiites are rare or absent in Proterozoic sequences – the spinifex-textured lavas of the Cape Smith belt have komatiitic basaltic compositions – but reappear in one notable example in the Cretaceous. The ~90 Ma ultramafic lavas of Gorgona Island are true komatiites which crystallised as spinifex-textured flows from liquids containing at least 20% MgO.

Formation of High-MgO Liquids

As shown in Figure 4A, magmas with highly magnesian, ultrabasic compositions form either through melting at high pressures, or by high percentages of mantle melting. The effect of increasing the pressure or depth of melting is to increase the stability of orthopyroxene relative to olivine, and that of garnet relative to the more magnesian mantle minerals (*see* **Earth:** Mantle). The consequence is the formation of ever more magnesian magma as the pressure increases. It has been shown that at high pressures, above about 8 GPa, near-solidus melts (liquids produced by melting of mantle peridotite at temperatures only slightly above the solidus) contain more than 30% MgO and have ultrabasic compositions. Increasing the percentage of melting has a similar effect. At shallower levels in the upper mantle, at pressures of ~0.5 to 5 GPa, the minerals that melt at low temperatures (plagioclase, spinel, garnet, and clinopyroxene) have relatively low MgO contents. As the degree of melting increases, the more magnesian minerals, olivine and orthopyroxene, progressively enter the liquid, increasing its MgO content. Figure 4A shows schematically how magmas with 30% MgO form either through deep melting near the solidus, or by higher degrees of melting at shallower levels. However, komatiites contain low concentrations of incompatible elements, which indicates that they formed through relatively high degrees of partial melting. Near-solidus melting therefore is an improbable

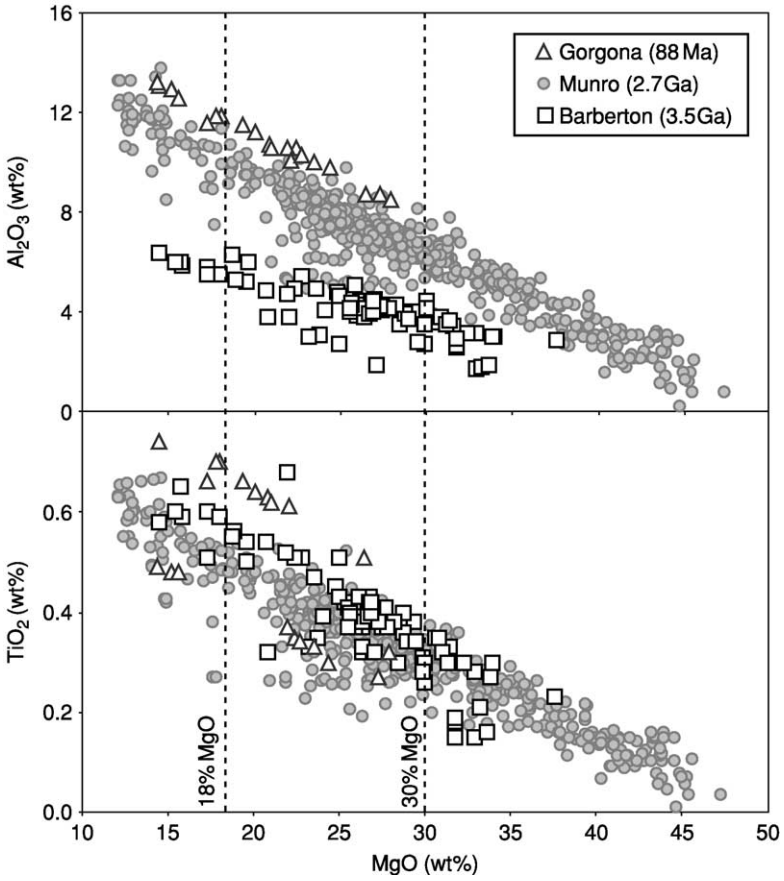


Figure 2 Variation diagrams illustrating the range of compositions of the main types of komatiite. The limit at 18% MgO separates komatiitic basalt from komatiite; the limit at 30% MgO indicates the probable maximum MgO content of komatiitic liquids. Rocks with more than 30% MgO are olivine cumulates.

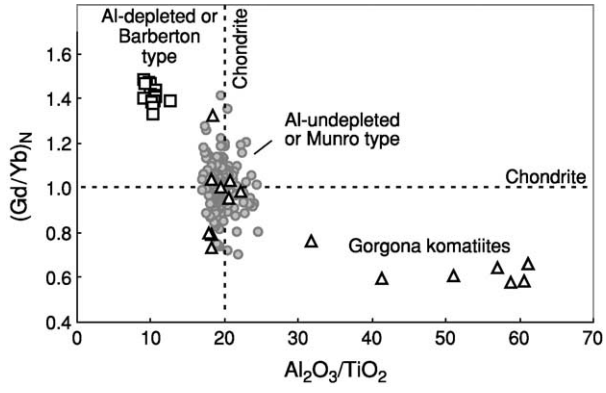


Figure 3 Al_2O_3/TiO_2 versus Gd/Yb showing the main types of komatiite. The high Gd/Yb and low Al_2O_3/TiO_2 of Barberton-type komatiite indicates that garnet was retained in the residue of melting and that the magmas formed at extreme mantle depths.

explanation for most komatiites: from Figure 5, we see that on the basis of Ce and Zr contents, they would have formed by 30 to 60% melting of mantle peridotite.

Eruption and Solidification

The viscosity of komatiite magma is very low, one to two orders of magnitude less than that of basalt. A viscosity of 1–2 Pascal-seconds was estimated for dry komatiite containing 28% MgO, which compares with 500–1000 Pascal-seconds for typical basalts. The low viscosity influences the way komatiite segregates from its source, rises through the lithosphere, and erupts on the surface.

Komatiite is a fragile magma. It is far hotter than surrounding rocks, especially when it passes through the lithosphere, and it has a strong capacity to interact with them. When it flows rapidly past crustal rocks, it is capable of thermally eroding and assimilating them; if it ponds in a crustal magma chamber, it will fractionally crystallise and lose its ultramafic character. For komatiite to reach the surface it must flow rapidly and continuously, without pausing on its way. Komatiite is also a relatively dense magma. Anhydrous komatiite, containing 28% MgO, has a density of about 2800 kg m^{-3}, significantly higher than

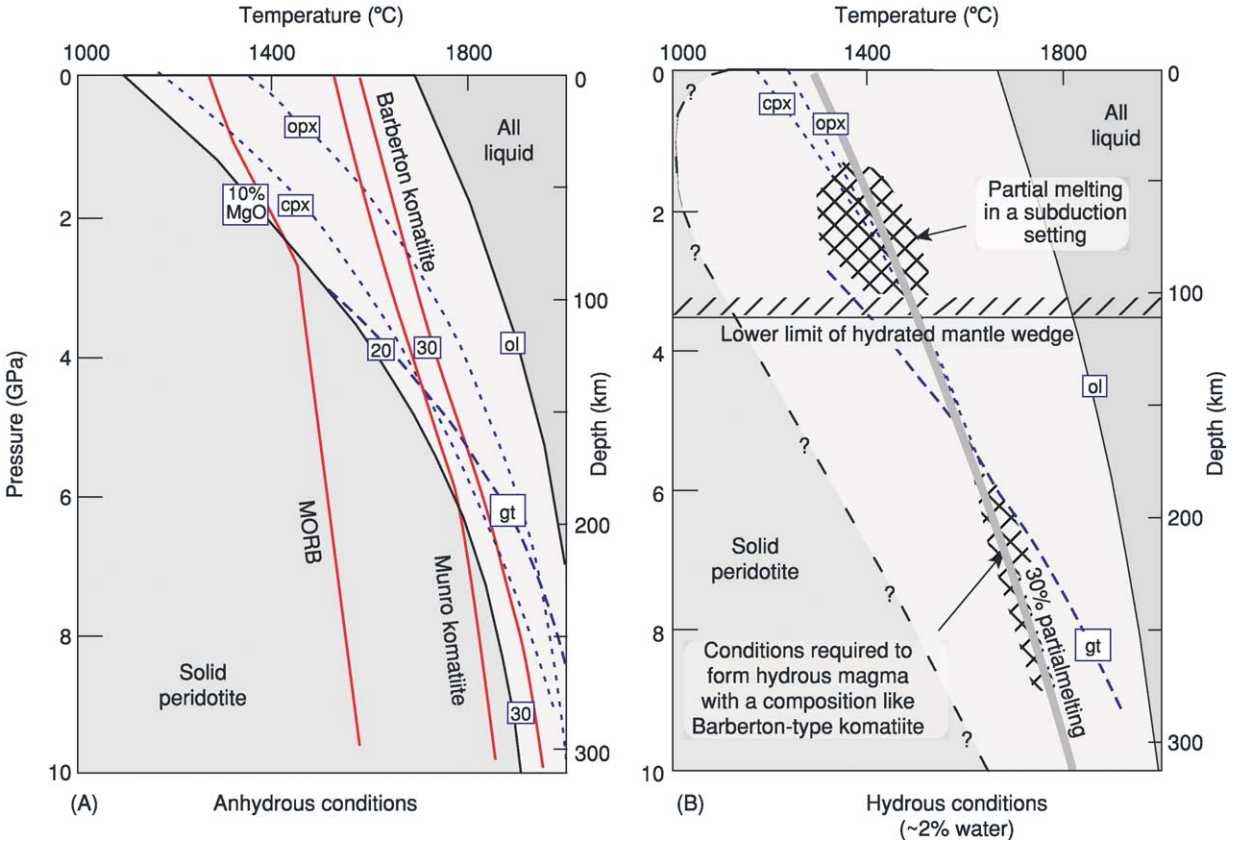


Figure 4 Melting behaviour of mantle peridotite and the formation of ultramafic magmas. (A) Anhydrous conditions. (Phase relations from Herzberg (1999)). The grey lines show the paths taken by mantle that undergoes partial melting. MORB forms from ambient, low-temperature mantle. Much high source temperatures, as in mantle plumes, are required to produce komatiites. The plumes intersect the solidus at great depths. The source of Munro-type komatiite starts to melt at about 200 km depth and it undergoes fractional melting. The relatively low pressure and the fractional melting process eliminates garnet from the source of these magmas. Barberton-type komatiites start to melt at greater depths – their source may have been molten as it transited the transition zone. These komatiites are formed by about 30% batch melting at depths greater than 200 km. (B) Hydrous conditions. (Phases relations from Asahara *et al.* (1998)). The position of the solidus is very uncertain but its location is less important than that of the 30% melting curve, which corresponds to the conditions under which Barberton-type komatiites form. Pressures greater than about 6GPa are needed to stabilize garnet in the residue of a 30% partial melt. In contrast, the subduction-zone komatiite of Grove *et al.* (1999) (shown as a star) forms at much shallower depths. Under these conditions garnet is not stable in the residue of fusion and the melt does not have the geochemical characteristics of Barberton komatiite.

that of many rocks in the upper crust. For komatiite to reach the surface and erupt, it must fill a continuous liquid column within rocks whose average density is greater than that of the komatiite itself. This would be the case when komatiite erupts in an oceanic setting where solidified basalt near the surface has a density similar to that of the komatiite liquid, and the cumulates or other intrusive rocks at deeper levels have higher densities. However, when komatiite traverses or erupts onto a granitic substrate, as in the Kambalda area in Western Australia, the high density of mantle rocks lower in the liquid column must counterbalance the low density of the granites.

We have very little idea how a komatiite behaves during eruption. The best analogue is probably the sheet flows of continental flood basalt sequences. On this basis, we can predict that komatiites probably erupted initially along fissures, as a series of lava fountains. The violence of this fountaining is difficult to judge. It will be enhanced by the low viscosity of the silicate liquid but mitigated by the high density. The primary control, however, is the volatile content in the komatiite magma, which probably is low in most komatiites.

We know from work in areas of good outcrop that once komatiites escape the vent they form highly mobile flows. The maximum length of a komatiite flow is unknown, our knowledge being limited by the quality and continuity of outcrop. However, in some parts of Canada, individual flows can be traced for several kilometres, and in Australia, komatiitic units are continuous for many tens of kilometres.

Thick, massive, olivine-rich units are present in most regions where komatiite lavas are abundant. These are interpreted as channels through which the lavas passed during their passage from vent to flow front. Flowage through these channels may have been very rapid and turbulent. Most of the komatiite flows preserved in greenstone belts represent lateral facies – small sheets or lobes of lava that spread out from a central feeder. These small pulses of lava may never have moved far from the feeder channel before ponding and crystallising beneath a thin elastic crust.

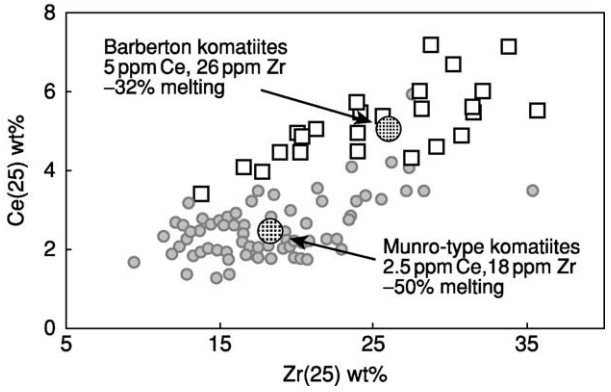


Figure 5 Concentrations of Ce and Zr in komatiites, normalized to 25% MgO by subtracting or adding olivine. The average compositions of the two groups are shown, together with calculated percentages of mantle melting. These were calculated assuming a primitive source for Barberton komatiites (1.6 ppm Ce and 9.7 ppm Zr) and a depleted source for Munro (1.3 ppm Ce and 9.7 ppm Zr). Residual mineralogy: ol – 50; opx – 30; cpx – 10; gt – 10. (Partition coefficients are from Green (1974).) (Principal sources of komatiite data are from Sun and Nesbitt (1978); Bickle *et al.* (1993); Jahn *et al.* (1983).)

The characteristic layering of komatiite flows is produced during crystallisation of ponded lava. Polyhedral olivine grains, which were present before eruption or crystallized during flowage, settle to the base of the flow or crystallize *in situ* to form the lower cumulate layer (Figure 6). At the same time, the spinifex-textured upper part of the flow crystallizes through downward growth of crystals from the crust of the flow. Faure *et al.*, showed in 2002 that the presence of a thermal gradient, such as exists at the margin of every flow and high-level intrusion, is instrumental in the formation of spinifex texture. The texture forms as a result of constrained, *in-situ* crystallization of olivine or pyroxene during moderately rapid cooling of low-viscosity ultramafic liquid.

Melting and Segregation of Komatiite Liquids

During mantle melting, if certain conditions are met, the silicate liquid segregates efficiently from its

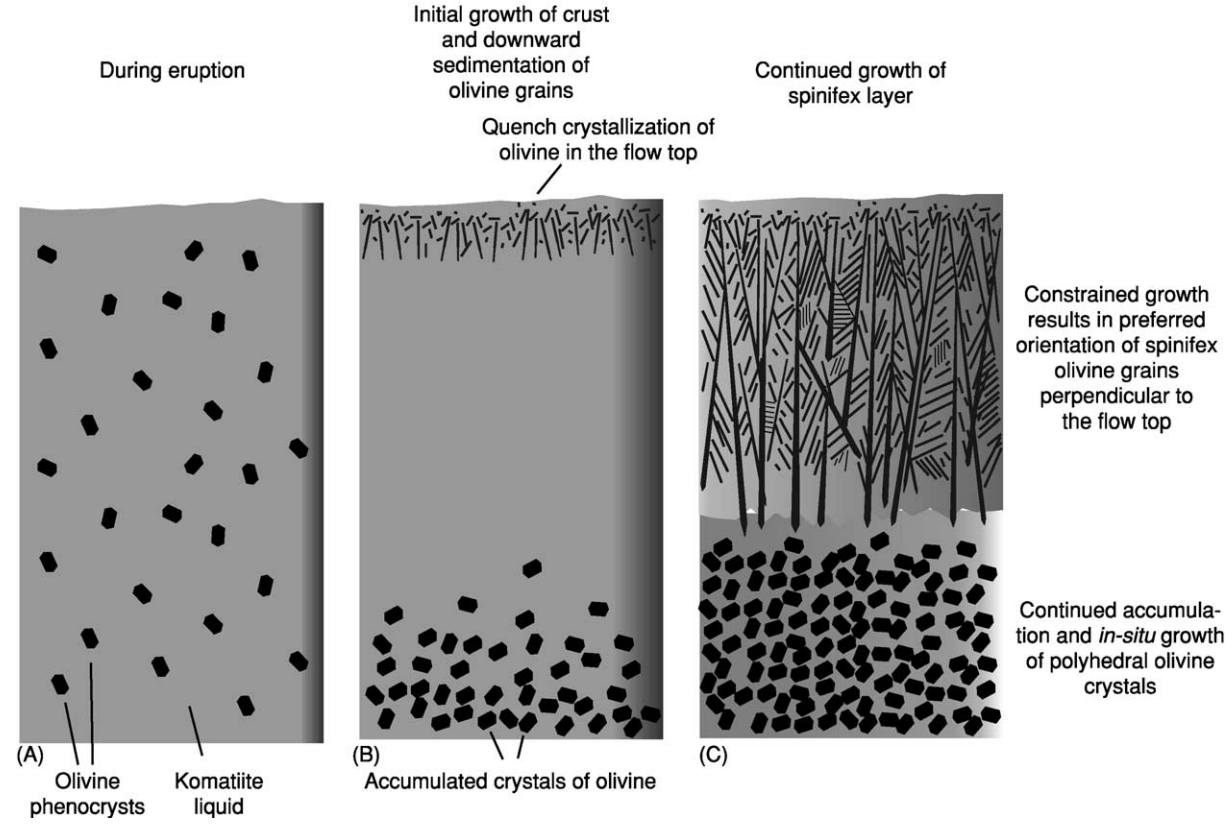


Figure 6 Diagram illustrating the crystallization of a layered spinifex-textured komatiite flow.

source, even at low degrees of melting. The conditions are: (i) the viscosity of the liquid is low; (ii) the solid matrix is deformable; and (iii) the density of the liquid is less than that of the unmelted residue. Although this is the normal situation for melting at low pressures where basaltic magmas are produced, it may not always be the case for komatiites. It has been shown through experimental studies that because silicate liquid is more compressible than solid silicate minerals, the density contrast between solid and liquid decreases as pressure increases. At pressures greater than about 8 GPa, which correspond to a mantle depth of about 250 km, the density of an ultramafic komatiite liquid exceeds that of olivine. It remains less, however, than that of garnet, the densest upper-mantle mineral. It has also been shown that at 8 GPa, komatiite magma forms through 30–50% melting, leaving a residue of olivine and majorite garnet. The density of the liquid is slightly less than that of the olivine-garnet residue. Although under static conditions there would be little impetus for the liquid to segregate from its source, mantle melting is normally due to adiabatic decompression in an ascending source, as in a mantle plume. As the source rises, the pressure decreases and the density difference between melt and solid increases. Eventually the density contrast becomes sufficiently large that komatiite of the Al-depleted Barberton type escapes from its source, probably in a single batch of high-degree mantle melt. This type of komatiite is probably one of the rare types of magma that forms through batch melting of the mantle Figure 7.

Al-undepleted or Munro-type komatiites lack the geochemical signature that signals melting in equilibrium with garnet. This does not necessarily require that the source was garnet-free; only that when the komatiite magma separated from its source, garnet was absent in the solid residue. Three processes can contribute to the elimination of garnet: low pressure, which destabilizes garnet; a high degree of melting, which eliminates low-temperature phases; and fractional melting, which preferentially removes the first-melting garnet-rich component from the source. In a rising mantle plume, all three processes may operate together.

Al-undepleted Munro-type komatiites are characterized by relative depletion of the more incompatible trace elements, such as the light rare-earth elements (LREE), Nb and Th. In some cases, notably for the Cretaceous Gorgona komatiites, the extent of LREE depletion varies widely within a suite of rocks of constant Nd isotopic composition. This pattern is a

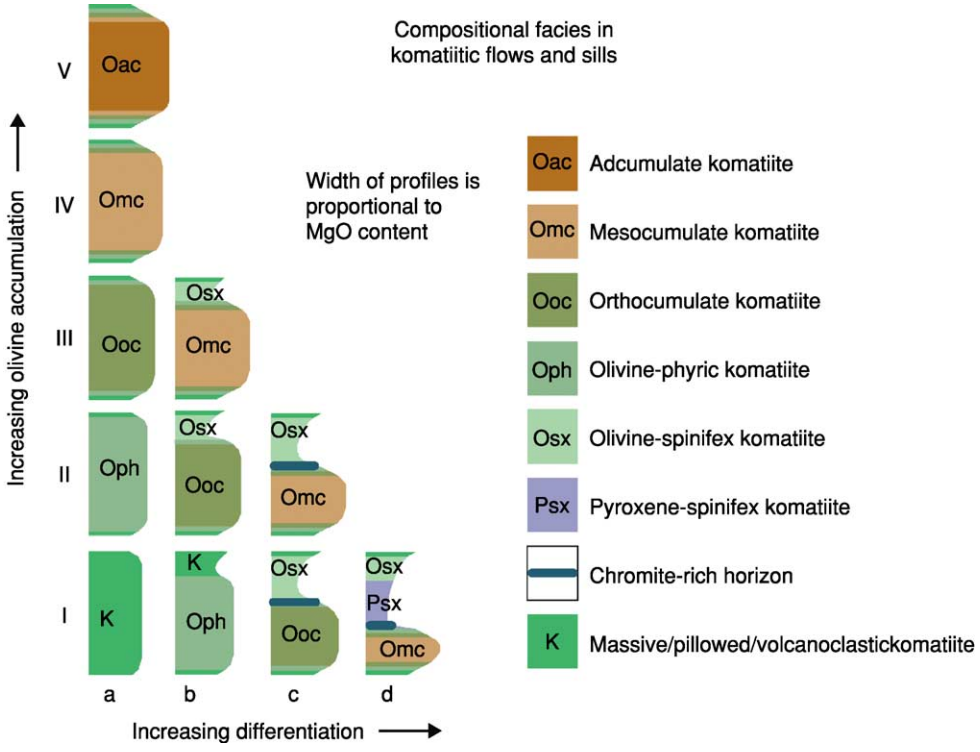


Figure 7 Classification of komatiitic flows and sills on the basis of the relative degrees of olivine enrichment and differentiation *in situ*. (Modified from Lesher *et al.* (1984).) UN, undifferentiated non-cumulate (massive, pillowed, or volcaniclastic); DN, differentiated non-cumulate; UC, undifferentiated cumulate; DC, differentiated cumulate; illustrating the crystallization of a layered spinifex-textured komatiite flow.

clear signature that the rocks formed through fractional melting. The source apparently started to melt at pressures lower than those that produced Barberton-type komatiites. Under these conditions, the komatiite liquid is less dense than the residual solid–perhaps the source initially contained a small water content, which entered the melt and lowered its density. This melt then escaped from the source, taking with it a high proportion of the incompatible elements or low-temperature components, including garnet, and the komatiites then formed through more advanced melting of the now-depleted source.

Wet Komatiites?

Grove and others and Parman and others have proposed that Barberton komatiite is a hydrous magma that forms by partial melting of metasomatized mantle above a subduction zone. In their model, the komatiite magma did not erupt as lava flows but crystallized in high-level sills. Arguments in support of the model are as follows.

i. The presence of water in the mantle source reduces the melting temperatures to 'reasonable' levels.
ii. The morphologies of olivine crystals in spinifex texture are best explained by crystallization in magmas that either contained a significant water content, or had recently exsolved water.
iii. The compositions of augite in spinifex-textured komatiites from the Barberton region resemble those of pyroxene in moderate-pressure hydrous experiments.
iv. Analyses of melt inclusions in chromite in komatiites from Zimbabwe suggest that the parental komatiite contained about 0.9% water.
v. Komatiites from the Boston Creek in Canada and in some other areas contain vesicles and minor amounts of magmatic amphibole.
vi. The trace-element compositions of some komatiitic basalts resemble those of modern boninites.

Many other geologists and geochemists believe that most komatiites are essentially anhydrous and that only some rare examples (excluding the Barberton komatiites) contain moderate water contents. The principal arguments are:

i. The Archaean mantle probably was hotter than the modern mantle because of higher heat production from more abundant radioactive elements, release of accretion energy and core formation. In this context, the ~1600°C liquidus of anhydrous komatiite is not 'unreasonably' hot.
ii. Spinifex texture has been reproduced in anhydrous experiments.
iii. Recent mapping has shown that Barberton komatiite, like komatiite in other regions, erupted on the surface as highly mobile, long-lived, non-explosive lava flows. This corresponds to the eruption behaviour of anhydrous magma.
iv. The extrusive setting of Barberton komatiite casts doubt on the interpretation of pyroxene compositions.
v. Melt inclusions in komatiitic olivine have low water contents.
vi. The trace-element compositions of most komatiites are very different from those of boninites.
vii. The major and trace-element contents of komatiites indicate that they formed at depths far below those of subduction zones (**Figure 4B**).

Ni-Cu-(PGE) Mineralization

Because they have relatively high Ni, Cu, and platinum-group element (PGE) contents and are capable of eroding S-rich crustal rocks, komatiites are capable of forming Ni-Cu-(PGE) sulphide deposits. The best-known examples, of Archaean age, are in the Kambalda region of Western Australia. These deposits are localized in the lower parts of thick dunitic units that are interpreted as lava channels. Komatiite lava flowing turbulently through the channels thermally eroded and melted S-rich floor rocks, leading to the segregation of Ni-Cu-PGE-rich immiscible sulphide liquids that accumulated at the base of the units to form the ore deposits (*see* **Mining Geology:** Magmatic Ores). Proterozoic deposits, such as those of the Cape Smith Belt in Canada, formed by a similar process within invasive lava channels.

See Also

Earth: Mantle. **Igneous Processes**. **Lava**. **Mining Geology:** Magmatic Ores.

Further Reading

Arndt NT (1982) Proterozoic spinifex-textured basalts of Gilmour Island, Hudson Bay. *Geological Survey of Canada* Paper 83–1A: 137–142.

Arndt NT (1994) Archean komatiites. In: Condie KC (ed.) *Archean Crustal Evolution*, pp. 11–44. Amsterdam: Elsevier.

Arndt NT and Nisbet EG (1982) What is a komatiite? In: Arndt NT and Nisbet EG (eds.) *Komatiites*, pp. 19–28. London: George Allen and Unwin.

Arndt NT and Nisbet EG (1982) *Komatiites*. London: George Allen and Unwin.

Asahara Y, Ohtani E, and Suzuki A (1998) Melting relations of hydrous and dry mantle compositions and the genesis of komatiites. *Geophysical Research Letters* 25: 2201–2204.

Bickle MJ, Ford CE, and Nisbet EG (1977) The petrogenesis of peridotitic komatiites; evidence from high-pressure melting experiments. *Earth Planetary Scientific Letters* 37: 97–106.

Echeverria LM (1980) Tertiary or Mesozoic komatiites from Gorgona Island, Colombia; field relations and geochemistry. *Contributions to Mineral Petrology* 73: 253–266.

Faure F, Arndt N, and Libourel G (2002) Crystallisation of plate spinifex texture at 1atm. pressure in a thermal gradient. *Geochim. Cosmochim. Acta* v. Goldschmidt Conference Abstracts, p. A225.

Green DH (1974) Genesis of Archaean peridotitic magmas and constraints on Archaean geothermal gradients and tectonics. *Geology* 3: 15–18.

Herzberg C (1999) Phase equilibrium constraints on the formation of cratonic mantle. In: Fei Y, Bertka CM, and Mysen BO (eds.) *Mantle Petrology: Field Observations and High-Pressure Experimentation,* pp. 13–46. Houston: The Geochemical Society.

Jahn BM, Gruau G, and Glickson AY (1982) Komatiites of the Onverwacht Group, South Africa: REE chemistry, Sm-Nd age and mantle evolution: *Contributions to Mineral Petrology* 80: 25–40.

Le Maitre RW, Bateman P, Dudek A, *et al.* (1989) *A Classification of Igneous Rocks and Glossary of Terms.* Oxford: Blackwell.

Lesher CM, Arndt NT, and Groves DI (1984) Genesis of komatiite-associated nickel sulphide deposits at Kambalda, Western Australia: A distal volcanic model. In: Buchanan DL and Jones MJ (eds.) *Sulphide Deposits in Mafic and Ultramafic Rocks,* pp. 70–80. London: Institution of Mining and Metallurgy.

Nesbitt RW, Jahn BM, and Purvis AC (1982) Komatiites: an early Precambrian phenomenon. *Journal of Volcanic Geothermal Research* 14: 31–45.

Nisbet EG (1982) The tectonic setting and petrogenesis of komatiites. In: Arndt NT and Nisbet EG (eds.) *Komatiites,* pp. 501–520. London: George Allen and Unwin.

Nisbet EG, Cheadle MJ, Arndt NT, and Bickle MJ (1993) Constraining the potential temperature of the Archaean mantle: a review of the evidence from komatiites. *Lithos* 30: 291–307.

Parman S, Dann J, Grove TL, and de Wit MJ (1997) Emplacement conditions of komatiite magmas from the 3.49 Ga Komati Formation, Barberton Greenstone Belt, South Africa. *Earth Planetary Scientific Letters* 150: 303–323.

Sun SS and Nesbitt RW (1978) Petrogenesis of Archean ultrabasic and basic volcanics: evidence from rare earth elements. *Contributions to Mineral Petrology* 65: 301–325.

Obsidian

G J H McCall, Cirencester, Gloucester, UK

Introduction

Obsidian is an extremely siliceous volcanic rock, found in lava flows and volcanic plugs, domes, and necks. It is closely related to rhyolite but has solidified from an extremely viscous magma as a glass with only minute crystallites within it. Like granite, it has a composition close to the eutectic, which explains the lack of crystallization (i.e., the glass solidified before crystals could form). Obsidian is of rare occurrence in volcanic suites, in which case it is mostly in calc-alkaline island-arc-type suites: this rarity is probably because of the fact that obsidian is the final residual product of differentiated magmatic suites, and the fact that the stickiness of the magma restricts it to protrusions such as domes and flows of small extent, which are squeezed out with difficulty. It has immense archaeological significance, having been used widely by ancient people for ornaments, arrowheads, knives, and scrapers. The Romans and Greeks quarried obsidian extensively for gemstones and the Aztecs mined it in the fourteenth through sixteenth centuries, presumably for ornaments. It is still listed as semiprecious gemstone, and the famous jeweller Peter Carl Fabergé (1846–1920) used it to create animal-shaped ornaments. Colour (black, brown, green, yellow, and red), translucency and transparency, reflectance, relative hardness, and sharp edges when fractured are all qualities that have made obsidian desirable through the ages.

Historical

The term 'obsidian' is a very ancient word for natural glasses. In the first century AD, Pliny the Younger wrote: "Among the various kinds of glass we may also reckon obsidian, a substance very similar to the stone which Obsidius discovered in Ethiopia. The stone is a very dark colour and somewhat transparent, but it is dull to the sight, and reflects, when attached as mirror to walls, the shadow of the object rather than the image". Thus the origin of the term clearly goes back to the Romans, and the first record was in Ethiopia. That the name of the rock comes from that of a person is, however, doubtful. John Hill in 1740 wrote, based on Theophrastus, a Greek philosopher who wrote extensively on plants, stones, and climatic

topics in about 300 BC, that "the antients had two or three of these dark marbles, of fine texture, of great use amongst them. They took a polish, were transparent to some degree when cut into thin plates, and reflected the image as our looking glasses do. The first kind was called Οψιανος απο της οψεος, which expressed its property of reflectivity and was afterwards written in the Latin as opsidianus or obsidianus".

This early derivation of the name does seem to agree better with the known fact that Greeks and Romans used obsidian as a gemstone and obtained it from the Island of Melos in the Aegean, where quarries have yielded it for 12 000 years. In 1773, the German mineralogist UFB Bruckmann wrote that obsidian was probably a black lava and geologist Leopold von Buch in 1809 noted that it flowed out, and was not cast out, from volcanoes. In 1822, the American geologist Parker Cleaveland wrote: "This variety has a strong resemblance to glass. Its fracture is distinctly conchoidal, with large cavities and strongly shining with a lustre more or less vitreous. The surface of the fracture often exhibits a striated or wavy appearance, and its appearance is a little unctuous. It scratches glass, gives fire with steel, but is brittle, and falls into sharp-edged fragments. Most commonly it is translucent at the edges, or opaque, but some varieties are translucent or in thin scales transparent. Its colour is black, either deep or pure, or tinged with brown, green, blue or grey, and sometimes passes to blue, green, brown or gray, even yellow or red. The darkest colours often discover a tinge of green by translucent light".

Composition

The Norwegian geologist and petrologist JHL Vogt in 1923 wrote that "compositions of eutectic or nearly eutectic proportions promote the formation of glass, since the eutectic has the lowest melting point; consequently, at that temperature the melt is more viscous than elsewhere on the curve, and points near the eutectic tend to reach solidifying point before reaching the crystallizing point. With relatively quick cooling the crystallization will be entirely or nearly restrained. Thus it is no accidental circumstance that by far the most obsidians have nearly the chemical composition of the granitic eutectite". As now used, the term 'obsidian' is applied to massive, usually dense, but often slaggy glasses of deep brown or black, grey, red, or mottled red and black colour. The viscosity of obsidian as a flow stems from branching and tangled chains of tetrahedral silicon and aluminium combined with oxygen. When solidified, obsidian is quite hard and its conchoidal fracture results in sharp, even cutting edges to the brown fragments. In many cases, the rock is spotted or banded. Spherulites and lithophysae occur in some obsidians, and may be abundant, also concentrated in certain layers. Normally obsidians are natural glasses of rhyolites, but any acid (siliceous) volcanic rock may solidify as similar glass by rapid cooling, and thus the terms 'trachyte' and 'dacite-obsidian' in common use, though strictly obsidians are of rhyolite composition.

The specific gravity of obsidian ranges from 2.30 to 2.58. The refractive index ranges from $n = 1.48$ to 1.53. The hardness on Moh's scale ranges from 5.6 to 7. The chemical compositions of various obsidians are given in Table 1; also shown in the lower part of the table are the CIPW norms (named after the petrologists Cross, Iddings, Pirsson, and Washington, in 1931). A norm is a means of converting a chemical composition of an igneous rock to an ideal mineral composition. In this way, similarities in rocks with contrasting mineral assemblages can be noted. Some of the factors considered are temperature, pressure, and mineral content; in the CIPW norm calculation, the magma is considered to be anhydrous and at low pressure.

Chemically, obsidian has a low water content, but even so, this is an order or more greater than is the case for tektites, which resemble obsidian and were once referred to as obsidianites. For example, water content of moldavites from Central Europe ranges from 0.006 to 0.010. Tektites also contain lechatelierite, an amorphous form of quartz that is never found in volcanic glasses (*see* **Tektites**).

Occurrences Worldwide

Obsidian Cliff, Yellowstone National Park

Obsidian Cliff in Yellowstone National Park (Wyoming, USA) is considered a typical occurrence. The chemical compositions of red and black obsidian samples from the site are given in Table 1. The composition is rhyolitic. The cliff forms a giant flow 120–160 m thick. The rock is locally columnar, and at the lower part is traversed by bands or layers of small grey spherulites, but cavities or lithophysae are almost absent. Higher up, the obsidian is less massive and contains large lithophysae (concentric shells of flattened fine material with a central cavity) parallel to the plane of flow.

Eolian Islands

Three recent obsidian flows from the island of Lipari have been described, being the youngest (from the sixth to eighth centuries AD) eruptives on the island (Figure 1). The Rocche Rosse flow is of obsidian

Table 1 Chemical analyses of obsidians

	Sample[a]						
Mineral/norms	*1*	*2*	*3*	*4*	*5*	*6*	*7*
SiO_2	75.52	74.70	76.20	75.23	74.37	74.05	73.84
TiO_2	—	—	—	—	—	—	0.14
Al_2O_3	14.11	13.72	13.17	12.36	12.65	14.67	13.00
Fe_2O_3	1.74	1.01	0.34	0.96	2.58	0.89	1.82
FeO	0.08	0.62	0.73	1.24	n.d	n.d	0.79
MnO	—	—	0.10	—	—	—	0.07
MgO	0.10	0.14	0.19	0	0.20	0.26	0.49
CaO	0.78	0.78	0.42	1.00	1.22	0.97	1.52
Na_2O	3.92	3.90	4.31	4.00	3.87	3.99	3.82
K_2O	3.63	4.02	4.42	4.62	4.57	5.11	3.92
H_2O	0.29	0.62	0.33	0.73	0.24	0.91	0.53
P_2O_5	—	—	—	0.27	—	—	0.01
FeS_2	0.11	0.40	—	—	—	—	0.02
Total	100.38	99.91	100.25	100.42	99.79	99.85	99.97
CIPW norms[b]							
Q	36.72	34.86	32.40	32.76	29.88	26.94	
Or	21.68	23.35	26.69	27.24	27.24	30.98	
Ab	33.01	33.01	36.16	33.54	33.01	32.54	
An	3.80	3.89	1.95	2.22	3.34	5.00	
C	2.24	1.53	0.51	—	—	0.71	
Di	—	—	—	0.75	2.20	—	
Hy	0.30	0.66	1.69	1.06	3.57	2.28	
Rut	0.23	1.39	0.46	1.39	—	—	
Hm	1.00	—	—	—	—	—	
Ap	—	—	—	0.67	—	—	

[a]n.d., Not determined. Samples: 1, red obsidian, Obsidian Cliff, Yellowstone National Park; 2, black obsidian, Obsidian Cliff, Yellowstone National Park; 3, obsidian, Obsidian Hill, Tewan Mountains, New Mexico; 4, obsidian, Cerro de Los Navajos, Mexico; 5, obsidian, Forgia Vecchia, Lipari; 6, obsidian, Tenerife; 7, average of 44 obsidians.
[b]See text for discussion of CIPW norms.

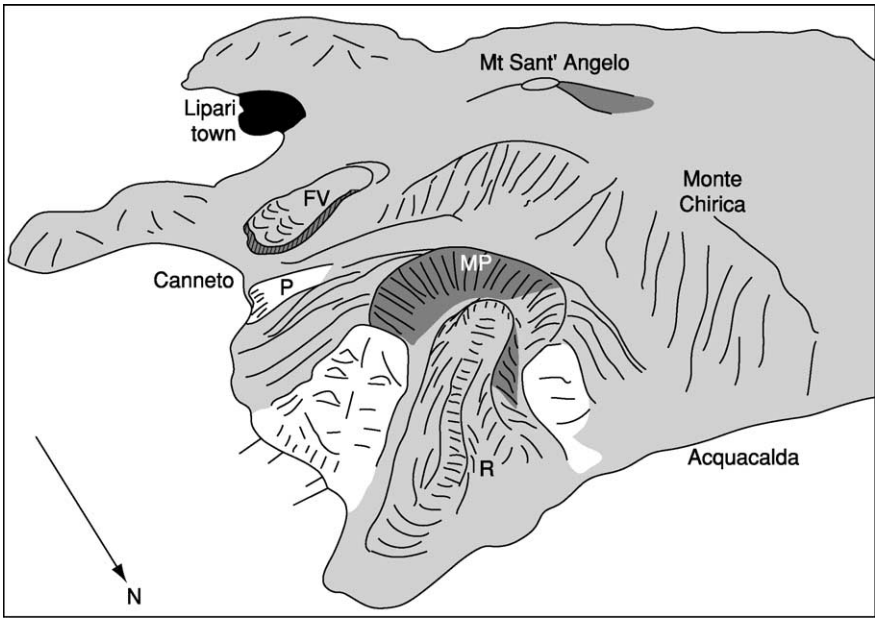


Figure 1 The island of Lipari, north of Sicily, showing the obsidian flows dating to the sixth to eighth centuries AD (R, Rocche Rosse; FV, Forgia Vecchia; MP, Monte Pilato; P, Pomiciazzo), the most recent eruptives on the island. Reproduced with permission from Guest JE, Cole PD, Duncan AM, and Chester DK (2003) *Volcanoes of Southern Italy.* Bath: Geological Society Publishing House.

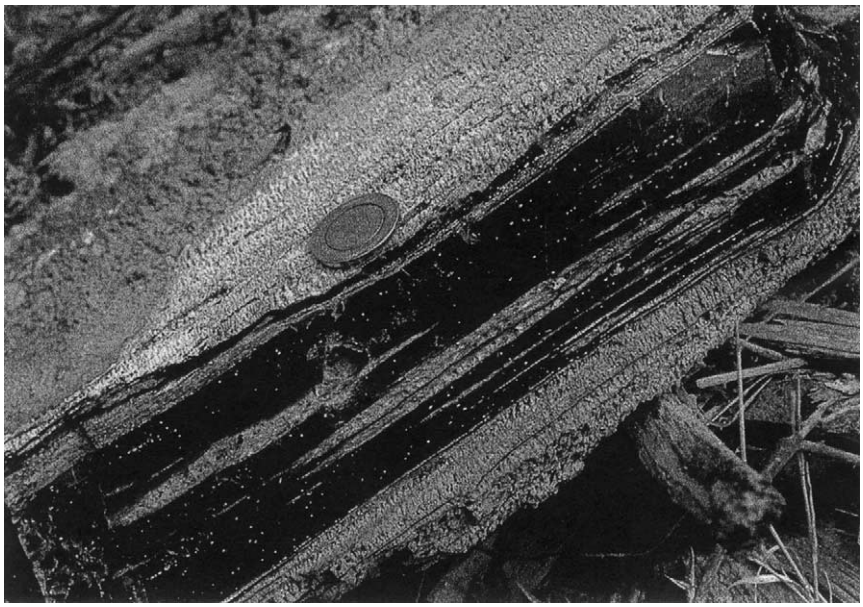


Figure 2 Close-up of the Rocche Rosse obsidian flow, Lipari, showing alternations of fresh and devitrified glass layers. Reproduced with permission from Guest JE, Cole PD, Duncan AM, and Chester DK (2003) *Volcanoes of Southern Italy*. Bath: Geological Society Publishing House.

block lava: the surface has arcuate surface ridges (ogives) and talus margins. The flow is several tens of metres thick and is typical of high-viscosity silicic lavas. There are devitrified layers and lenses alternating with the black obsidian (Figure 2). There are also obsidian flows on the nearby island of Vulcano.

Armenia

The Quaternary Gutansar volcano in Armenia shows a progression from basalts and andesites, through liparites, transitional liparite obsidians, to massive obsidian flows. There is a fissure extrusion of black obsidian at the base, changing gradually to brownish and greyish slightly swelled glass, then to perlite and pumice. The obsidian extrusion has a fanlike structure (Figure 3). The distribution of the obsidian bodies indicates very shallow magma chambers. Such viscous magmas increase in volume and heave up continuously to the surface with regular accelerations, the process being one of tectovolcanic displacements.

Newberry Caldera, Oregon

Newberry Caldera in the Oregon Cascades (USA), 8 km in diameter, lies atop of a volcano initiated in the Pliocene, but its main activity occurred during the Pleistocene and the last eruptions occurred just over 2000 years ago. The obsidian flows lie within the caldera and are among the most recent eruptives. They present expanses of barren, shiny black obsidian, almost devoid of vegetation, and surfaces sharp as razors can cut the legs of the unwary geologist. There are four separate steep-sided flows of glistening black obsidian. The Big Obsidian Flow (Figure 4) covers 1.5 km^2 and has its own dome of pumiceous obsidian plugging the vent. Two smaller flows straddle a north-east-trending fissure from which they were erupted. Another small flow was erupted from a vent on the caldera wall. This development at Newberry Caldera is undoubtedly the most spectacular development of obsidian flows known, but the terrain is extremely difficult inside the caldera.

California

Obsidian in the central eastern Sierra Nevada mountains of California, USA, at Mono Craters, occurs in a flow of viscous lava, as lenses and bands interleaved with rhyolite (Figure 5). The obsidian is very regularly banded (Figure 6). Two other spectacular obsidian flows occur further north, at Glass Mountain, California (Figure 7), in the Medicine Lake region of the Cascade Range, overlooking the Nevada and Oregon borders. These flows are less than 1000 years old and are associated with small obsidian domes. They display remarkable flow structures, of curved pressure ridges akin to those on the surface of glaciers.

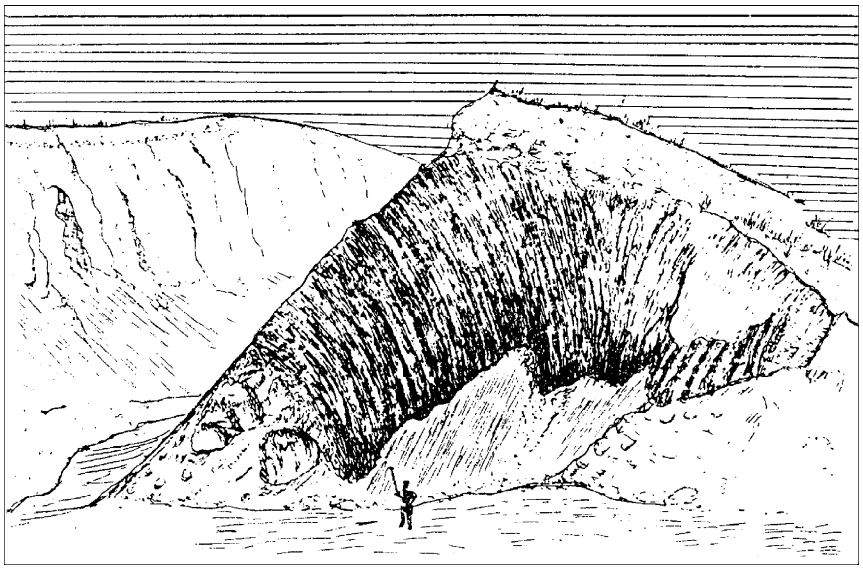


Figure 3 Fan-like fluidal structure in an obsidian extrusion, Gutansar volcano, Armenia. Reproduced from Mkrtchian SR, Paffengoltz KN, Shirinian KG, Karepetian KI, and Karepetian KG (1971) *Late Orogenic Acid Volcanism of Armenian SSR*. Yerevan: Publishing House of the Academy of Sciences of the Armenian SSR.

Figure 4 'Big Obsidian Flow', the Newberry Caldera, Oregon, showing also the obsidian dome from which it emanates. The flow is about 5 km long from source to front. Note the corrugations, or 'corda', which are pressure ridges as on a glacier, each of an amplitude of ~5 m. Reproduced with permission from Green J and Short NM (1971) *Volcanic Landforms and Surface Features: A Photographic Atlas and Glossary*. New York, Heidelberg, and Berlin: Springer-Verlag.

Obsidian as a Semiprecious Stone

The physical attributes of obsidian might seem to make it an unlikely material for ornamental use, but it is in fact classified as a semiprecious stone (Figure 8). The small carved rhinoceros in Figure 8 is an example of work by Fabergé. Obsidian has been described and illustrated by Bauer and Bouska, who noted that it is a product of rapid solidification of molten magma and is amorphous and without cleavage. These authors mentioned the large obsidian formations in the Eolian

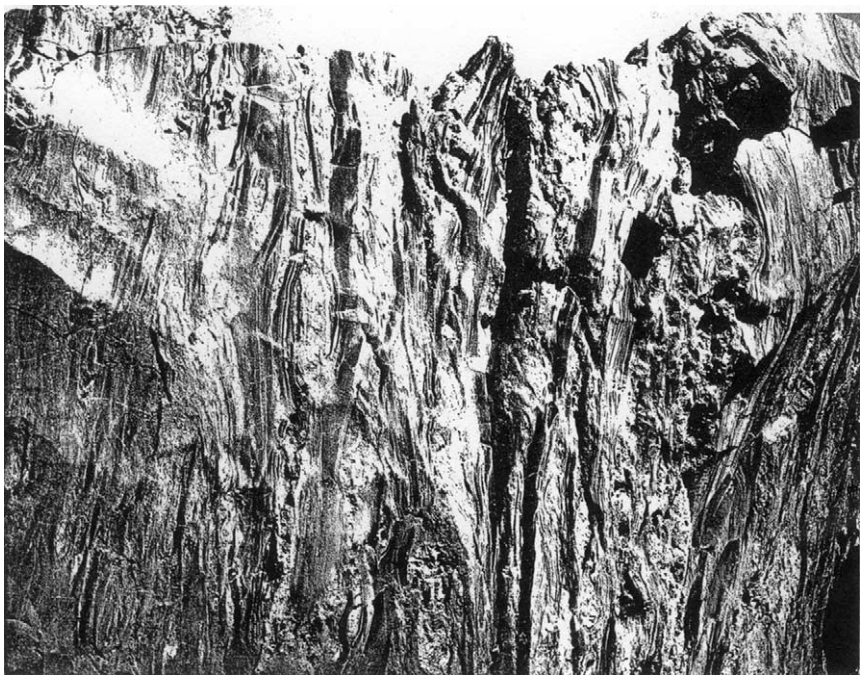


Figure 5 Banded rhyolite and obsidian (black), Panum Crater, Mono Craters, California (height of face is 1.5 m). Reproduced with permission from Green J and Short NM (1971) *Volcanic Landforms and Surface Features: A Photographic Atlas and Glossary*. New York, Heidelberg, and Berlin: Springer-Verlag.

islands, especially Lipari, and those in Oregon, Arizona, Utah, Hungary, Slovakia, the Caucasus, the Urals, and New Zealand. Most obsidians reveal abundant minute crystallites under the microscope. Obsidians from Mexico, the United States, and the Lake Sevan area of Armenia display an outstanding sheen and silky lustre, due to the presence of such crystallites. The pre-Colombian Indians of Central America, besides fashioning knives, scrapers, and spearpoints from it, used it for ornamentation. The Mexican Indians have long been famous for fashioning decorative objects and amulets, necklaces, and bracelets of obsidian (jet is used in a similar way around Whitby, Yorkshire; it is not obsidian, but is lignitic and is believed to be derived from water-logged driftwood in shales).

As a gemstone, obsidian is cut in various ways. Perfectly transparent varieties are faceted, mainly into steep cuts, but brilliants are quite common. Materials of lower transparency are faceted into cabochons, or table-cuts. Large irregular cabochons are commonly cut from Mexican or North American obsidian and contain tiny crystallites of the tetragonal polymorph of silica, cristobalite. A very attractive obsidian with grey, eye-like spots in the red, haematite-coloured groundmass comes from Gyumishkoe in Armenia. Pitchstone is more siliceous than obsidian, and lacks its lustre, and tachylite also resembles it, but is basaltic in composition.

Obsidian Artefacts

Obsidian is one of the most useful materials known to archaeology. Towards the end of the Stone Age, humans increasingly sought obsidian for making cutting tools. Skilled artisans made knives and daggers and ornamental bowls and polished mirrors; even the making of less sophisticated tools such as scrapers and arrowheads required expertise in chipping off flakes. Such tools appear abundantly at archaeological sites in Europe and Asia, in South, Central, and North America, and in the Pacific Islands. Obsidian is rare in Australia, though obsidian artefacts have been found in Queensland. It is mainly found in quite recent island-arc volcanic belts, and these do not traverse Africa, though the original obsidian find in Roman times was reportedly in Ethiopia. Because obsidian hydrates rapidly, most fresh obsidian is less than 10 million years old.

Archaeological Tracing Methods

Obsidian served early humans well for making cutting tools, due to its relative hardness and the manner in which it fractures into pieces of sharp-edged

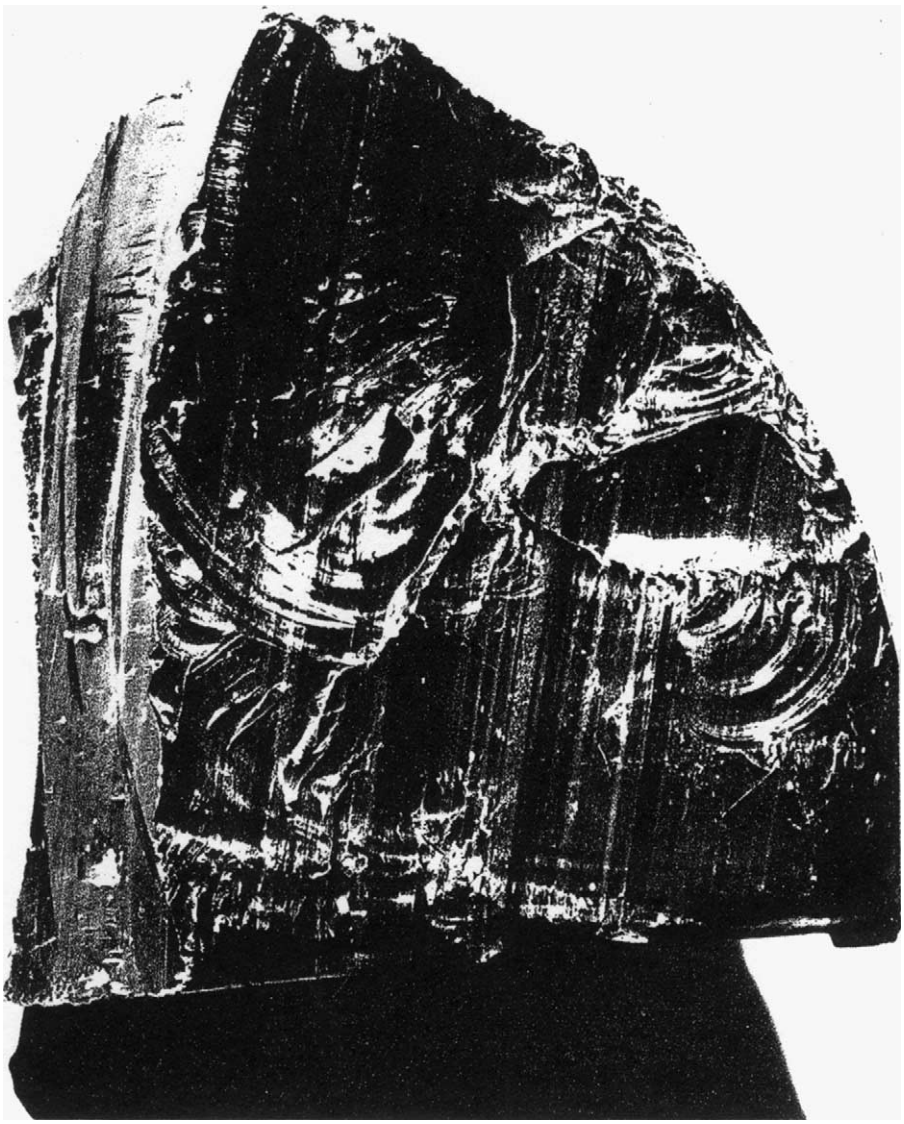


Figure 6 Banded obsidian, Mono Craters, California (maximum dimension of specimen, 12 cm). Note the conchoidal fracture. Reproduced with permission from Green J and Short NM (1971) *Volcanic Landforms and Surface Features: A Photographic Atlas and Glossary*. New York, Heidelberg, and Berlin: Springer-Verlag.

wedges. Because it is isotropic, there is no preferred direction for striking off flakes. The age of the artefacts, as distinct from the glass of which they are made, can be determined by radiocarbon dating or other means such as stratigraphic position or associated potsherds. The hydration state of the obsidian can be used for dating the glass, because the rock hydrates at a steady rate with the passage of time, though this would seem to be a method liable to error. By any means, dating of obsidian artefacts has helped to delineate ancient trade routes extending for thousands of kilometres.

Newly developed geochemical techniques have made it possible to trace the source of the glass utilized in artefacts. The aim of these tracing techniques is to trace the obsidian from the archaeological sites back to the parent lava flow, by geochemical matching. Minor and trace elements that have a wide compositional range variation in obsidians are utilized, especially manganese and sodium. Such variations are due to the trace element composition differences in the parent rocks, which were partially melted in the mantle or crust to produce the magma. Variations also derive from changes due to crystals forming on the way up to the surface (e.g., in the magma chamber); partition of minor or trace elements is imposed between crystals, which separate off and melt. Successive flows from the same magma chamber are erupted

Figure 7 Two obsidian flows and several obsidian domes, Glass Mountain, California. The northern of the two flows emanates from an obsidian summit dome with a minute axial orifice for gas escape, and there is a succession of very small pancake-like effusions from a chain of centres in the upper left-hand corner of the photograph. The pressure structures on the flows, their glassy nature, and their tongue-like form, with steep sides and fronts on account of their viscosity, are features clearly illustrated. Reproduced with permission from Green J and Short NM (1971) *Volcanic Landforms and Surface Features: A Photographic Atlas and Glossary*. New York, Heidelberg, and Berlin: Springer-Verlag.

at different times in the magma chamber crystallization, and so they too will show differences in minor or trace element ratios. Each obsidian flow thus has a unique minor and trace element 'fingerprint'. Geochemical matching begins with obtaining a number of different minor and trace element values for each source flow, as a baseline. Having identified the minor trace element values of the likely source flow, the geochemist can with confidence match it with obsidian artefacts from archaeological sites and delineate the trade routes. In J Glover's *Geological Journeys*, three cases were described in which this has been done successfully.

In various northern and central locations of what is now the United States are found the Hopewell sites, which are remarkable burial sites built by the ancestors of Native Americans between 100 BC and AD 500. Obsidian artefacts are found there, but there are no developments of the volcanic rock nearby. Possible sources on the Pacific Coast or in Alaska, Yellowstone National Park, New Mexico, and Mexico have been investigated by JB Griffin and colleagues. Neutron activation analyses confirmed two groups with Na/Mn ratios of 150 and 90. These results fingerprinted two flows at Yellowstone National Park in Colorado. Figure 9 shows the path over which the obsidian had to travel, over 2400 km eastward, to the Hopewell sites. The process may have involved intermediate stages of bartering across this route.

Artefacts in the south-west Pacific have also been analysed. Emission spectroscopy analysis results on flakes from the Santa Cruz Islands, part of the nation of the Solomon Islands (one of the most northern group of the chain previously called the New Hebrides, now Vanuatu), show that the chemistry matches obsidians cropping out on New Britain, not the basaltic volcanic rocks of Santa Cruz. Radioacarbon dating shows that the flakes were deposited in Santa Cruz about 1000 BC, so the population of the islands 3000 years ago must have carried obsidian in small craft by sea east–south-eastwards over at least 2000 km (Figure 10). Work in 1996 by R Service has shown, further, that obsidian from New Britain and the Admiralty Islands (a group close to New Britain) was transported earlier to a

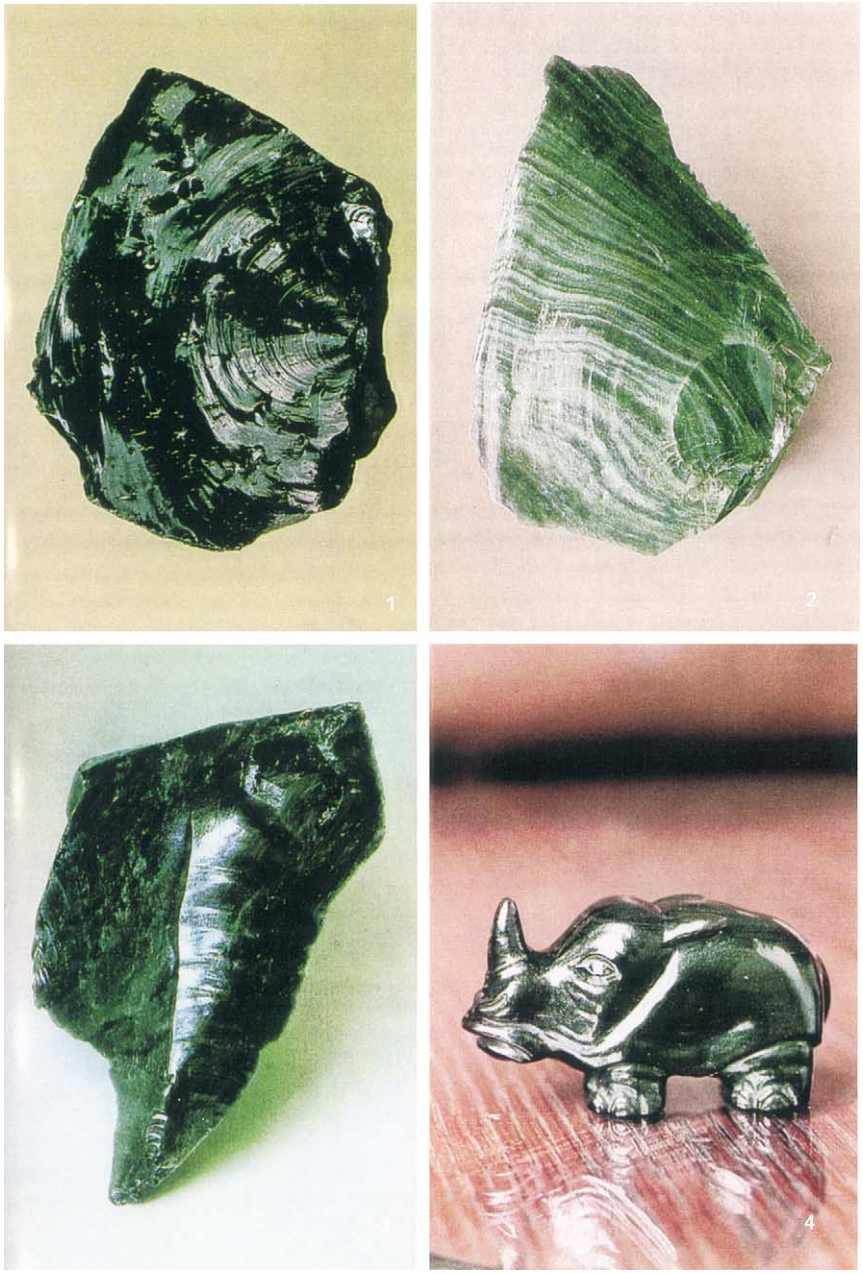


Figure 8 Gem-quality obsidian: (1) black, massive obsidian from the Caucasus (60 × 45 mm); (2) obsidian with a silvery, silky sheen, Yerevan, Armenia (120 × 80 mm); (3) Neolithic tool made of obsidian, Cejkov, Slovakia (140 × 100 mm); (4) rhinoceros figurine by Carl Fabergé, Russia (length, 60 mm). Reproduced with permission from Bauer J and Bouska V (1983) *Precious and Semi-Precious Stones.* London: Octopus.

6000-year-old site in northern Borneo, in the reverse direction (Figure 10).

An obsidian trade in the Middle East and Near East has been described by O Williams-Thorpe. About 8000 BC, humans in the Fertile Crescent of the Middle East, extending from the Lebanon to Iraq, for the first time left the hunter–gatherer mode of life, instead domesticating animals and growing cereals and vines – this has been called the Neolithic revolution. Obsidian was being traded extensively through this region during this period, presumably mostly for cutting and scraping tools, and here again the movements have been geochemically traced.

One further source of artefacts has been quarries and mines. Quarries were used as a source of obsidian by the Romans on Melos, and in Mexico, on a volcanic peak near Veracruz, the Aztecs mined obsidian from AD 1350 to 1520. The obsidian was removed

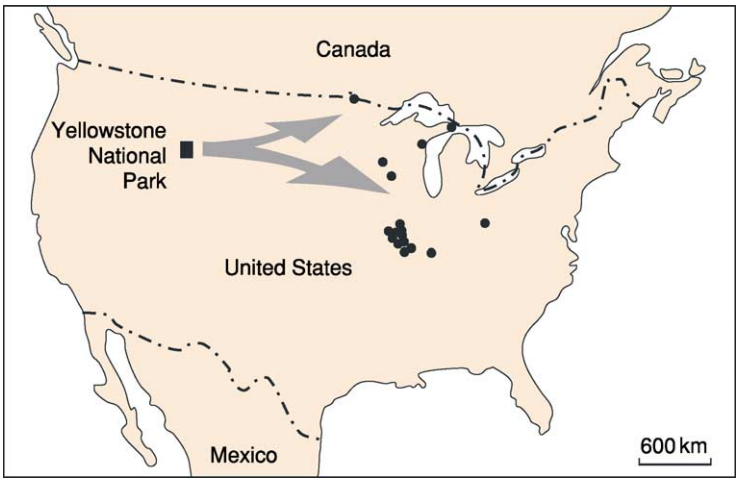


Figure 9 The path of transport of obsidian from Yellowstone, Colorado, to the Hopewell Indian sites. Reproduced with permission from Glover J (2003) *Geological Journeys: From Artifacts to Zircon.* Perth, Western Australia Division: Geological Society of Australia.

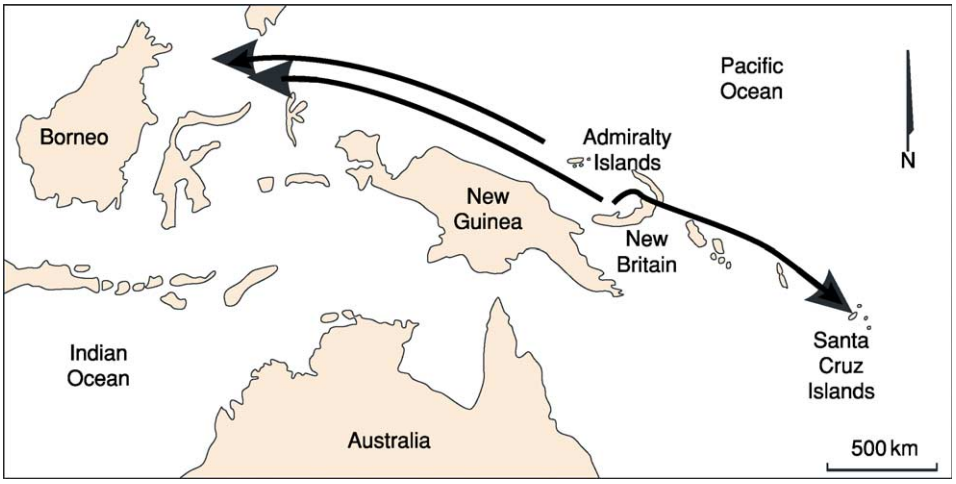


Figure 10 The trade routes for obsidian from New Britain and the Admiralty Islands to Santa Cruz Islands and Borneo, respectively. Reproduced with permission from Glover J (2003) *Geological Journeys: From Artifacts to Zircon.* Perth, Western Australia Division: Geological Society of Australia.

by driving passageways 70 m into the rock. Using levers of sharpened pine poles, the Aztecs exploited weaknesses and fractures in the obsidian rock, even using hammer stones of obsidian.

See Also

Earth: Mantle; Crust. **Geoarchaeology**. **Lava**. **Plate Tectonics**. **Volcanoes**.

Further Reading

Ambrose WR and Green RC (1972) First millennium BC transport of obsidian from New Britain to the Solomon Islands. *Nature* 237: 31.

Bauer J and Bouska V (1983) *Precious and Semi-Precious Stones*. London: Octopus.

David B, Bird R, Fullagar R, and Little L (1993) Glassy obsidian artifacts from North Queensland: the Nolan's Creek source and some archaeological occurrences. *The Artifact* 15: 25–30.

Dixon J, Cann JR, and Renfrew C (1968) Obsidian and the origins of trade. *Scientific American* 218(3): 38–46.

Glover J (2003) *Geological Journeys: From Artifacts to Zircon.* Perth, Western Australia Division: Geological Society of Australia.

Green J and Short NM (1971) *Volcanic Landforms and Surface Features: A Photographic Atlas and Glossary.* New York, Heidelberg, and Berlin: Springer-Verlag.

Griffin JB, Gordus AA, and Wright GA (1969) The identification of sources of Hopewellian deposits in the Middle West. *American Antiquity* 34: 1–14.

Guest JE, Cole PD, Duncan AM, and Chester DK (2003) *Volcanoes of Southern Italy*. Bath: Geological Society Publishing House.

Johannsen A (1952) *A Descriptive Petrology of Igneous Rocks: Volume 2, The Quartz-Bearing Rocks*. Chicago: University of Chicago Press.

Mkrtchian SR, Paffengoltz KN, Shirinian KG, Karepetian KI, and Karepetian KG (1971) *Late Orogenic Acid Volcanism of Armenian SSR*. Yerevan: Publishing House of the Academy of Sciences of the Armenian SSR.

Peterson NV and Grob EA (1965) *State of Oregon Lunar Geological Field Conference Guide Book*. Corvallis: University of Oregon Department of Geology and New York Academy of Sciences.

Service RF (1996) Rock chemistry traces ancient traders. *Science* 274: 212–213.

Torrence R (1986) *Production and Exchange of Stone Tools: Prehistoric Obsidian in the Aegean. New Studies in Archaeology*. Cambridge: Cambridge University Press.

Williams-Thorpe O (1995) Review article: obsidian in the Mediterranean and Near East: a provenancing access story. *Archaeometry* 37: 215–238.

IMPACT STRUCTURES

R A F Grieve, Natural Resources Canada, Ottawa, ON, Canada

Introduction

The systematic study of the terrestrial impact record is a recent endeavour in geology. The first terrestrial impact structure was recognized in 1906. By the late 1930s, the number of terrestrial impact structures had risen to about 20 but there was considerable controversy over their origin. The recognition, however, of diagnostic mineralogical indicators of, so-called, shock metamorphism established reliable criteria for the occurrence of extreme transient and dynamic pressures in the geologic environment and, hence, the occurrence of an impact event. These criteria have been developed further by observation and experiment over the intervening years and have stood the test of time and new observations. Terrestrial impact structures are essentially the sole source of ground truth data on large-scale natural impact events, with respect to the character and spatial distribution of impact-related lithologies and structure, particularly in the third dimension. Although the terrestrial impact record serves as an analogue for impact cratering processes on the other terrestrial planets, impact was not generally regarded, until recently, as a process of importance to the evolution of the Earth. This changed dramatically with the evidence for the occurrence of a major impact at the Cretaceous–Tertiary (K–T) boundary 65 My ago. Although originally contentious to many in the larger geoscience community, the involvement of a major impact at the K–T boundary and the related mass-extinction event in the biosphere is now generally accepted.

General Character of the Record

The terrestrial impact record is incomplete. It is characterized by inherent biases, largely related to the high level of endogenic geological activity of the Earth. At the time of writing, the known impact record consists of approximately 170 known impact structures or crater fields, and some dozen impact events, which are recorded in the stratigraphic record, some of which are related to known impact structures: for example, the modavite tektites with the Ries impact structure, Germany and the Ivory Coast microtektites with the Bosumtwi impact structure. Some of these tektite–microtektite strewn field cover a considerable area of the Earth's surface, e.g., the Australasian strewn field covers an area in excess of 50×10^6 km^2.

Due to the effects of erosion, the terrestrial impact record contains a mixture of topographic forms, and it is more appropriate to use the generic term impact structures than impact craters, which by definition require a negative topographic form. The discovery rate of new impact structures is about five per year, with the rate of discovery having increased substantially in the past three decades (**Figure 1**). A current listing of known terrestrial impact structures can be found at http://www.unb.ca/passc/ImpactDatabase.

The spatial distribution of known terrestrial impact structures is biased towards the stable interior or cratonic areas of the North American, Australian, and European continents (**Figure 2**). These are areas with low rates of erosional and tectonic activity, over extended periods of geological time. They are, thus, the best available surfaces for the preservation of

impact structures in the terrestrial geological environment. They are also areas where there have been active programmes to search for and study impact structures.

Erosion and sedimentation result in characteristics of the terrestrial impact record that are not as dominant on other planetary bodies. For example, approximately one-third of known impact structures are buried by post-impact sediments. Most buried impact structures were detected initially as geophysical anomalies and later drilled, for economic or scientific purposes, which confirmed their impact origin. Although a few submerged impact structures are known, they occur on the relatively shallow, continental shelves. No impact structures are known from the true ocean floors. This reflects the relatively young age and the generally poor resolution of geological knowledge of the ocean floors.

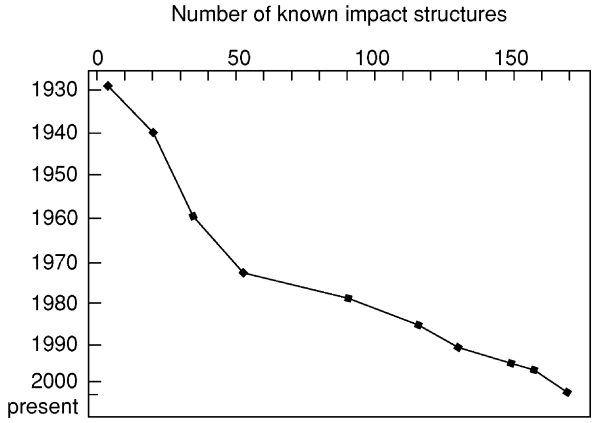


Figure 1 Number of known terrestrial impact structures recognized with time. Note the increase in the rate of discovery after the 1970s, due to the establishment of shock metamorphism (see text for details) as a reliable criterion for identifying the impact origin of specific structures.

There are also biases in the ages and sizes of known terrestrial impact structures. The majority are <200 My old. This reflects problems of preservation and, to a lesser extent, recognition in the highly active geological environment of the Earth. At larger diameters, the cumulative size–frequency distribution is similar to the production distribution observed on the other terrestrial planets. In the terrestrial case, however, this distribution more likely represents a steady-state condition between the formation and removal of impact structures from the record. At diameters below approximately 20 km, the cumulative size–frequency falls off, with an increasing deficit of structures at smaller diameters. This drop-off is an inherent property of the terrestrial record, as it has remained even with the addition of new structures to the known record over the past 30 years. The deficit of small craters is due to atmospheric crushing of smaller impacting bodies, the relative ease with which smaller structures can be buried or eroded, and the intrinsic difficulty in recognizing smaller structures.

Morphology

Small impact structures have the form of a bowl-shaped depression, with an upraised rim, and are known as simple craters (Figure 3). The rim, walls, and floor define the so-called apparent crater. At the

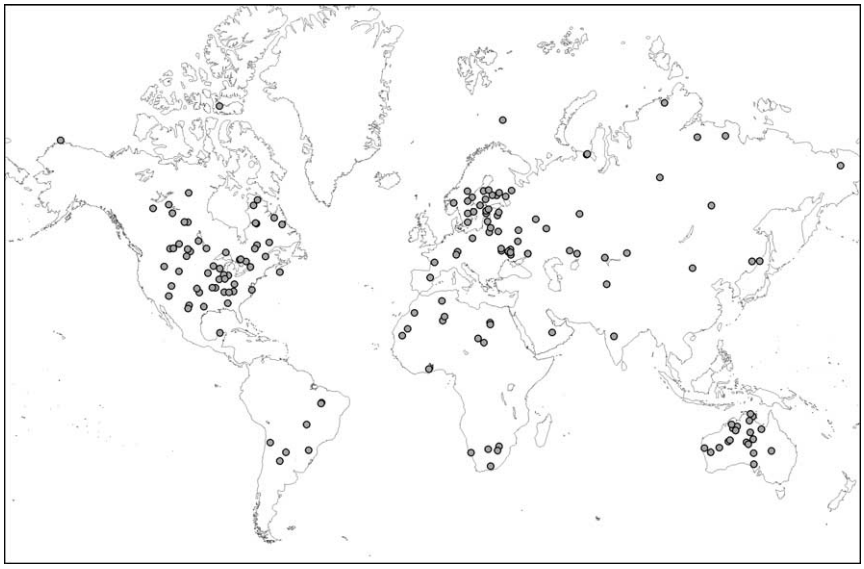


Figure 2 Spatial distribution of known terrestrial impact structures. Note current concentrations in the stable interiors of North American, Australian, and European continents.

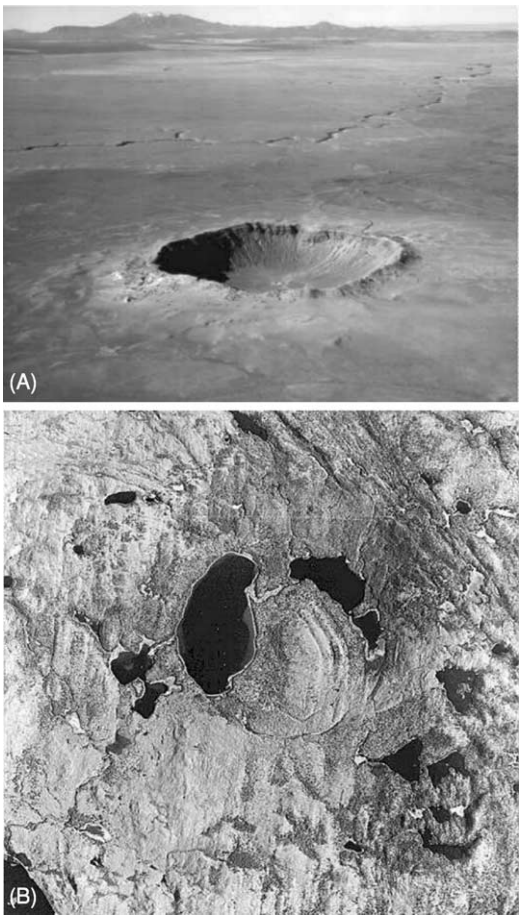


Figure 3 (A) Oblique aerial view of 1.2-km-diameter, 50 000-year-old simple crater, Meteor or Barringer Crater, Arizona, USA. (B) Vertical aerial view of 3.8-km-diameter, 450 ± 30-My-old, Brent Crater, Ontario, Canada. Note how this ancient crater has no rim, has been filled by sediments and lakes, and is a generally subtle topographic feature.

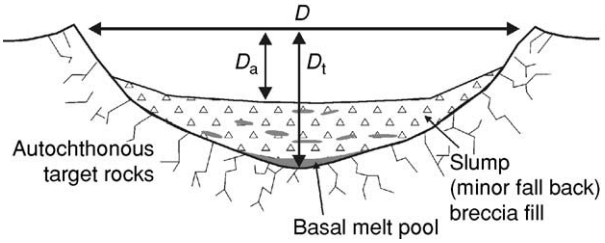


Figure 4 Schematic cross-section of a simple crater. D is the diameter and D_a and D_t are the depths of the apparent and true crater, respectively.

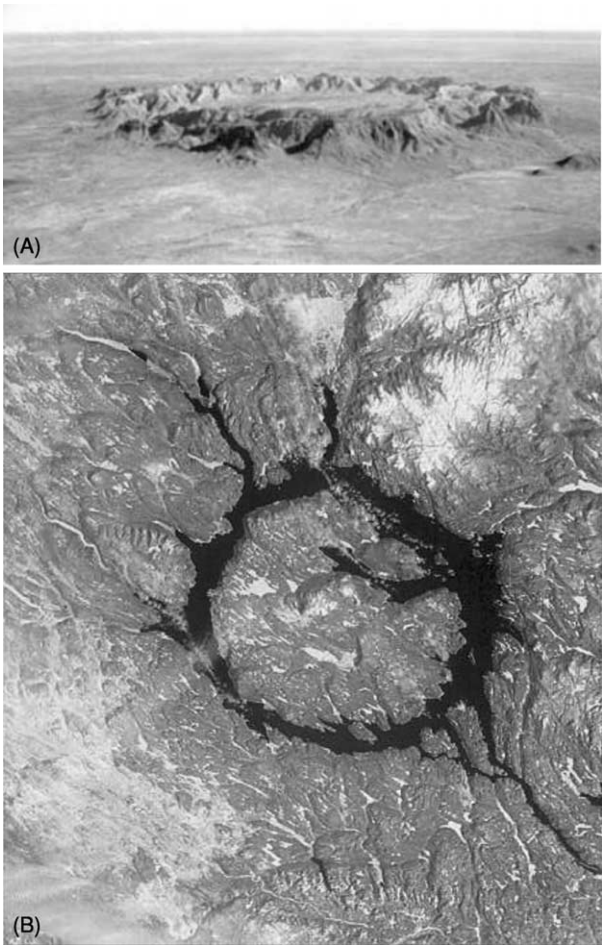


Figure 5 (A) Oblique aerial photograph of the Gosses Bluff impact structure, Australia. Note that all that is visible of the originally 22-km, 142.5 ± 0.8-My-old structure is a 5-km annulus of hills, representing the eroded remains of a central uplift. (B) Shuttle photograph of the Manicouagan impact structure, Canada, 100 km in diameter and 214 ± 1 My old. Note that the annular trough (with a diameter of ~65 km) is filled by water.

rim, there is an overturned flap of ejected target materials. Beneath the apparent floor is a lens of brecciated target material that is roughly parabolic in cross-section (Figure 4). This breccia lens is allochthonous and polymict. In places, this breccia lens contains highly shocked and melted target materials. Beneath this breccia lens, parautochthonous, fractured target rocks define the walls and floor of what is known as the true crater. The depth to the base of this breccia lens is roughly one-third of the rim diameter and the depth to the top of the breccia lens is about one-sixth. Shocked rocks in the parautochthonous materials of the true crater floor are confined to a small central volume at the base of the true crater.

With increasing diameter, simple craters show increasing evidence of wall and rim collapse and evolve into complex craters (Figure 5). Complex craters on Earth first occur at diameters greater than 2 km in sedimentary target rocks but not until diameters of 4 km or greater in stronger, more coherent, igneous or metamorphic, crystalline target rocks. The rim of complex craters is a structural feature, corresponding

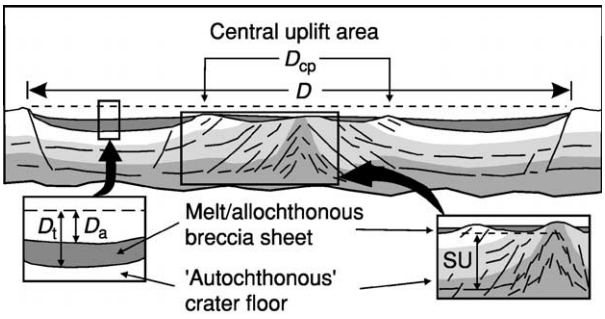


Figure 6 Schematic cross-section of complex impact structure. Notation as described in the legend to **Figure 4** with SU corresponding to structural uplift and D_{cp} to the diameter of the central uplift. Note preservation of beds in outer annular trough of the structure, with excavation limited to the central area.

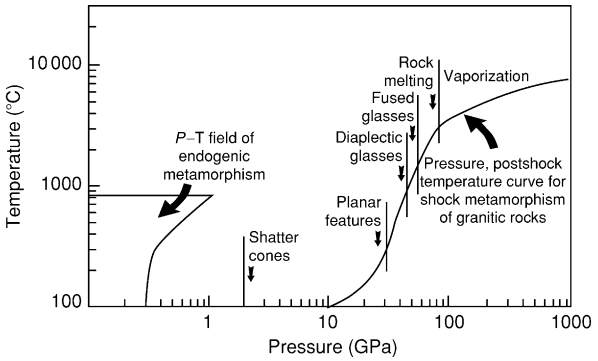


Figure 7 Temperature and pressure range of shock metamorphic effects compared to that of endogenic terrestrial metamorphism. Planar features include planar deformation features and planar fractures. Note scale is logarithmic.

to a series of fault terraces. Interior to the rim lies a down-faulted annular trough, which is partially filled by a sheet of impact-melt rock and/or polymict allochthonous breccia (**Figure 6**). Only in the central area of the crater is there evidence of substantial excavation of target materials. This central region is structurally complex and can be occupied by a central peak (**Figure 5**), which is the topographic manifestation of a much broader and extensive area of structurally uplifted target rocks that occurs beneath the centre of complex craters.

With increasing diameter, a fragmentary ring of interior peaks appears, marking the transition from complex craters to impact basins. There have been claims that the largest known terrestrial impact structures, e.g., Chicxulub, Mexico; Sudbury, Canada; and Vredefort, South Africa, have multiring basin forms. Although certain of their geological and geophysical attributes form annuli, it is not clear that these correspond, or are related in origin, to the obvious topographical rings observed, for example, in lunar multiring basins.

A small number of relatively young, and, therefore, only slightly eroded, complex impact structures (e.g., Haughton, Canada; Ries, Germany; Zhamanshin, Kazakhstan) do not have an emergent central peak or other interior topographical expression of a central uplift. These structures are in mixed targets of platform sediments overlying crystalline basement. This difference in form is probably a target rock effect but it has not been studied in detail.

Geology of Impact Structures

Although an anomalous circular topographic, structural, or geological feature may indicate the presence of an impact structure, other terrestrial geological processes can produce similar features. The burden of proof for an impact origin for a specific structure generally lies with the occurrence of shock metamorphic effects. Only small, young, simple structures, where the impacting body has been slowed by atmospheric deceleration, preserve physical evidence of the impacting body. These are generally fragments of iron or stony-iron meteorites. Stony meteorites are weaker and small stones are crushed as a result of atmospheric interaction. Larger impacting bodies (>100–150 m in diameter) survive atmospheric passage with undiminished impact velocity.

On impact, a portion of the impacting body's kinetic energy is partitioned into kinetic energy in the target rocks, which results in the formation of a craterform. The remainder of the impact kinetic energy is partitioned into increasing the internal energy of the target rocks. The target rocks are compressed by the passage of a shock wave and experience extremely high transient pressures and temperatures. For example, the peak pressure experienced in crystalline target rocks by the impact of a stony body, such as a chondritic asteroid, at 25 km s^{-1} is approximately 100 GPa. During shock compression, considerable pressure–volume work is done. On decompression, not all this mechanical work is recovered. This excess work is manifest as waste heat, leading to the heating, melting, and even vapourization of part of the target rocks and the impacting body, destroying it as a physical entity.

The combined effect of this compression and heating is the production of a series of irreversible changes that occur in individual minerals and rocks, which are known collectively as shock metamorphic effects. Shock metamorphic effects are produced at pressures and temperatures well beyond those in endogenic terrestrial metamorphism (**Figure 7**). The physical conditions upon impact are a function of initial impact parameters. Projectile and target rock type and impact velocity determine peak pressures on

impact, and projectile size determines the absolute radial distance at which particular shock metamorphic effects occur. Shock metamorphic effects are also produced on vastly different time-scales from endogenic metamorphic effects, and disequilibrium is the rule, not the exception.

Impact Melting

Impact-melted lithologies occur as glass bombs in crater ejecta, as dykes within the crater floor and walls, as glassy to crystalline lenses within the breccia lenses of simple craters, or as coherent annular sheets (Figure 8) lining the floor of complex craters. When crystallized, impact-melt sheets have igneous textures, and may, therefore, resemble endogenic igneous rocks. An important textural characteristic, however, of impact-melt rocks is the presence of mineral and rock fragments, which exhibit shock metamorphism to different degrees. The size of such fragments ranges from millimetres to several hundreds of metres and gradational changes in fragment content are observed, with highest concentrations towards the lower and upper contacts of coherent impact-melt sheets.

The composition of impact-melt rocks reflects the wholesale melting of a mix of target rocks, as opposed to partial melting and/or fractional crystallization relationships for endogenous igneous rocks. The composition of impact-melt rocks can be reproduced by a mixture of the various target rock types, in their appropriate geological proportions. Such parameters as $^{87}Sr/^{86}Sr$ and $^{143}Nd/^{144}Nd$ ratios also reflect the preexisting target rocks, although other isotopic systems, e.g., Ar^{39}/Ar^{40}, reflect remelting at the time of impact. In general, even relatively thick impact-melt sheets are chemically homogeneous over distances up to tens of kilometres. Differentiation is not a characteristic of impact-melt sheets (with the exception of the extremely thick, $\geq$2.5 km, Sudbury igneous complex, at the Sudbury Structure, Canada).

Enrichments above target rock levels in siderophile elements and Cr have been identified in some impact-melt rocks. These represent an admixture of up to a few percent of meteoritic material from the impacting body. In some melt rocks, the relative abundances of the various siderophiles have constrained the composition of the impacting body to the level of meteorite class. In other melt rocks, no siderophile anomaly has been identified. The latter may be due to the inhomogeneous distribution of meteoritic material or to differentiated and, therefore, non-siderophile-enriched impacting bodies, such as basaltic achondrites. High-precision chromium and osmium-isotopic analyses have also been used to detect a meteoritic signature at terrestrial impact structures.

Fused and Diaplectic Glasses

Shock-fused minerals are characterized by flow structures and vesiculation (Figure 9). Peak pressures required for shock melting of single minerals are 40 to 60 GPa, for which postshock temperatures exceed the melting points of typical rock-forming minerals. Under these conditions, the minerals in the rock will melt immediately and independently, after the passage of the shock wave. Melting is mineral selective, producing unusual textures in which one or more minerals show typical melting features; whereas others, even juxtaposed ones, do not. One of the most common fused glasses observed at terrestrial impact structures is that of quartz, i.e., lechatelierite.

Conversion to an isotropic, dense, glassy phase is a shock metamorphic effect unique to framework silicates. These phases are called diaplectic (from the Greek 'destroyed by striking') glasses, and are produced by breakdown of long-range order of the crystal lattice without fusion. Based on shock recovery experiments, the formation of diaplectic glass occurs between 30 and 45 GPa for feldspar and 35 to 50 GPa for quartz. The morphology of the diaplectic glass is

Figure 8 Approximately 80-m-high outcrop of coherent impact-melt rocks at the Mistastin complex impact structure, Canada. These rocks resulted from the melting of the target rocks by shock pressures in excess of approximately 60 GPa or 600 kbars (Figure 7).

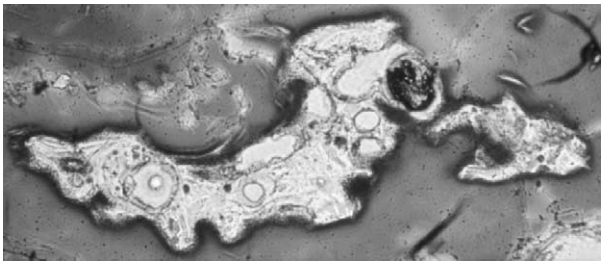


Figure 9 Photomicrograph of fused glass (lechatelierite) of the mineral quartz, from the Ries impact structure, Germany. Field-of-view = 2.5 mm.

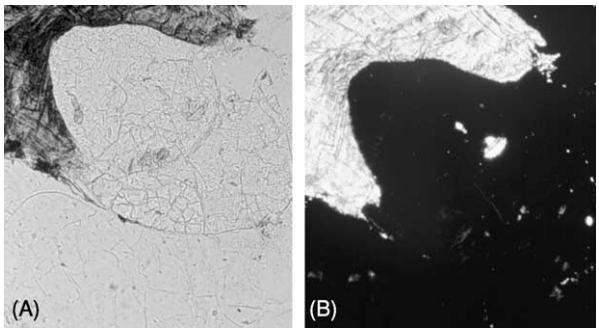


Figure 10 Diaplectic mineral glasses. Photomicrograph of granite from the Mistastin complex impact structure with biotite (upper left), quartz (centre), and plagioclase (bottom and far right). Field-of-view = 2.5 mm. (A) Plane light. (B) Crossed polars. Quartz and plagioclase (black) are isotropic, while retaining original mineral shape; i.e., they are diaplectic glasses. This requires approximately 40 GPa shock pressure (Figure 7).

the same as the original mineral crystal and shows no evidence of fluid textures. Maskelynite, the diaplectic form of plagioclase, is the most common example from terrestrial rocks; diaplectic glasses of quartz and of alkali feldspar also occur (Figure 10).

High-Pressure Polymorphs

Shock can result in the formation of metastable polymorphs, such as stishovite and coesite from quartz and diamond and lonsdaleite from graphite. Coesite and diamond are also products of endogenic terrestrial geological processes, including high-grade metamorphism, but the paragenesis and the geological setting are completely different from that in impact structures. In terrestrial impact structures, stishovite and coesite polymorphs occur in small or trace amounts, as very fine-grained aggregates, and are formed by partial transformation of the host quartz.

Planar Microstructures

The most common documented shock-metamorphic effect is the occurrence of planar microstructures in tectosilicates, particularly quartz (Figure 11). The utility of planar microstructures in quartz is a function of the ubiquitous nature of the mineral and its stability, in the terrestrial environment, and the relative ease with which they can be documented optically. Planar deformation features in quartz have various orientations and are produced under pressures of ~10 to ~35 GPa (Figure 7), with their crystallographic orientation providing a measure of the recorded shock pressures.

Shatter Cones

Shatter cones (Figure 12) are the only known diagnostic shock effect that is megascopic in scale. They

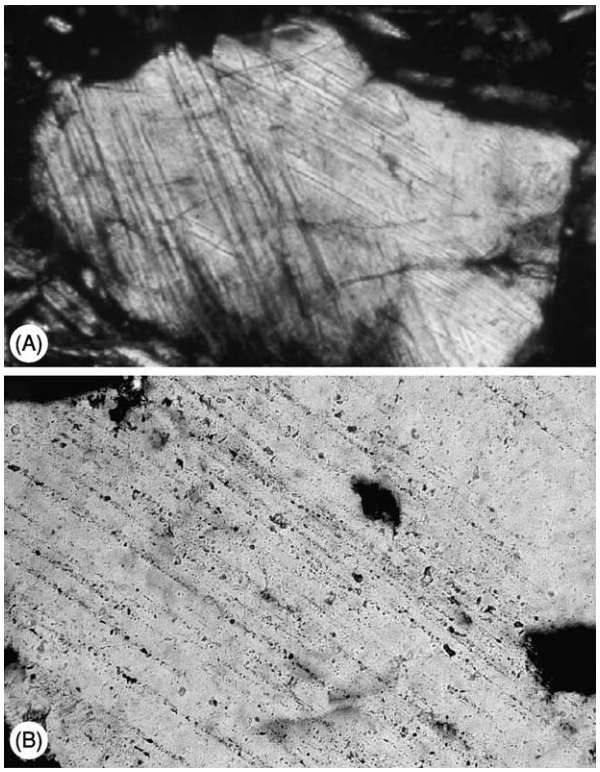


Figure 11 Photomicrographs of planar deformation features in quartz crystals: (A) nondecorated features, New Quebec impact structure, Canada; and (B) decorated features, Charlevoix impact structure, Canada. Field-of-view = 1 mm crossed polars.

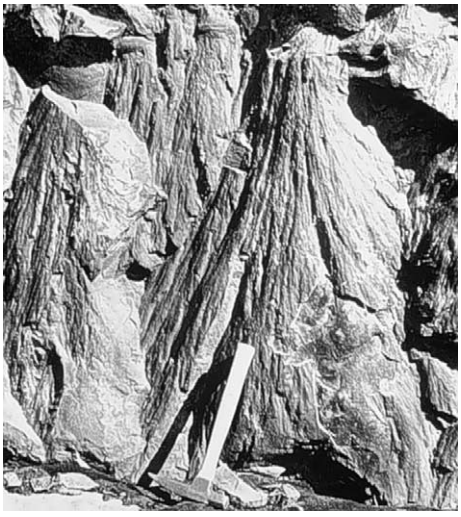


Figure 12 Shatter cones in quartzites at the Sudbury impact structure, Canada.

are unusual, striated, and horse-tailed conical fractures ranging from millimetres to tens of metres in length. Shatter cones are initiated most frequently in rocks that experienced moderately low shock pressures, 2–6 GPa, but have been observed in rocks

that experienced ~25 GPa. Shatter cones are best developed in fine-grained, structurally isotropic lithologies, such as carbonates and quartzites. They do occur in coarse-grained crystalline rocks but are less common and more poorly developed.

Impacts and Earth Evolution

The basic working hypotheses for the formation of the terrestrial planets is the accretion of small bodies by collision and the subsequent growth to larger planetary embryos. This early stage of planetary accretion is believed to be followed by a stage involving truly giant impacts between the embryos. In the context of planetary formation, therefore, impact is the most fundamental process (*see* **Solar System:** Moon).

In the case of the Earth, the last embryonic planetary collision may have resulted in the formation of the Earth's Moon. Numerical simulations of a glancing impact of a Mars-sized body with the proto-Earth result in the formation of an Earth-orbiting accretionary disk from which a Moon-like body can form. The key is that the bulk of the material that forms the accretionary disc is originally terrestrial or impacting body mantle material. Condensation and reaccretion result in volatile and siderophile-depleted material, which goes to form the Moon. Although the impact hypothesis for the origin of the Moon is consistent with the constraints of the dynamics of the Earth–Moon system and the geochemical nature of the Moon, it remains a model. The consequences of such a giant impact for the proto-Earth, however, would have been severe. They would have included massive remelting and the likely loss of the original atmosphere.

After planetary formation, the subsequent high rate of bombardment by the remaining tail of accretionary debris is recorded on the Moon and other terrestrial planets that have preserved portions of their earliest crust. This is not preserved on Earth, because of the high level of endogenic geological activity and the resultant relatively young surface of the Earth. In the case of the Moon, a minimum of 6000 craters with diameters >20 km are known to have been formed during the early period of approximately 4.5–3.8 Ga. There are also approximately 45 known impact basins ranging in diameter from Bailey at 300 km to the south pole Aitken basin at 2600 km. Throughout geological time, the Earth has received more impacts than the Moon. The Earth is a physically larger target for incoming bodies and has a much larger gravitational cross-section to attract incoming bodies. There has been considerable speculation as to the potential effects of these basin-sized impacts on the Earth. By analogy with the lunar case, it is likely that few terrestrial surface rocks would have survived intact through this period of heavy bombardment. Basin-sized impacts on the early Earth may have also affected the existing atmosphere, hydrosphere, and the potential development of the biosphere.

The oldest known terrestrial impact structures are Sudbury and Vredefort. They have reconstructed original diameters of approximately 250–300 km, and ages of 1.85 and 2.0 Ga, respectively. Although it is expected that such large events would have had a deleterious effect on the climate and biosphere, no direct evidence is known at present. These structures, however, also affected the local geology in a manner that is to human benefit. Both impact structures are the sites of world-class ore deposits.

An extensive literature concerning the evidence for a major impact event at the K–T boundary and its association with a mass extinction in the terrestrial biosphere 65 Ma exists. The unequivocal physical evidence for impact contained in K–T boundary deposits consists of planar microstructures in quartz, feldspar, and zircon, and the occurrence of stishovite, impact diamonds, high-temperature magnesioferrite spinels (believed to be vapour condensates), and various melt spherules, generally altered, but including the tektite-like glass spherules in Haiti and other Caribbean sites. The chemical evidence consists of a global siderophile anomaly in K–T boundary deposits, indicative of an admixture of meteoritic material.

Although originally contentious, there is now little doubt that the Chicxulub structure in the Yucatan peninsula, Mexico is the K–T impact structure. Chicxulub is buried by some 1 km of younger sediments but evidence for impact, in the form of planar microstructure quartz and feldspar in deposits interior and exterior to the structure, as well as impact-melt rocks, has been documented. In addition, the geochemistry of K–T tektite-like glasses from Haiti matches the mixture of lithologies found at the Chicxulub site. Isotopic ages for the impact-melt rocks at Chicxulub of 64.98 ± 0.05 Ma are indistinguishable from the K–T tektites at 65.07 ± 1.00 Ma (*see* **Mesozoic:** End Cretaceous Extinctions).

The original hypothesis for the killing mechanism for the K–T mass extinction suggested global darkening and cessation of photosynthesis due to ejecta in the atmosphere. Soot has also been identified in K–T deposits and ascribed to global wildfires. A considerable thickness of anhydrite ($CaSO_4$) occurs in the target rocks at Chicxulub. Impact heating of anhydrite would produce sulphur aerosols in the atmosphere. Modelling suggests that these sulphur aerosols would reduce light levels below those needed for photosynthesis for 6–9 months. In addition, if most of the aerosols were in the form of SO_2, solar

transmission would drop to 10–20% of normal (a cloudy day) for a few decades. It may be, therefore, that the devastating effects of the K–T impact are due to the character of the target rocks at Chicxulub. Whatever the case, the temporal association of an extremely large impact crater with a worldwide ejecta layer and a global mass-extinction event is well established. The cause–effect relationship is less well established and is the subject of current research.

The frequency of K–T-sized events on Earth is on the order of one every 100 ± 50 My. Smaller, but still significant, impact events occur on shorter time-scales. Dust loadings from the formation of impact craters as small as 20 km could produce light reductions and temperature disruptions on relatively short time-scales. Such impacts occur on Earth with a frequency of a few every million years or so and are not likely to have an effect upon the biosphere. The most fragile component of the present environment, however, is human civilization. There is little doubt that if it lasts long enough it will suffer severely or may even be destroyed by an impact event. The greatest threat (the product of probability and the expected death and destruction) is likely from ocean-wide impact-induced tsunamis, which occur on time-scales of tens of thousands of years. For example, model calculations indicate that the impact of even a relatively small body (approximately 200 m in diameter) would result in a wave in the open ocean that would still be 10 m high 1000 km from the point of impact. Potential impact events do, in fact, occur on human time-scales. For example, the Tunguska event on 30 June 1908 was due to the atmospheric explosion of a relatively small, <100 m, body at an altitude <10 km.

Terrestrial impact structures are the manifestation of extremely high-energy, transient geological events and, as such, have resulted in unusual local geological environments. Some of these have produced significant economic deposits. About 25% of known terrestrial impact structures have some form of economic deposit, and about half of these are currently exploited or have been exploited in the recent past. They range from local to world-class (e.g., reserves of 1.6 billion tonnes Ni–Cu ores at Sudbury). They also include hydrocarbon deposits (e.g., reserves of 50 million barrels of oil and 60 billion ft^3 of gas at Ames, Oklahoma, USA). The Ames impact structure produced the structural trap and also provided the source rocks, which are locally developed postimpact oil shales within the crater. In addition hydrocarbon production at Ames includes nontraditional reservoirs, such as the fractured crystalline rocks of the central uplift. The world-class Campeche Bank oilfield, Gulf of Mexico, produces most of its hydrocarbons from K–T breccias, with 10% porosity, resulting from seismic activity related to the nearby Chicxulub impact. Proven reserves are 30 billion barrels of oil and 15 trillion ft^3 of gas and exceed the entire onshore and offshore US reserves.

Summary Remarks

The detailed study of impact events on Earth is a relatively recent addition to the spectrum of studies engaged in by the geological sciences. More than anything, it was preparations for and, ultimately, the results of the lunar and planetary exploration programme that provided the impetus and rationale for their study. The terrestrial record of impact has made important contributions to our understanding of impact processes. Answers to many questions are known to, at least, the first order. Many details require clarification, however, and some problems, for example, the characteristics of superficial impact deposits, such as ejecta, are difficult to address in the active terrestrial geological environment.

It is apparent that impact can no longer be considered a process of interest only to the planetary community. It is a process that has fundamentally affected terrestrial evolution. For example, without the K–T impact, the present-day biosphere may have been quite different. Similarly, without a Mars-sized impact forming the Moon, and the resulting tidal forces on the Earth, one can only speculate how the littoral zone, the most important area in terrestrial ecosystem, would have evolved and been populated. Impact is the most catastrophic geological process known and is fundamental to the nature of the Solar System. Impact events have happened on Earth throughout geological time and will happen again. Although the occurrence of a large impact event has a low probability on the time-scale of human civilization, the consequences of its occurrence could be globally disastrous.

See Also

Earth Structure and Origins. **Engineering Geology:** Natural and Anthropogenic Geohazards. **Mesozoic:** End Cretaceous Extinctions. **Solar System:** Asteroids, Comets and Space Dust; Meteorites; Mercury; Venus; Moon; Mars. **Tektites**

Further Reading

Australian impact structures (1996) *AGSO Journal of Australian Geology and Geophysics* 16(4): 371–625.

Frankel C (1999) *The End of the Dinosaurs*. Cambridge: Cambridge University Press.

French BM (1998) *Traces of Catastrophe: A Handbook of Shock Metamorphic Effects in Terrestrial Meteorite*

Impact Structures. Lunar and Planetary Institute Contribution 954. Houston: Lunar and Planetary Institute.
Grady MM, Hutchison R, McCall GJH, and Rothery RA (eds.) (1998) *Meteorites: Flux with Time and Impact Effects*. Geological Society of London Special Publication 140. London: Geological Society of London.
Gehrels T (ed.) (1994) *Hazards due to Comets and Asteroids*. Tucson: University of Arizona Press.
Hodge PW (1994) *Meteorite Craters and Impact Structures*. Cambridge: Cambridge University Press.
Hoyt WG (1987) *The Coon Mountain Controversies*. Tucson: University of Arizona Press.
Melosh HJ (1989) *Impact Cratering: A Geologic Process*. Oxford: Oxford University Press.
Plado J and Pesonen LJ (eds.) (2002) *Impacts in Precambrian Shields*. New York: Springer-Verlag.
Plucker-Ehrenbrink BE and Schnitz B (eds.) (2001) *Accretion of Extraterrestrial Matter throughout Geologic Time*. Norwell, MA: Kluwer Academic Press.
Spudis P (1993) *The Geology of Multi-ring Impact Basins*. Cambridge: Cambridge University Press.

INDIAN SUBCONTINENT

A B Roy, Presidency College, Kolkata, India

Introduction

The Indian Sub-Continent constitutes a distinctive geographic entity: the countries included in the Sub-Continent, including Bangladesh, India, Nepal, and Pakistan, virtually cut off from the rest of Asia by lofty mountain chains. Almost half of the Sub-Continent's boundary in the north is bordered by the Himalayas and its associated branches: the Sulaiman and Kirthar Ranges to the Hindukush in the northwest, and the Naga Hills and the Arakan Yoma constituting the Indo-Myanmar Arc (**Figure 1**). The Himalayas extend for over 2500 km, from the Pamir in the west to the Mismi Hills in the east. The landscape changes sharply on the southern side of the great mountain ranges, where the high mountain terrain descends down to a huge plain land of unconsolidated sediments, known as the Indo-Gangetic Alluvial Plain (IGAP). The IGAP, which constitutes the enormous flood plains of the Indus and the Ganga River systems, also includes the narrow basin of the Brahmaputra River in the east and the Thar Desert in the west (including the North Gujarat Plain). South of the IGAP lies the rocky landmass of Peninsular India.

Except for the Himalayas and associated mountain ranges, the Indian Sub-Continent is traditionally considered as a shield of Precambrian rocks with a younger cover. The geophysical data, especially the seismic, Bouguer gravity anomaly, and heat flow patterns, suggest that much of the shield area attained considerable 'mobility' during post-Precambrian time. In this respect at least, the Indian Shield appears distinctly different from the better-known shield areas like the Canadian or the Ukrainian Shields.

The history of geological evolution of the Indian Sub-Continent is quite long and complex and, broadly speaking, took place in two stages. The first stage covered the entire Precambrian, which was the period of growth and final cratonisation. During the second stage of its evolution, the cratonised Indian Shield underwent considerable reconstitution that ultimately produced the present-day geomorphology as well as the tectonic character of the region.

The following description discusses the regional geology of the Sub-Continent mainly in terms of the history of evolution of the crust, starting from the evolution of the Archaean basement and the formation of Proterozoic basins to the different stages of the Phanerozoic reconstitution. Evidence for all of these come from smaller regions, which preserved these stages of crustal development. The description of regional geology of the Indian Sub-continent will be divided into the following heads:

i. Precambrian Indian Crust;
ii. Geology of the Gondwana basins;
iii. Between the Jurassic break-up and Himalayan collision.
iv. Geology of the Himalayas;
v. Quaternary sedimentation and neotectonics.

Precambrian Indian Crust

The Peninsular India lying south of the Indo-Gangetic Alluvial Plain is an old landscape, a considerable part of which is covered by Phanerozoic rocks, the Gondwanas, Deccan Traps, marine Mesozoic-Tertiary formations, and Recent alluvium. Tectonically speaking, the peninsula is a shield area that has remained free of any orogenic deformation since the Cambrian. The entire region was considered a 'terra incognita', even up to about three decades ago. Detailed studies in several interdisciplinary fields helped to characterise

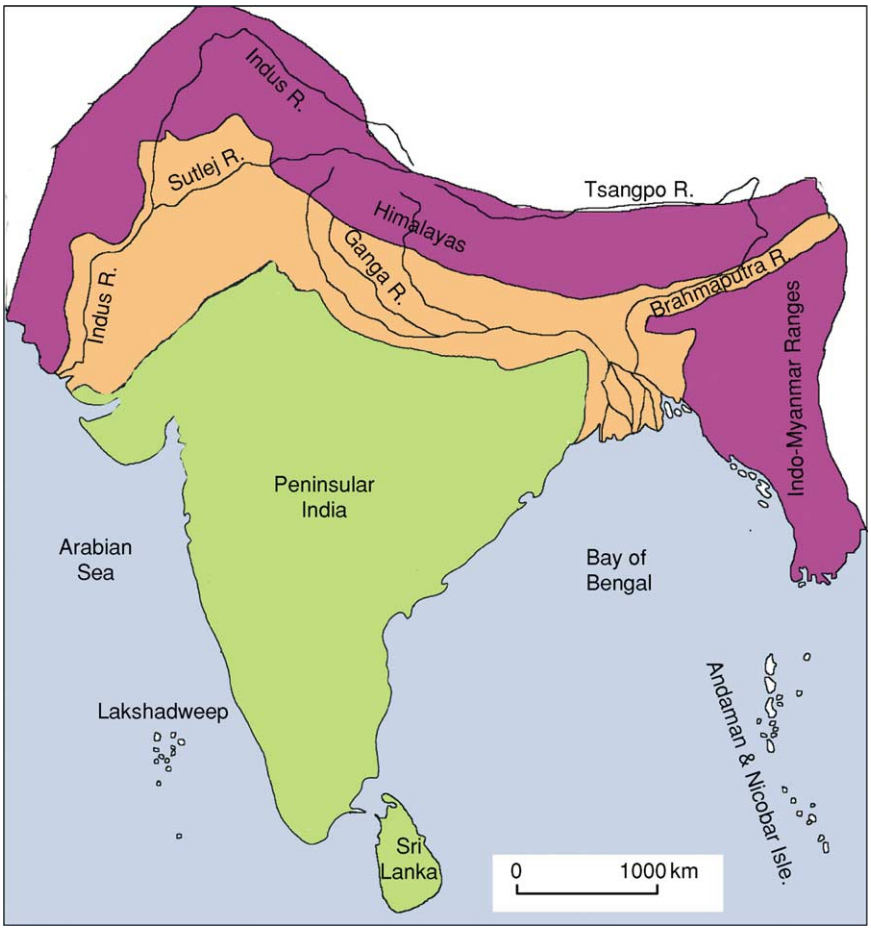


Figure 1 Physical map of the Indian Sub-Continent.

the Indian Shield as a coherent unit, comprising a number of Precambrian crustal blocks sutured together along joins marked by prominent lineaments. The Precambrian rocks occur as an isolated fault-bounded rectangular block at the Shillong Plateau. Tectonic affinity of this small block east of the Bengal Basin remains unclear. The crustal blocks include four cratons (Dharwar, Bastar, Singhbhum, and the Aravalli-Bundelkhand) and two granulite terrains (Eastern Ghats and Southern Granulite Terrains) and are characterised by distinctive evolutionary history and metallogenic patterns (Figure 2). The most ancient rocks constituting the basement in each craton are between 3.3 Ga and 3.5 Ga old. Before becoming a craton, all these crustal blocks were mobile belts. Typical cratonic platformal sediments cover large areas in the different cratons, apart from the Singhbhum and the granulite belts.

Dharwar Craton

Divided into the Eastern and the Western belts, this craton is marked by the evolution of a number of greenstone belts, called the schist belts, over sialic basements. No universally acceptable stratigraphical succession is known from the craton, which dates from the Archaean (3.5 Ga to 2.5 Ga). The principal structural trend of the Dharwar Craton is north-south, showing an eastward convexity (Figure 3). A change of metamorphic grade is indicated by the presence of very low grade rocks in the north to a transitional granulite facies in the south. A number of mineral deposits occur in the different schist belts. Important ones are iron-ores (as BIFs), chromite, nickel (as sulphides), copper, and gold.

The eastern Dharwar Craton by comparison shows more extensive Late Archaean reconstitution and granitic activities. The chronology of growth of the Dharwar Craton is given in Table 1.

Bastar Craton

The Bastar Craton also known as the Bhandara Craton or the Bastar-Bhandara Craton, occupies a quadrangular area with well-defined boundaries. Lithologically, the craton comprises patches of Archaean supracrustal rocks engulfed in a vast expanse of gneiss-granite of various ages, well-defined

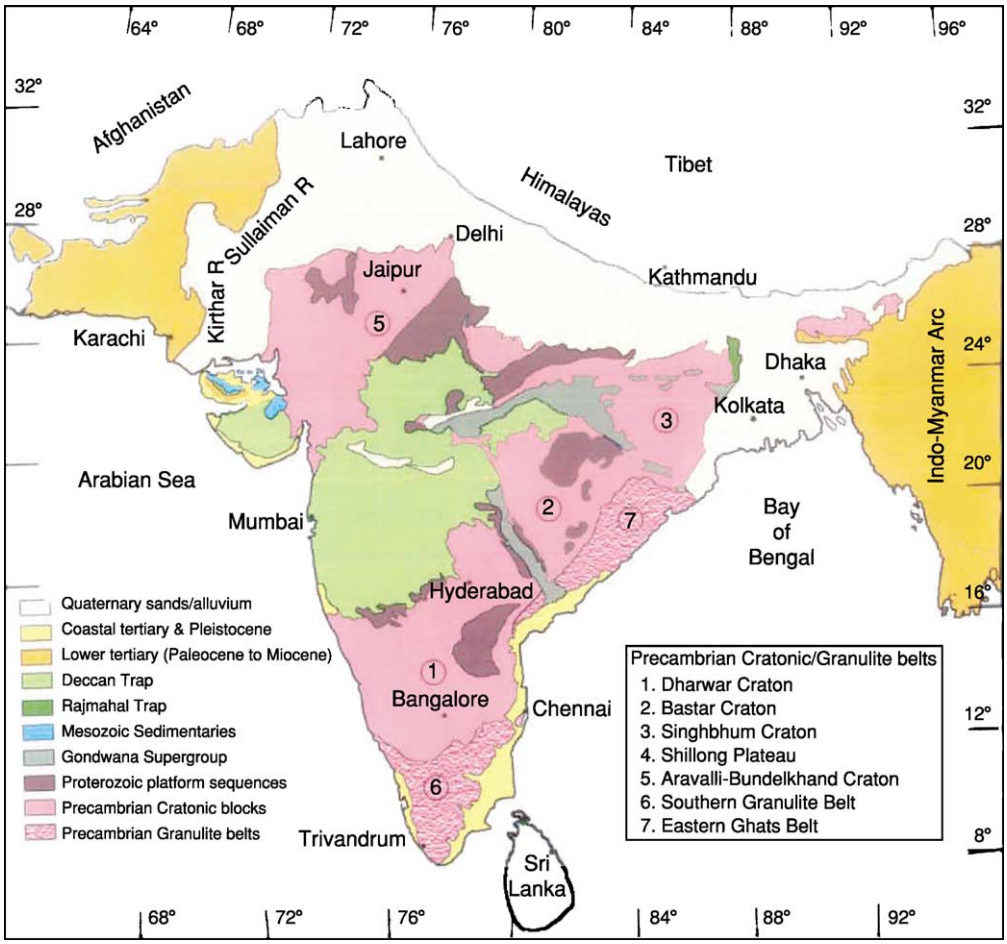


Figure 2 Generalised geological map of the Indian Sub-Continent.

Proterozoic fold belts (comprising the Sausar, Sakoli, and Dongargarh Groups), and cratonic platformal cover rocks of the Chattisgarh Group and its equivalents. The Deccan Trap covers it in the north (Figure 4).

The antiquity of the Bastar Craton is indicated by the occurrence of 3.5 Ga old trondhjemite gneiss. The oldest supracrustal rocks host large deposits of iron-ore (of BIF type). Barring the general high-grade metamorphic character of rocks, the supracrustals show features of dismembered greenstone sequences. There are occurrences of granulite-facies rocks in the Bastar Craton. The geochronological data, although poor suggests a number of episodes of reconstitution of the basement rocks during both the Archaean and the Proterozoic.

Important mineral deposits in the Bastar Craton include iron-ore in the Bailadela Group, manganese ores in the Sausar Group, and copper ores in the Malanchkhand Granite.

Singhbhum Craton

The most significant aspect of the recent works carried out during the last four decades is the recognition of an old Archaean nucleus, which was enlarged through the accretion of younger bodies around it. The components making up the Archaean nucleus include: (i) the oldest Archaean supracrustals, called the Older Metamorphic Group (OMG), and a variety of tonalitic gneisses that intruded it; (ii) massifs of the Archaean granitoids; and (iii) the supracrustals called the Iron Ore Group (IOG). The pre-IOG events are shown by the addition of juvenile materials during 3.3 and 3.0 Ga. The presence of a pre-OMG basement of still older age is inferred, based on the geochemical character of the OMG ortho-amphibolites. The granitic activities at around 2.6 Ga marked the closure of Archaean history in the craton. The stratigraphic status of the very lowly metamorphosed platformal sequence, the Kolhan Group, which occurrs in close association with the IOG, remains a matter of debate.

North of the Archaean nucleus lies arcuate Proterozoic fold belts (also known as the North Singhbhum mobile belts) which show a prominent northward convexity (Figure 5). Close to the southern boundary lies the prominent intracrustal shear zone that

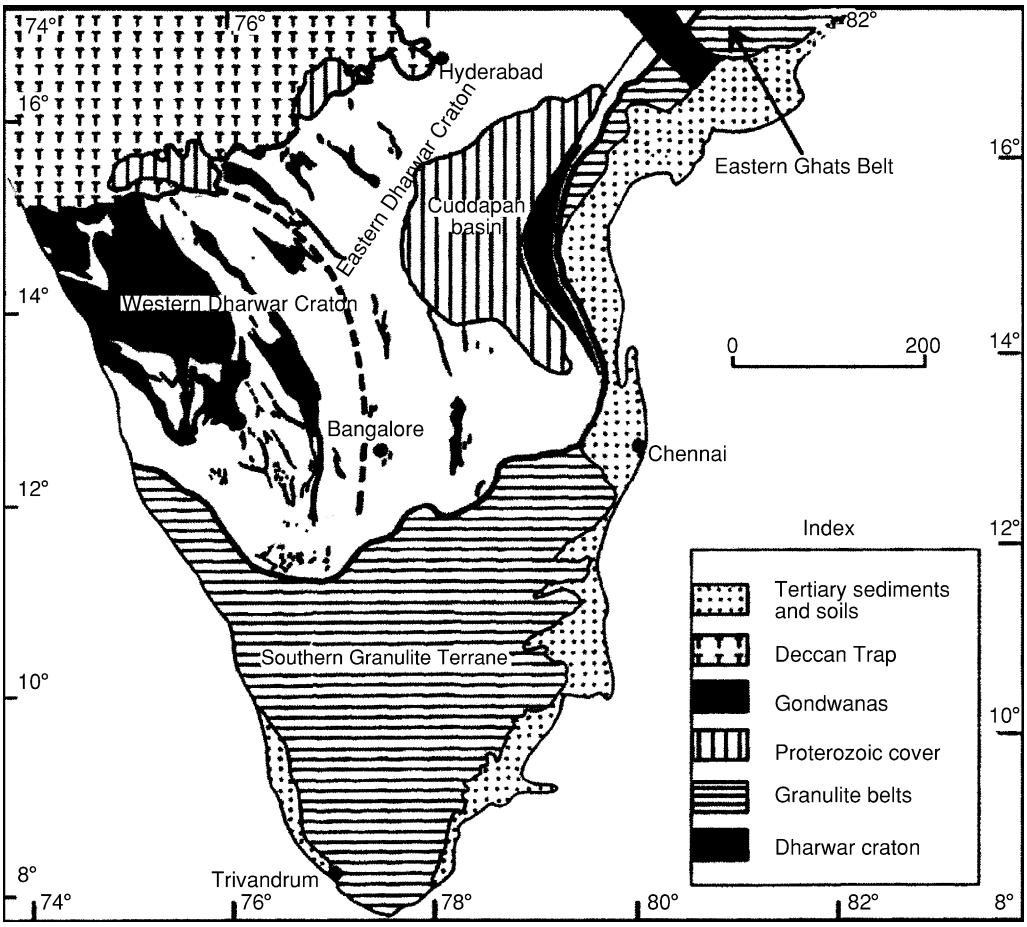


Figure 3 Generalised geological map of the Dharwar Craton and the Southern Granulite Terrain (adapted with permission from Radhakrishna 1983; © Geological Society of India, Bangalore).

Table 1 Chronology of evolution of crust of the Dharwar Craton (adapted from Radhakrishna and Vaididdhyanadhan, 1997)

2.5 Ga	Closepet granite and granulites
	Dharwar type schist belts (2.9–2.6 Ga) (Holenarasipur, Javanhalli, Bababudan, Chitradurga, Shimoga, Sandur)
3.0 Ga	Kolar-type auriferous schist belt (Kolar and Hutti)
>3.0 Ga	Sargur Group
~3.5 Ga	Formation of earliest basement

virtually wraps the northern half of the oval outcrop of the Singhbhum nucleus. There is hardly any unanimity on the stratigraphic succession of the Proterozoic fold belts. A simplified succession of the Singhbhum Craton is given in Table 2.

Aravalli-Bundelkhand Craton

Separated by the Son-Narmada Lineament from the three major cratonic blocks, the Dharwar. Bastar, and the Singhbhum, this large cratonic block comprises two distinctive crustal blocks, the Aravalli and the Bundelkhand, divided by an important tectonic boundary, the Great Boundary Fault (Figure 6). The Bundelkhand block comprises a gneissic basement formed between 3.3 and 2.5 Ga. The Proterozoic supracrustals include the Gwalior, the Bijawar, and the Mahakoshal Groups. The block did not undergo any tectothermal event later than 1850 Ma. By contrast, the Aravalli part of the craton shows continuous geological orogenic histories that ended at around 850 Ma (Table 3).

This craton contains rich deposits of phosphorite, in addition to lead-zinc and copper ore bodies.

Southern Granulite Terrane

The popular view on the evolution of this terrane assumes a model of amalgamation of microterranes. The network of dominantly dextral ductile shear zones emanating from the Cauvery Shear zones is considered central to the problem of assembly of Precambrian crust of the SGT (Figure 3). Geochronological data suggest the formation of granulite crust mostly during the Archaean, followed by crustal

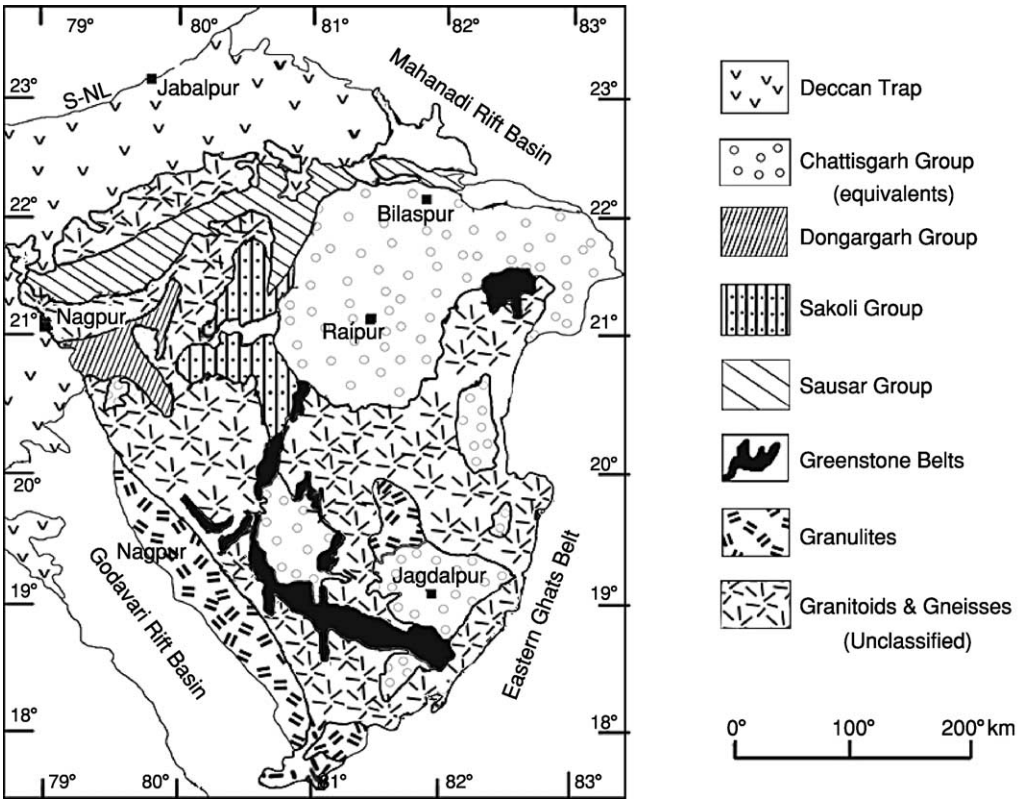


Figure 4 Geological map of the Bastar Craton. S-NL-Son-Narmada Lineament (adapted with permission from Roy *et al.*, 2000; © Geological Survey of India).

reworking during the Palaeoproterozoic and (in some only) during the Neoproterozoic.

Eastern Ghats Granulite Belt

This linear belt of deep sections is subdivided into four crustal provinces with widely different geological evolutionary histories. The northernmost part of the belt bordering, the Bastar Craton, formed in the late Archaean. The other domains provided indication of reconstitution by tectonothermal events in Mesoproterozoic, Grenville, and Pan-African times. Considerable similarity is reported between the Eastern Ghats Belt and the Rayner Province of East Antarctica in terms of their geological evolution.

Cratonic Basins

Several cratonic basins have developed in different parts of the Peninsular Shield area, of which the two most important ones are the Cuddapah Basin and the Main Vindhyan Basin. Both basins evolved quite early, in all probability during the late Palaeoproterozoic, signifying cratonisation of the Archaean crust prior to the basin opening. A break in sedimentation is indicated by the intrusion of kimberlite pipes in the older sequence and the deposition of diamond-bearing conglomerate at the base of the upper upper formations. The youngest cratonic basin developed in the west of the Aravalli-Bundelkhand Craton over a basement of the late Neoproterozoic plume-related volcanic to plutonic Malani Group.

Geology of Gondwana Supergroup

Definition, Classification, and Distribution

After a break in sedimentation for over 200 million years between the Ordovician and the Early Permian, the deposition of sediments in the Indian Sub-Continent started with the formation of tillites and glacial boulder beds in close association with Permian marine beds. This was accompanied by the deposition of fluvial and fluvio-lacustrine sediments in linear intracontinental rift basins. These sediments, along with intercalated plant remains that ultimately turned into coal seams, constitute the Gondwana Supergroup.

Precise definition of the Gondwana formations as formal stratigraphic units has suffered because of the inclusion of rocks deposited in diverse geological conditions into its ambit, and also because of the overemphasis given to the floral evidence. The Gondwana Supergroup has been redefined to include

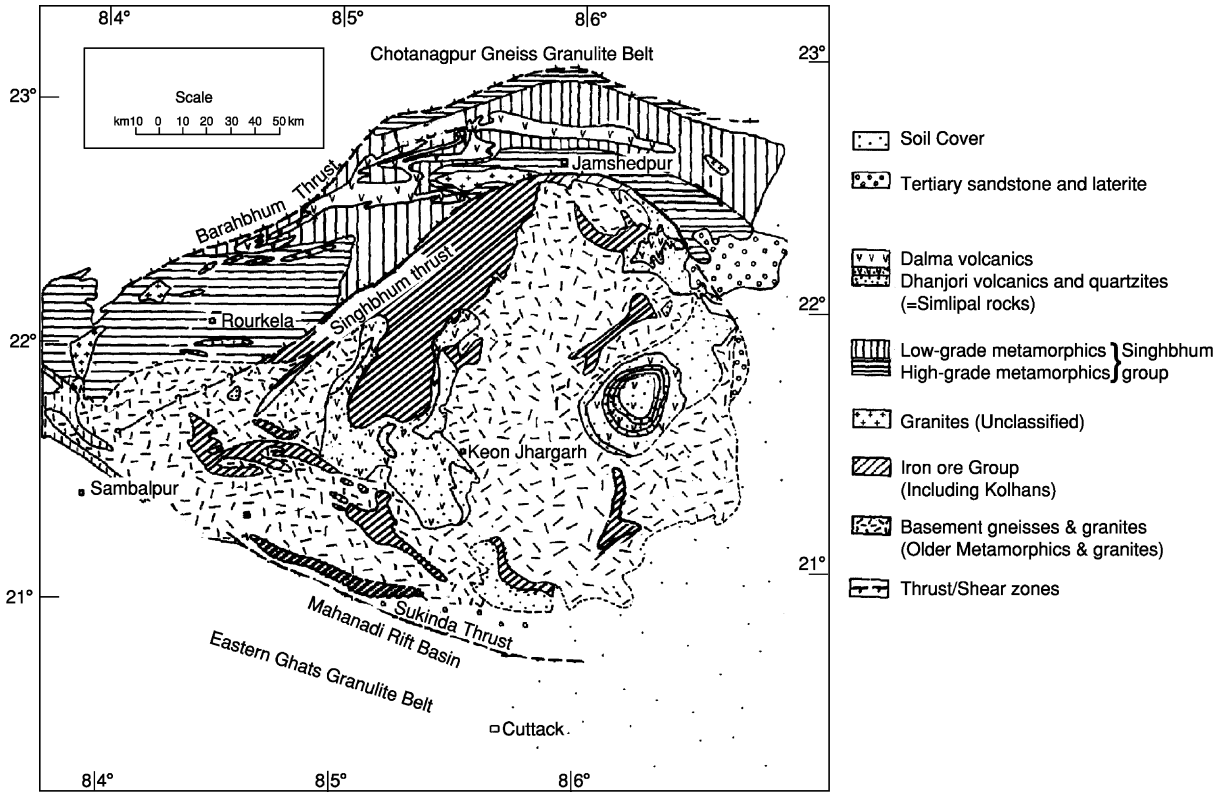


Figure 5 Geological map of the Singhbhum Craton (adapted with permission from Mahadevan TM, 1994; © Geological Society of India, Bangalore).

dominantly continental rift-basin deposits formed during the Early Permian and Middle Jurassic with minor marine inputs. Adoption of such a definition helps in the finer tuning of the stratigraphic succession by removing the contemporaneous and other Jurassic-Creataceous formations, which are correlatable when the Jurassic Gondwana broke up and the Cretaceous-Eocene plume outburst occurred. Recent studies favour adoption of smaller subdivisions of the Gondwana Supergroup instead of the two- and three-fold classifications based on floral assemblages.

Redefinition of the Gondwana Supergroup helps to identify three belts of the Gondwana basins. These are:

i. E–W trending Damodar Valley basins. Included in this belt is the narrow sub-parallel belt of subsidiary basins occurring north of the main belt.
ii. NW–SE trending Son-Mahanadi Valley basins. The belt widens and thickens in a northwesterly direction, meeting the extension of the Damodar Valley basins.
iii. NW–SE trending Pranhita-Gondwana Valley basins.

Stratigraphic classification of the Gondwana formations developed in the three belts is given in Table 4. The basal Talchir unit includes glacial tillites with shale and shale-siltstone rhythmites. All the other units comprise sandstone with shale as the dominant lithology. A number of coal seams (along with carbonaceous shale) occur in the Karharbari, Barakar, and Ranigani formations. The Barren Measure, which occurs between the Barakar and the Raniganj is (as the name implies) devoid of any coal seam. The Panchet and the Supra-Panchet are also coal-free sequences. The latter is termed the Mahadeva Formation in the Satpura Range.

Sedimentation, Basin Morphology and Evolution

Evidence from fossil flora and palynological studies indicate repeated changes in climate during the Gondwana sedimentation, which is also reflected in the nature of repetitive cycles of lithofacies. After the initial cold glacial condition, the climate warmed up along with increased humidity, which varied between medium and very high. This highly humid condition favoured enormous growth of plants. The en masse deposition of plant remains ultimately led to the formation of coal seams. The depositional condition changed to shallow freshwater with increased evidence of desiccation and salinity during later Panchet time.

Palaeoslope studies in the Damodar Valley provide inconsistent results, but a general north-northwesterly to northerly palaeoslope is suggested for these

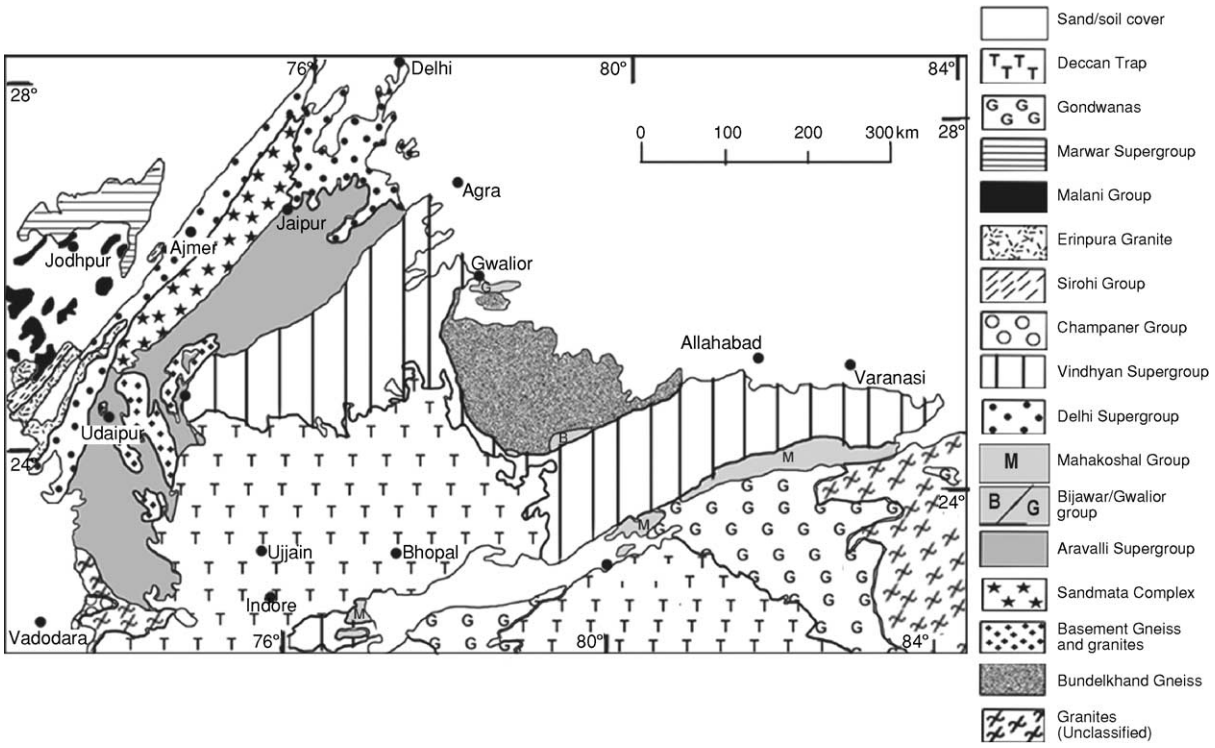


Figure 6 Geological map of the Aravalli-Bundelkhand Craton (adapted from Soni *et al.*, 1987; © Geological Society of India, Bangalore and Roy and Jakhar, 2002; © AB Roy, 2002).

Table 2 Generalised stratigraphic succession of Precambrian rocks in the Singhbhum Craton (adapted from Mahadevan, 1993 and Mukhopadhyay, 2001)

Phanerozoic covers (soils and laterites and Gondwana formations)	
Dolerite dikes and sills	
Gabbro-Anorthosite suites	
Kolhan Group[a] (?)	
Dalma lavas, Dhanjori lavas with quartzite[b] (Simplipal lavas and quartzites)	
Singhbhum Group – Pelitic and arenaceous metasediments with mafic sills	
Iron Ore Group	Upper Shales with sandstones and volcanics Banded hematite jasper with iron ore Mafic tuffs and lavas Tuffs, felsic volcanics and tuffaceous shales Sandstone and conglomerate (local)
Granitoids	(3.0–2.6 Ga)
Older metamorphic tonalitic gneisses	(3.3–3.4 Ga)
Older Metamorphic Group	Pelitic schist, quartzite, para- and ortho-amphibolite
Sialic basement of unknown age	

[a]Some authors consider the Kolhan Group as a component of the Archaean Iron Ore Group.
[b]Dhanjori Group is placed by some below the Singhbhum Group.

Table 3 Precambrian tectonostratigraphic framework of the Aravalli cratonic block (adapted from Roy and Jakhar, 2002)

Marwar Supergroup (cratonic sediments)	Youngest Proterozoic
Malani Group (volcanics, granites, sediments)	780–680 Ma
Erinpura Granites	~850 Ma
Sirohi Group (metasediments)	
Intrusion of Gabbro-diorites	~1000 Ma
Synorogenic granites in the Delhi Fold Belts	~1450 Ma
Delhi Supergroup (metasediments and metavolcanics)	
Granulite exhumation in Sandmata Complex	1725–1625 Ma
Synorogenic granites (Darwal Granite)	~1850 Ma
Aravalli Supergroup (metasediments and minor metavolcanics)	
Basement Gneiss (Mewar Gneiss Complex) and granites	~2500– ~3300 Ma

basins. In contrast, the sedimentation both in the Pranhita-Godavari and Son-Mahanadi Valley basins possibly took place along linear fluvial basins of river systems draining southeast to northwest in the direction of the palaeoslope.

Gondwana sedimentation during the initial stage (primarily during the Talchir sedimentation) took place in eroded topographic depressions. Based on

Table 4 Generalised correlation of the Indian Gondwana formations (adapted from Shastri *et al.*, 1977)

Age	*Damodar Valley*	*Mahanadi-Son Valley*	*Pranhita-Godavari*
JURASSIC			
Lower			Kota
TRIASSIC			
Upper	Supra-Panchet	Parsora Mahadeva	Dharmavaram
			Maleri
Middle		Pali	
			Bhimavaram
Lower	Panchet		Yerapalli
			Mangali
PERMIAN			
Upper	Raniganj Barren-Measure	Kamthi	Kamthi
			Motur
Lower	Barakar	Barakar	Barakar
	Karharbari	Karharbari	
	Talchir	Umaria-Talchir	Talchir

the similarities in sedimentary succession in several of the now-isolated basins, the presence of a master basin for the deposition of the Gondwana sediments has been suggested. Faulting, as well as topographic relief during the post depositional phase, presumably caused the varying morphology of the individual basins. Popular opinion, however, insists on a rift origin of the Gondwana basins, which at a later stage developed into half or full grabens. The progressive, as well as repetitive movements along block faulted basement slabs underlying the basin, controlled the sedimentation in different Gondwana basins. The present-day basin geometry is a combined effect of faulting in three stages, at the initiation, during, and post-dating sedimentation.

Between Jurassic Breakup and the Himalayan Collision

The Gondwana sedimentation, which began in the Permian, continued until the Lower Jurassic. The next major global event that grossly affected Indian continental block was the breakup of Gondwana at around 165 Ma. The initial separation resulted in marine incursions and deposition of sediments in western Rajasthan and in the Kachch region along WNW–ESE trending rift basins. The deposition of continental sediments, which had earlier stopped in different Gondwana basins before the Lower Jurassic was also resumed, at least in certain cases. The Gondwana breakup event is also responsible for the development of arrays of fracture systems in the Indian Continental block (Figure 7). Geomorphologically expressed as lineaments, these fracture systems, which developed either as new sets of fractures or as reactivated old tectonic grains, helped to significantly change the geophysical character of the Indian crust in a variety of ways.

As the Indian continental block (along with Madagascar and Seychelles, and Antarctica remaining attached to it) moved northward following the dismemberment of Gondwana, it was affected successively by the outbursts of four plume heads centred at Marion, Reunion, Crozet, and Kerguelen Islands. The manifestations of the Crozet Plume outbursts are virtually unknown. The Marion Plume outbursts resulted in the separation of Madagascar from the Indian continental block during 80 to 90 Ma (Figure 8). Evidence for this comes from the occurrence of 88–90 Ma old acid as well as mafic rocks in different parts of central and North Kerala, St Mary's Island off the Karnataka coast and also from Madagascar. The Rajmahal Traps and the Sylhet Traps are the manifestations of Kerguelen Plume activities, which lasted from 130 Ma to 110 Ma. The plume caused the separation of Antarctica from the Indian continental block and induced Cretaceous marine ingress both in the south-eastern part of the peninsula as well as in the south-east of the Bengal basin.

The timing and the passage of impingement of the Reunion Plume are well recorded in the form of different features that developed between Kohistan in the north to the Lakhsadweep-Maldive Islands in the south. The important features that developed during between 70 Ma and 64 Ma include:

i. the formation of narrow sedimentary basins, locally having volcanic inputs (many of which turned into oil-bearing formations) in western Rajasthan and northern Gujarat;
ii. extensive mafic (continental tholeiite basalt) volcanicity in western and central India;
iii. intrusions of plutonic alkaline masses of diverse composition which were also responsible for the development of isolated patches and linear belts

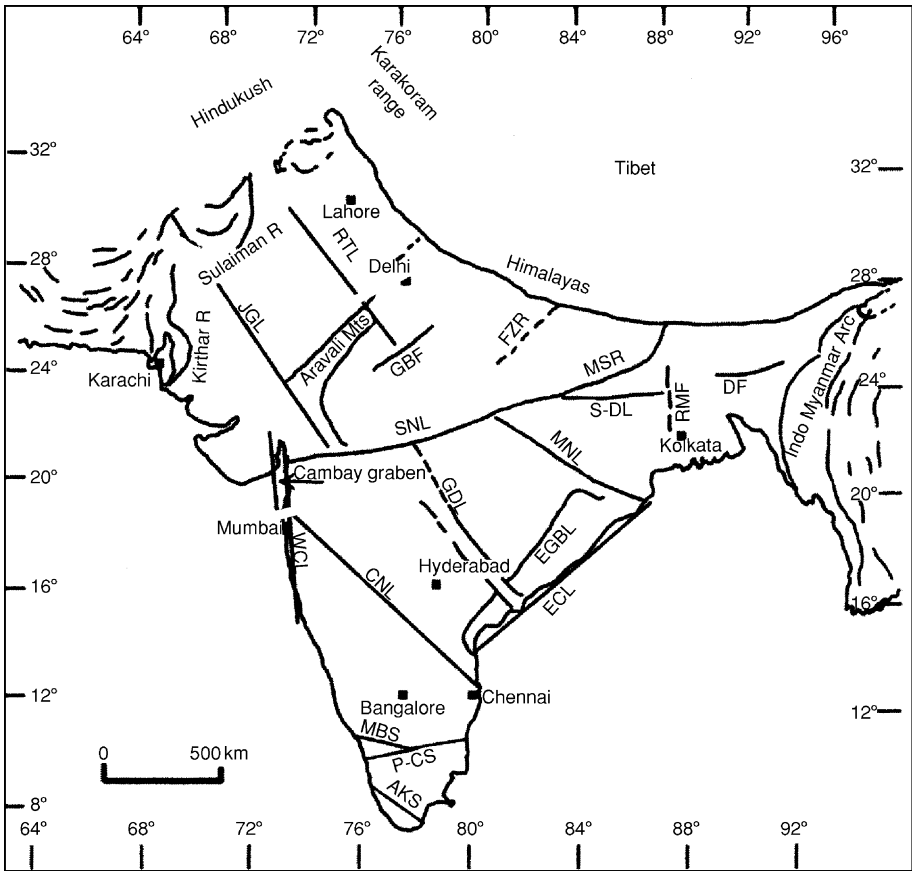


Figure 7 Generalised map showing the distribution of major lineament-fractures in the Indian Sub-Continent (adapted with permission from Roy, 2004; © Gondwana Research, International Association for Gondwana Research, Japan). Legend: AKS-Achankovil Shear Zone: CNL-Chennai-Nasik Lineament: DF-Deuki Fault: ECL-East Coast Lineament; GDL-Godavari Lineament; GBF-Great Boundary Fault; FZR-Faizabad Ridge: JGL-Jaisalmer-Barwani Lineament; MBS-Moyar-Bovani Shear Zone; MNL-Mahanadi Lineament; MSR-Munger-Saharsa Lineament; P-CS-Palghat-Cauvery Shear Zone; RMF-Rajmahal Fault; SNL-Son-Narmada Lineament; S-DL Son-Damodar Lineament; RTL-Raisinghgarh-Tonk Lineament; WCL-West Coast Lineament.

of positive Bouguer gravity anomaly and zones of high heat flow; and

iv. the development of fractures/lineaments. The palaeomagnetic data from the Deccan flood basalt suggests their development in two stages during C30N and C29R.

Geology and Evolution of the Himalayas

The Himalayas, the grand crescent-shaped mountain ranges with a prominent southward convexity, fringes the entire northern margin of the Indian Sub-continent. The world's loftiest and youngest mountain ranges extend for over 2500 km from south of the Indus Valley beyond Nanga Parbat (height 8114 m) in the west to Namcha Barwa (height 7755 m) in the east. Topographically, the Himalayas is bent sharply at the western end to join with the Sulaiman and Kirthar Ranges. With a similar sharp

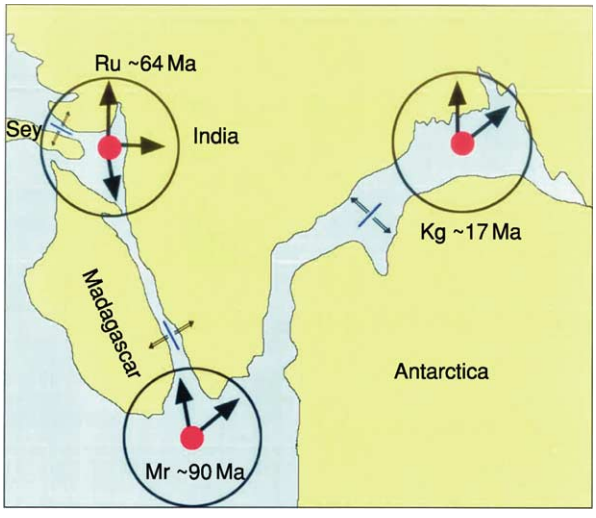


Figure 8 Simplified map of India, Antarctica and Madagascar, showing timing of the three different plume outbursts around the Indian Sub-Continent. (Adapted with permission from from Raval and Veeraswamy, 2003; © Gondwana Research, International Association for Gondwana Research, Japan.)

bending at the eastern end, the mountain ranges join the north-trending Indo-Myanmar Arc represented by the Naga Hills and the Arakan Yoma (Figure 9).

The Himalayas are divided axially into the following five units, each showing distinctive litho-tectonic character and evolutionary history:

i. The Sub-Himalayas. 10 to 50 km wide belt of Late Tertiary molasse sediments age constituting the Siwalik Group. The belt also includes the older Murree formations and their equivalent, the Dharamshalas.
ii. The Lesser Himalayas. 60 to 80 km wide belt predominantly comprising Proterozoic low-grade metamorphosed rocks overlain by thrust sheets of granites and metamorphic rocks.
iii. The Higher (or Great) Himalayas. 10 to 15 km thick belt of dominantly Precambrian metamorphites and young granites of Cenozoic age. This is also the zone of highest uplift.
iv. Trans-(or Tethyan) Himalayas, a belt of dominantly shelf (usually fossil-bearing) sediments of Late Proterozoic to Cretaceous age, bounded by the Indus-Tsangpo Suture Zone (ITSZ), a relatively narrow belt of ophiolites and associated sediments. Though not a thrust contact, the ITSZ is an important tectonic contact welding the Indian continental block with the Tibetan block Immediately north of the ITSZ is a belt of 40 Ma to 100 Ma old granitoids, known as the Trans-Himalayan batholith granites.

The tectonic architecture of the Himalayas is built on three prominent intracrustal thrusts. From north to south these thrusts are:

i. The Main Central Thrust (MCT) which separates the crystalline rocks of the Higher Himalayas from the low-grade metamorphites of the Lesser Himalayas.
ii. The Main Boundary Thrust (MBT) which regionally separates the Lesser Himalayas from the Sub-Himalayas.
iii. The Himalayan (Main) Frontal Thrust (HFT or MFT) which demarcates the tectonic and physiographic boundary between the Siwaliks and the Indo-Gangetic Alluvial Plain.

A unique feature of the Himalayas is its crustal thickness, which rises from about 35 km in the IGAP to a thickness of between 65 and 80 km over

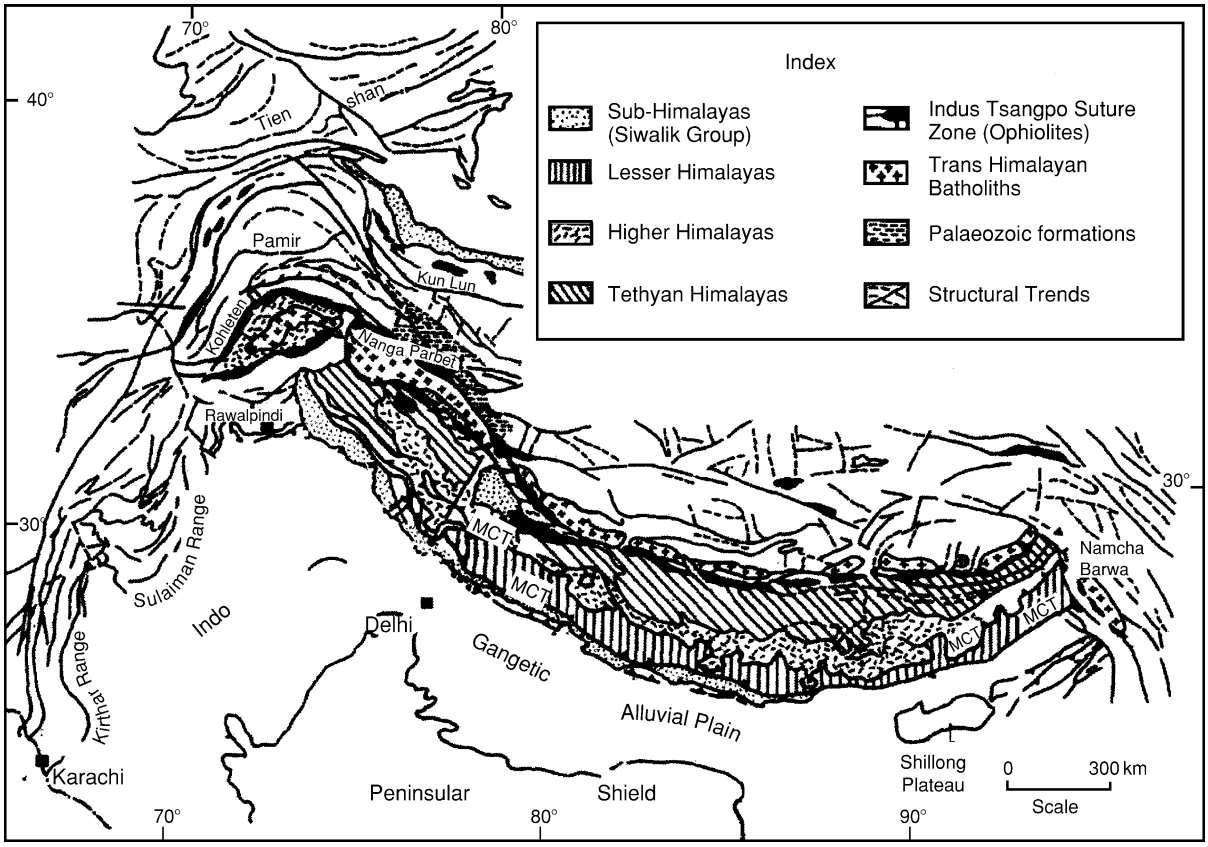


Figure 9 Geological map of the Himalayas showing important tectonic elements. MBT-Main Boundary Thrust; MCT-Main Central Thrust. (Adapted with permission from Mahadevan, 1994; © Geological Society of India, Bangalore.)

the Higher Himalayas. The increasing crustal thickness is reflected in the MOHO dip, which has been estimated to be 7°–8° N under the Sub-Himalayas up to over 15° further north.

Lithotectonic Units and Sedimentation

A considerable part of the Higher Himalayan mass is the Precambrian rocks which also constitute Peninsular India. An important component of the Lesser Himalayas is the Lower Palaeozoic (*ca.* 500 ± 50 Ma) dominantly per-alkaline anorogenic granites, which are popularly termed the Pan-African. The sedimentary rocks suggest continuity of sedimentation in some parts in the northern shelf sea across the Precambrian-Cambrian boundary into the Lower Palaeozoic, especially during the Cambrian. Records of continuity of sedimentation in these basins during the Ordovician and Silurian are, however, quite equivocal, although both periods are represented by fossils. Deposition of shelf facies sediments was resumed during Carboniferous in zones of marine incursions along intracratonic rift basins that had developed in the Salt Range and in the Kashmir region. The deposition in some basins continued until the Triassic. There are records of contemporary volcanicity in the Pir Panjal Ranges and in other places in the east. Both marine and continental sediments are correlatable with those of the Gondwana deposits which occur in parts of Nepal and the Sikkim Himalayas.

Continuous sedimentation from the Cambrian to the Eocene, with a number of breaks, is recorded in the Tethys belt. Fossil records from different parts of this belt indicate that the extent of these breaks were not of uniform duration.

The closure of the Tethys Ocean by the Eocene caused a brief pause in sedimentation, which was resumed around mid-Miocene times in two important basins. One in the north opened as a major intermontane (back arc) basin in the suture zone, leading to the deposition of Indus Group. The Siwalik Group was deposited in the southern foreland basin that developed in front of the rising Himalayas from around 18 Ma. The Siwalik Group bears rich fossil records of plants, molluscs, fishes, reptiles, and mammals. The last Himalayan upheaval at around 1.7 Ma caused shifting of the depocentres to the south to build up the flood plains of the Indo-Gangetic Alluvial Plain.

While the closure of Tethys Sea marked the end of sedimentation in the north, the marine shelf sedimentation continued both along the eastern and western margins of the Indian continental block, in the Naga Hills and Arakan Yoma in the east and the Sulaiman and the Kirthar Ranges in the west. Sedimentation in these basins, which began in the Eocene, continued at least until the Oligocene.

Himalayan Tectonics

The earliest Himalayan deformation coincided with the final closure of the Tethys Sea at around 50 Ma, affecting the rocks on either side of the suture zone. There was a distinct southward polarity of the deformation across the Tethys to the Higher Himalayan Crystalline Complex. A series of south-directed recumbent folds and thrusts were produced in the Higher Himalayas, resulting in thickening of the crust, with attendant Barrovian metamorphism, anatexis and the generation of leucogranites. The southward transmission of thrust nappes by the MCT, continued till around 22 Ma. This was also the time when the Barrovian metamorphic isograds underwent inversion.

Almost simultaneously with the piling of the fold-thrust nappes in the Higher Himalayas, the Indus molasse basin in the north and the Siwalik molasse basin in the south developed as rapidly subsiding troughs.

Like the piggy-back thrusting model, the southward transmission of the fold-thrust nappes, which was initially along the MCT, was later carried on by MBT in the south. The HFT (or MFT) which overrides the Recent sediments was the last thrust to form in the Himalayan tectonics.

The Himalayas represent a classic example of continent-continent collision. The very similar tectonic pattern observed over the entire length of the Himalayas is primarily an expression of the impact of two continental blocks. Complexities noted in the western end of the Himalayas arise because of the development of an island arc complex (Kohistan-Dras Island Complex) prior to its collision with the Karakoram microplate (possibly during the mid-Crataceous).

Palaeomagnetic data indicate an initiation of the continental collision at equatorial latitudes, resulting in the progressive suturing from the Paleocene in the north-western Himalayas until the Eocene in the eastern Himalayas. Continued convergence and indentation of the Indian continental block with southern Asia (or Tibet) up to the Early Miocene, resulted in the doubling of the crustal thickness over a large region of the Himalayas, the Pamir-Hindukush and Tibet. The total area of the thickened crust may account for about 2000 km of crustal shortening in the entire orogen. As to the origin of the Himalayan arc, palaeomagnetic observations seem to favour a steady-state model of formation of the arcuate bending of the mountain ranges due to Late Tertiary

counterclockwise rotational underthrusting of the Indian continental block beneath the Tibetan Plateau after the latest Miocene.

Quaternary Sedimentation and Neotectonics

The Quaternary geology, which began with the waning phase of the Siwalik sedimentation, came to an end with the most recent upheaval of the Himalayas. The depocentres had by then shifted to their subsiding southern parts which ultimately evolved as the Indo-Gangetic Alluvial Plains (IGAP). Geographically, apart from the Indus-Ganga flood plains, the IGAP also include the narrow basin of the Brahmaputra River in the east and the Thar Desert (along with the North Gujarat Plain) in the west. The alluvial sediments over the entire IAGP belt range between 400 m and 800 m, with a maximum thickness of about 6 km along the edge of the Himalayas. The belt is divided into a number of sub-basins by several submerged ridges (basement highs) lying across it.

Quaternary sediments outside the IGAP occur in the Narmada and Tapti Basins in Peninsular India, and along the eastern coastlines. Thick laterite formations (some of which contain rich bauxite deposits) were produced at this time in parts of central India, eastern Ghats and the Konkan coasts in the western Ghats.

The formation of the Thar Desert in the east of the Indus Basin, which had a fluvial pre-history, is linked with the establishment of the monsoon system over the Sub-Continent by the mid-Pleistocene, with the high rising Aravalli Mountains producing the rain-shadow zone to its west. The series of saline lakes that dot the entire desert-land were formed by the segmentation and blocking of river channels due to neotectonic movements.

The Quaternary neotectonic movements caused spectacular geomorphic changes in the entire Sub-Continent, primarily through movements of fault-bounded blocks. The Rann of Kachch in northern Gujarat is a classic example of regional uplift during historical times. The development of the Ganga-Brahmaputra-Megna Delta Complex (also known as the Sundarban Delta) is a very important geological landform feature which evolved in three stages of tectonically-influenced delta sedimentation processes during the Late Pleistocene. A point of interest in the delta sediments is the arsenic toxicity in the groundwater aquifer within the Holocene sediments.

See Also

Gondwanaland and Gondwana.

Further Reading

Dasgupta S and Sengupta P (2003) Indo-Antactica correlation: a perspective from the Eastern Ghats Granulite Belt, India. In: Yoshida M, Windley BF, and Dasgupta S (eds.) *Proterozoic East Gondwana: Supercontinent Assembly and Breakup*. Geological Society of London, Special Publications 206: 131–143.

Mahadevan TM (1994) *Deep Continental Structure of India: A Review.* Geological Society of India Memoir 28.

Mahadevan TM (2002) *Geology of Bihar & Jharkhand.* Bangalore: Geological Society of India.

Mukhopadhyay D (2001) The Archaean Nucleus of Singhbhum: the present state of knowledge. *Gondwana Research* 4: 307–318.

Naqvi SM and Rogers JJW (1987) *Precambrian Geology of India.* New York: Clarendon Press. Oxford: Oxford University Press.

Radhakrishna BP (1983) Archaean granite-greenstone terrain of the South Indian Shield, 1–46. In: Naqvi SM and Rogers JJW (eds.) *Precambrian of South India.* Geological Society of India Memoir 4.

Radhakrishna BP and Vaidyanadhan R (1997) *Geology of Karnataka.* Bangalore: Geological Society of India.

Raval U and Veerasswamy K (2003) India Madagascar separation: Break-up along a pre-existing mobile belt and chipping of the craton. *Gondwana Research* 6: 467–485.

Roy, Abhinaba, Raamachandra HM, and Bandopadhyay BK (2000) Supracrustal belts and their significance in the crustal evolution of Central India. *Geological Survey of India, Special Publication* 55: 387–406.

Roy AB (2004) The Phanerozoic reconstitution of Indian Shield as the aftermath of break up of the Gondwanaland. *Gondwana Research* 7(2): 387–406.

Roy AB and Jakhar SR (2002) *Geology of Rajasthan: Northwest India: Precambrian to Recent.* Jodhpur: Scientific Publishers (India).

Sastry MVA, Acharyya SK, Shaw SC, *et al.* (1977) Stratigraphic lexicon of Gondwana Formations of India. *Geological Survey of India Miscellaneous Publication* 36, 170.

Searle MP, Windley BF, Coward MP, *et al.* (1987) The closing of Tethys and tectonics of the Himalaya. *Geological Society of America Bulletin* 98: 678–701.

Soni MK, Chakraborty S, and Jain VK (1987) Vindhyan Supergroup – A review. *Geological Society of India Memoir* 6: 87–138.

Subbarao KV (1999) Deccan Volcanic Province. *Geological Society of India Memoir* 43(1) and 43(2), 947.

Valdiya KS (1998) Dymanic Himalaya. Hyderabad: Universities Press.

JAPAN

J Tazawa, Niigata University, Niigata, Japan

Introduction

Japan, lying off the eastern coast of Eurasia, has an area of 377 819 km^2; slightly larger than the United Kingdom, and comprises four major islands, Hokkaido, Honshu, Shikoku, Kyushu, and some thousands of smaller islands. Honshu accounts for over 60% of the total area. The Japanese Islands are about 3800 km long, ranging from Kunashiri (45·33·N, 148·45·E) to Yonakuni (24·27·N, 122·56·E) (Figure 1).

The Japanese Islands consist of two arc-trench systems, the North-east Japan and South-west Japan Systems. The NE Japan System, related to the subduction of the Pacific Plate, extends from the Chishima (Kuril) arc-trench to the Izu-Ogasawara (Bonin) arc-trench, through the NE Japan Arc-Japan Trench. The SW Japan System, related to the subduction of the Philippine Sea Plate, extends from the SW Japan Arc-Nankai Trough to the Ryukyu arc-trench. Two back-arc basins, the Okhotsk Sea and the Japan Sea, are placed behind the NE Japan System. Moreover, another new, weakly-developed convergent plate boundary (*see* **Tectonics:** Mountain Building and Orogeny) extends along the eastern margin of the Japan Sea from Sakhalin to central Honshu. The Itoigawa-Shizuoka Tectonic Line (west border of the 'Fossa Magna') is considered to be the boundary between the North American Plate and the Eurasian Plate. Thus Japan exhibits the interaction of four plates, the Pacific, North American, Philippine Sea, and Eurasian Plates. Japan's current tectonic setting has existed only since mid-Tertiary (Early Neogene) time.

Japan was born at the site of a trench, namely, a subduction zone bordering the eastern margin of North China (Sino-Korea) in the Early Ordovician (about 500 Ma). Since then, the Proto-Japan has experienced successive accretions, with episodic high-P/T metamorphism, thrusting, and large scale strike-slip faulting and therefore the geology of Japan is very complicated.

Geological exploration of Japan dates from the pioneer work of E. Naumann in 1875–1885. He first described the geological features of the Japanese Islands, and proposed the 'Fossa Magna' and the Median Tectonic Line. Between 1940 and 1960, T. Kobayashi and M. Minato put forward syntheses on the tectonic development of the Japanese Islands based on the Geosyncline concept. In contrast to the classical syntheses, K. Kanmera, A. Taira, and T. Kimura expressed their new progressive opinions in the 1970 to the 1980s, after the establishment of the Plate Tectonic theory. In the early 1990s, three models of the tectonic framework of the Japanese Islands and its development were proposed: (i) the microcontinent model by K. Ichikawa and others; (ii) the nappe model by Y. Isozaki and S. Maruyama; and (iii) the strike-slip model by J. Tazawa. Since then, an enduring controversy has continued between the three models.

In this article, the tectonic division and evolution of the Japanese Islands are based on the strike-slip model. The 'Economic Geology' of the area is omitted, since most of the mines in Japan are no longer working.

Topography

Japan is very mountainous, with mountains occupying about 71% of the total land area (Figure 2). Altogether, 532 of these mountains are over 2000 m high; Mt Fuji, the highest, rises to 3776 m. The high mountains are concentrated in central Honshu, just west of the 'Fossa Magna', and form three echelon ranges, the Northern, Central, and Southern Japan Alps from north to south. Other ranges coincide with the Palaeozoic to Mesozoic tectonic and regional metamorphic belts or the Quaternary volcanic belts. Trenches are particularly deep along the Pacific Plate boundary, reaching about 7000 m in the Chishima and Japan Trenches, and 8000 m in the Izu-Ogasawara Trench. The Nankai Trough along the Philippine Sea Plate is not as deep as a trench, being less than 6000 m. The Japan Sea, one of the back-arc basins, is roughly divided into northern and southern parts: the Northern Japan Sea has a flat floor, the Japan Basin, of about 3500 m: the Southern Japan Sea has an uneven floor, with the Yamato Bank, Yamato Basin, Tsushima Basin, and Korean Plateau.

Volcanism

About 200 Quaternary volcanoes, of which more than 60 are still active, occur in the Japanese Islands (Figure 3). These volcanoes are distributed in two belts, the East Japan and West Japan Volcanic Belts, and are related to the subduction of the Pacific Plate and the Philippine Sea Plate, respectively. The East Japan Volcanic Belt extends parallel to the Chishima-Japan-Izu-Ogasawara trenches, from the Chishima Islands to the Izu-Ogasawara Islands, through Hokkaido and northern Honshu. Mt Fuji, a typical

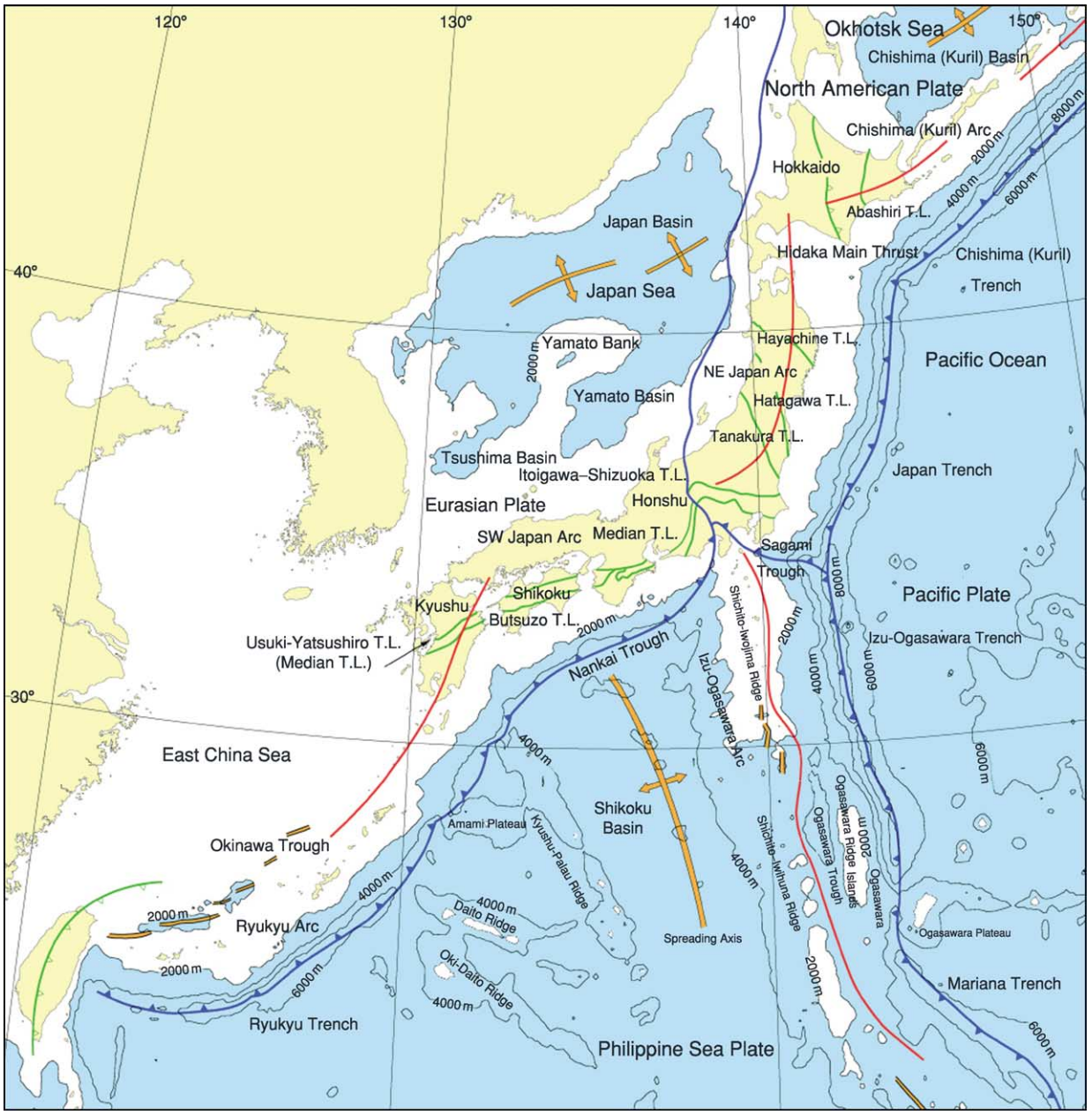


Figure 1 Location map of Japan and the surrounding regions. (After Editorial Committee for the Geology of the Japanese Islands (2002).)

stratovolcano, belongs to this belt. The West Japan Volcanic Belt runs parallel to the Nankai Trough and the volcanoes assigned to this belt occur in the Japan Sea side of western Honshu and Kyushu. Some volcanoes, such as Mt Usu in Hokkaido and Mt Aso in Kyushu, formed large calderas after the eruption of vast amounts of pyroclastic material. In general, the volcanic rock types range from basalt, andesite, dacite, and rhyolite of the tholeiitic series on the east, near the volcanic front, to andesite and basalt of more alkalic or high-alumina basalt types on the west, far from the volcanic front.

Earthquakes

Figure 4 shows the hypocentre distribution of deep-focus earthquakes, over M 2 and deeper than 30 km in Japan. Most of the deep-focus earthquakes occur along the seismic plane (Wadati-Benioff zone) of both arc-trench systems, the NE Japan and SW

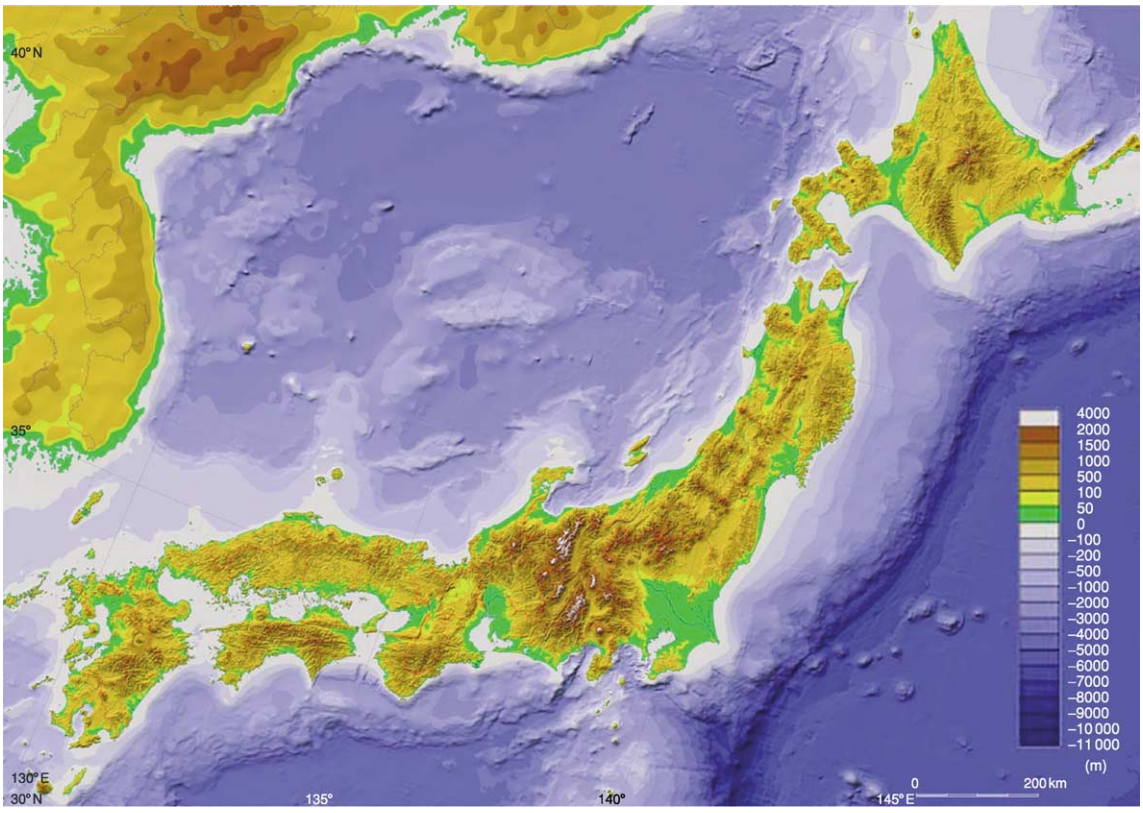


Figure 2 Topographical map of Japan. (After Editorial Committee for the Geology of the Japanese Islands (2002).)

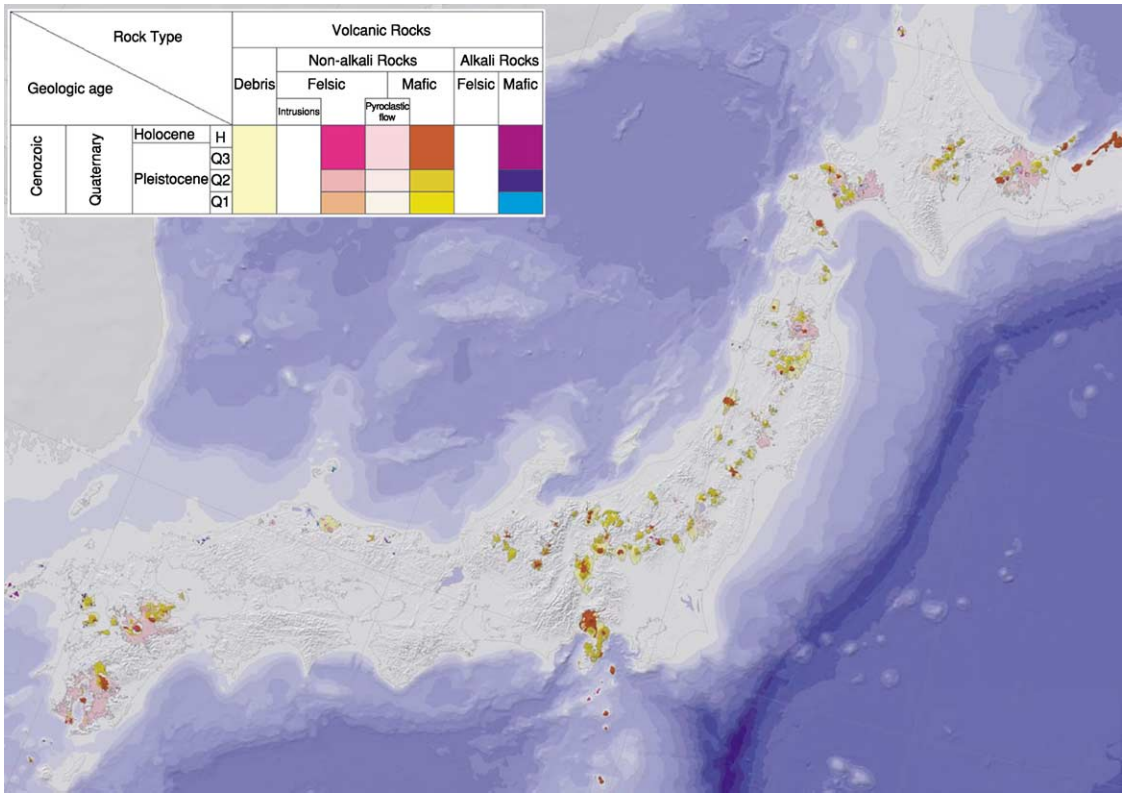


Figure 3 Distribution of Quaternary volcanoes and volcanic rocks in Japan. (After Editorial Committee for the Geology of the Japanese Islands (2002).)

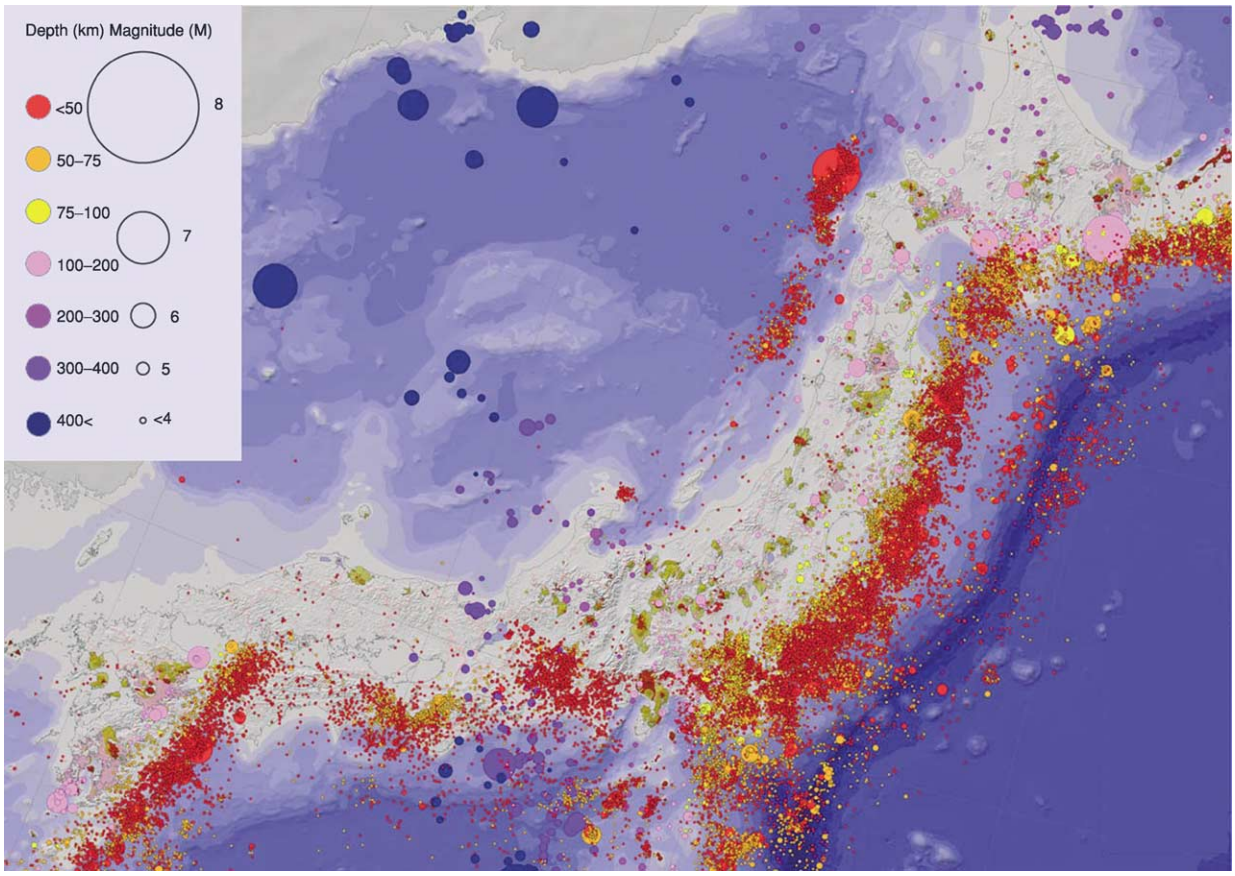


Figure 4 Map showing the distribution of deep-focus earthquakes and Quaternary volcanoes, overprinted on **Figure 3**. (After Editorial Committee for the Geology of the Japanese Islands (2002).)

Japan Systems. Occasionally, shallow submarine earthquakes generate large tsunamis along both the Pacific and Japan Sea sides, particularly in northern Honshu and the Ryukyus. The Great Kanto Earthquake (*M* 8.2; 1 September 1923) killed over 140 000 people and destroyed a third of Tokyo and most of Yokohama. The Meiji Sanrikuoki Earthquake (*M* 8.6; 15 June, 1896) caused a great tsunami of 38 m high, and killed about 22 000 people in the Sanriku coast of Iwate Prefecture, Pacific side of northern Honshu.

In the NE Japan System, the distribution of hypocentres along the seismic plane is smooth and continuous. However, the dipping angle of the seismic plane, i.e., the boundary between the Eurasian Plate and the subducting Pacific Plate varies from place to place. The average angle is about 30° under northern Honshu, but it is steeper under Hokkaido and the Izu-Ogasawara Islands. The maximum depth of deep-focus earthquakes is estimated to be about 700 km. The distribution of hypocentres in the SW Japan System is irregular and they are intermittently distributed beneath central Honshu and Shikoku. The maximum depth in Kyushu is estimated to be about 200 km.

Geological Outline

The geology of Japan (**Figure 5**) has been complicated by long-ranging subduction-related tectonic movement since the Early Ordovician. The tectonic framework of the Japanese Islands is difficult to recognize owing to the thick covering of the Neogene and Quaternary sediments, which include volcanic rocks.

The Pre-Neogene rocks of the Japanese Islands are divided into the following four terranes: (i) the South Kitakami Terrane (Early Ordovician to Late Devonian accretionary terrane, comprising the Hida Gaien, South Kitakami, and Kurosegawa Belts); (ii) the Akiyoshi Terrane (Middle to Late Permian accretionary terrane, comprising the Akiyoshi, Maizuru, Suo, Ultra Tanba, and Joetsu Belts); (iii) the Mino Terrane (Early Jurassic to Early Cretaceous accretionary terrane, comprising the

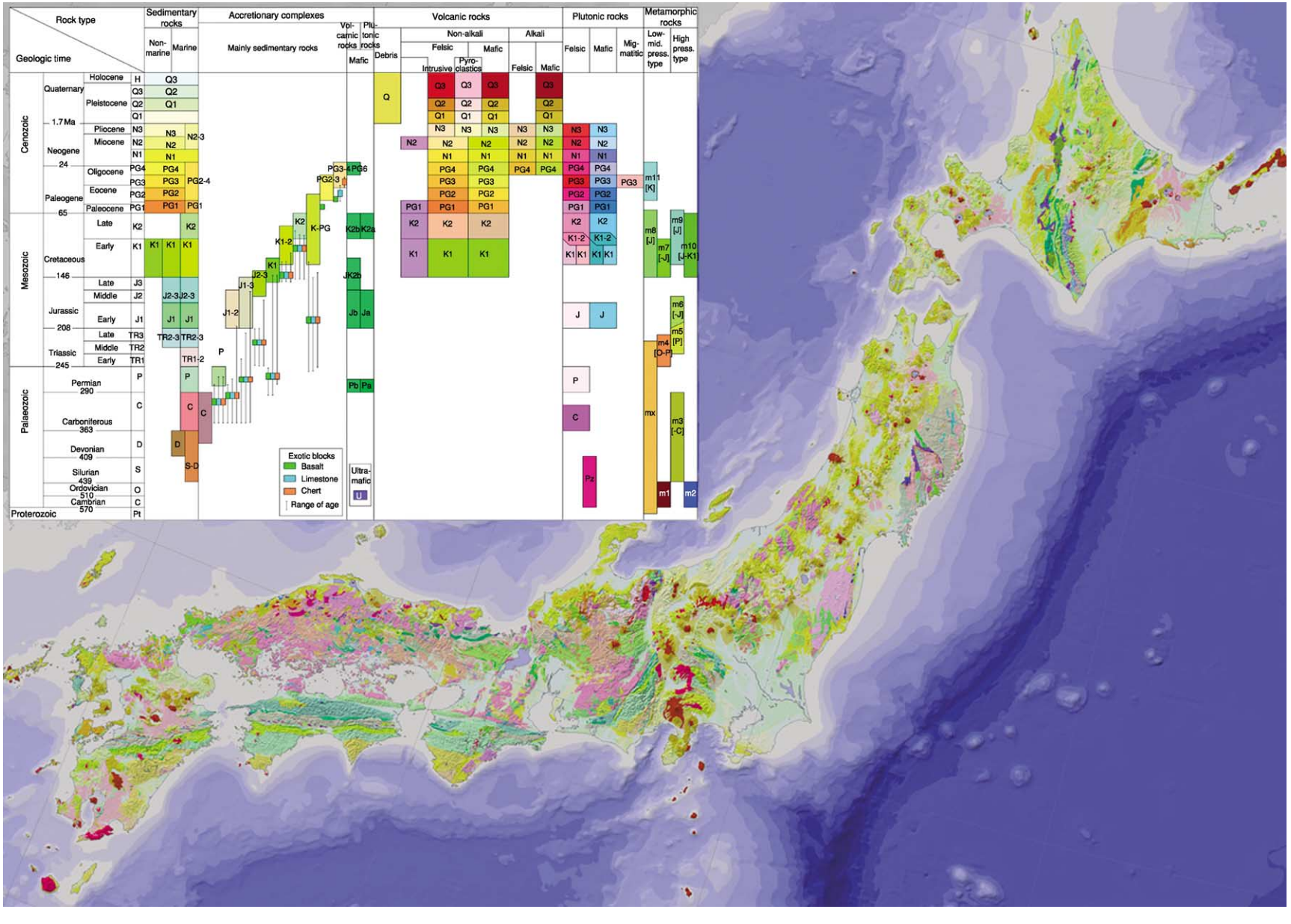


Figure 5 Geological map of Japan. (After Editorial Committee for the Geology of the Japanese Islands (2002).)

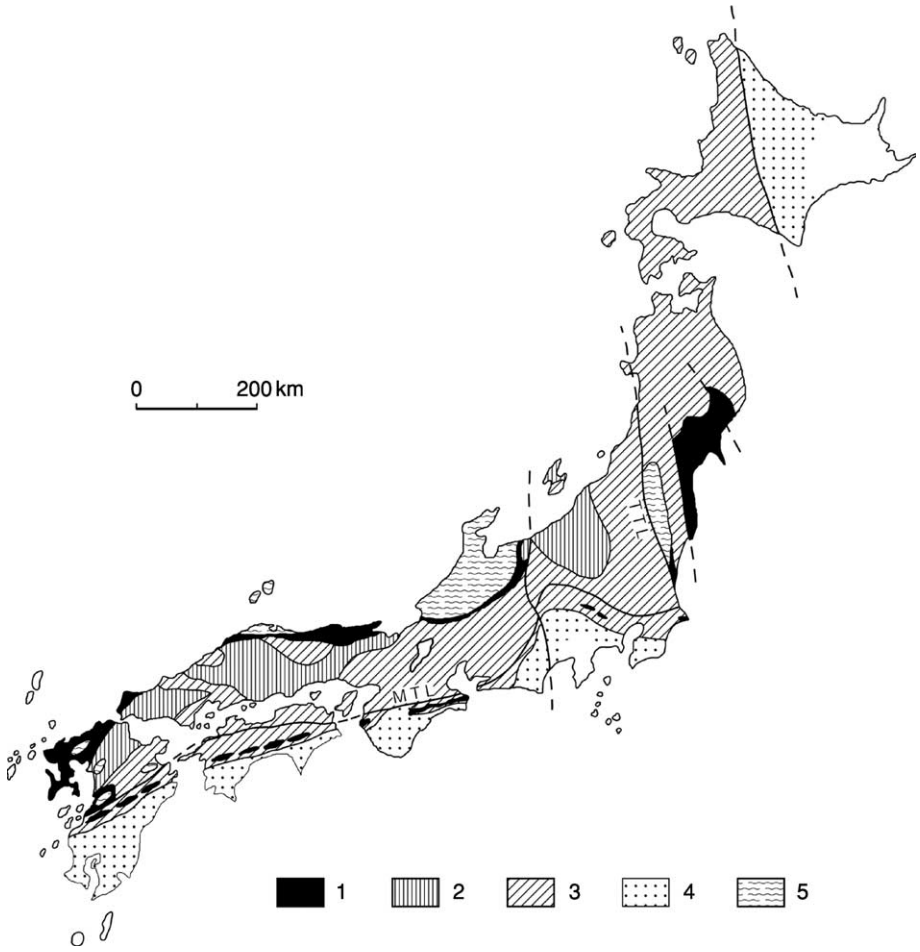


Figure 6 Tectonic map of the Japanese Islands, showing the distribution of pre-Neogene terranes and nappes. MTL: Median Tectonic Line, TTL: Tanakura Tectonic Line, Legend 1: South Kitakami Terrane, 2: Akiyoshi Terrane, 3: Mino Terrane, 4: Shimanto Terrane, 5: Hida-Abukuma Nappe. (After Tazawa (2004).)

Mino, Kamuikotan, North Kitakami, Chizu, Ryoke, Sanbagawa, Northern Chichibu, and Southern Chichibu Belts and the Gozaisho metamorphic rocks); and (iv) the Shimanto Terrane (Late Cretaceous to Early Neogene accretionary terrane, comprising the Shimanto, Hidaka and Tokoro Belts and part of the Idonnappu Belt) (Figure 6). These terranes are arranged in a NE–SW direction, subparallel with the extension of the Japanese Islands, and younging towards the south-east, from the Japan Sea side to the Pacific side. Besides the four terranes, several nappes (Hida Nappe, Abukuma Nappe, Higo Nappe, etc.) in the Hida-Abukuma Nappe, consisting of Palaeozoic and Mesozoic gneisses, schists, amphibolites, crystalline limestones, and granites, all of which experienced medium-pressure metamorphism in the Triassic (250–220 Ma), thrust over the South Kitakami, Akiyoshi, and Mino Terranes towards the east to south-east. These metamorphic rocks were probably derived from the Qinling-Dabie Suture Zone, a collision zone between the North China and South China blocks.

Tectonic Evolution

The tectonic framework of the Japanese Islands was formed through long-term subduction onto mostly the eastern margin of North China, since the Early Ordovician. The Silurian to Permian marine faunas (brachiopods, cephalopods, and trilobites) and the Permian floras indicate affinities to those of central Asia (Tien Shan, north Xinjiang), Inner Mongolia, North-east China, and Far East Russia (Primorye), except for the Permian Tethyan-Panthalassan fusulinoid and brachiopod faunas from exotic limestone blocks in the Jurassic melange

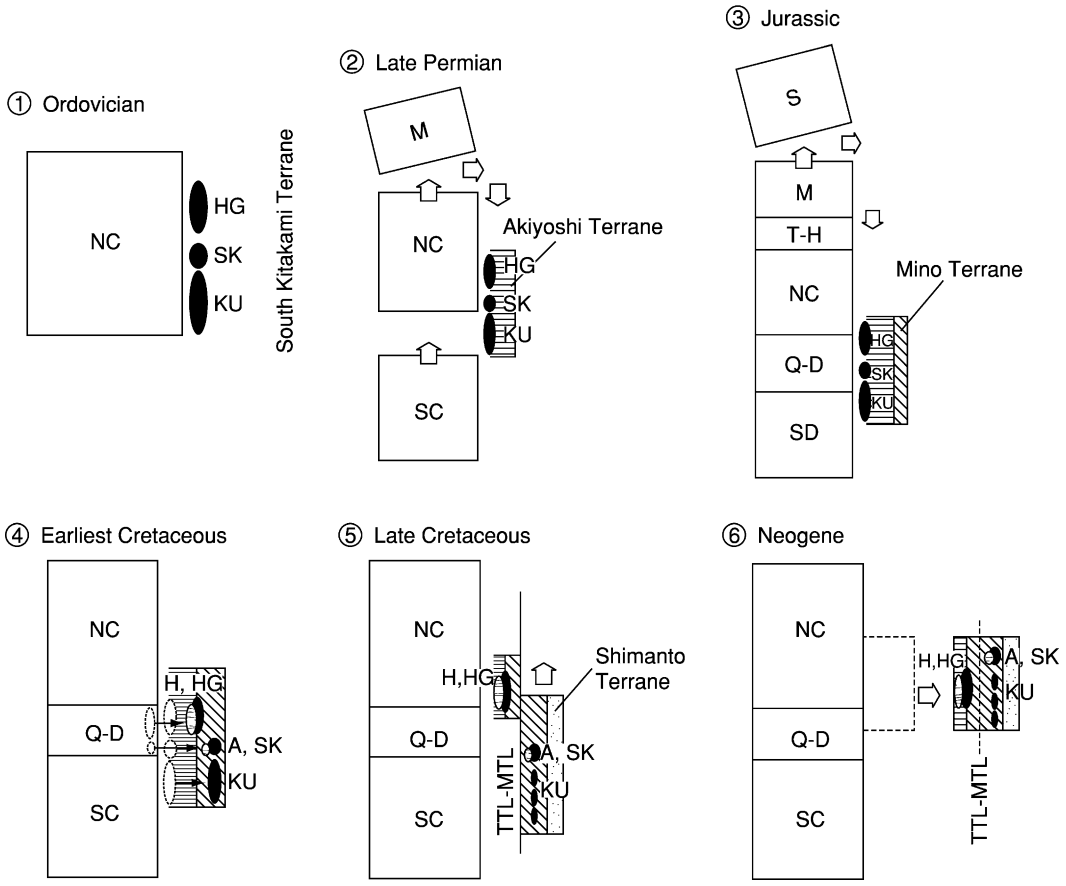


Figure 7 Generalized kinematic model for the tectonic development of the Japanese Islands in Ordovician to Neogene times. HG: Hida Gaien, KU: Kurosegawa, M: Mongolia, NC: North China, Q-D: Qinling-Dabie, S: Siberia, SC: South China, SK: South Kitakami, T-H: Tienshan-Hinggan, TTL-MTL: Tanakura-Median Tectonic Line. (After Tazawa (2004).)

of the Mino Terrane. For example, the Permian brachiopod faunas of the Hida Gaien and South Kitakami Belts in the South Kitakami Terrane are characterized by a mixture of Boreal and Tethyan elements, and are similar to those of Inner Mongolia, North-east China, and Primorye. Figure 7 shows the tectonic evolution of the Japanese Islands together with some continental blocks in east Asia.

The oldest trench, related to the accretion of Ordovician to Late Devonian sediments in the Hida Gaien, South Kitakami, and Kurosegawa Belts, all of which belong to the South Kitakami Terrane (Figure 7(1)), changed to a continental shelf in the Early Carboniferous. Subsequently, a new subduction zone developed in front of the continental shelf in the Middle Permian. The Middle to Late Permian sediments of the Akiyoshi Terrane accumulated in this trench (Figure 7 (2)). After that, the South Kitakami Terrane (together with the Akiyoshi and Mino Terranes) moved towards the south along the dextral strike-slip faults on the eastern margin of North China in the Late Permian to the Late Jurassic. This dextral strike-slip movement may have been caused by the collision of the Mongolian and North China blocks in the Late Permian to Triassic (Figure 7(2)) and the Siberian and Mongolian blocks in the Jurassic (Figure 7(3)). Next to the dextral strike-slip faulting, a thrusting towards the east took place in this region in the latest Jurassic to earliest Cretaceous. This thrusting resulted in the emplacement of several nappes, such as the Hida Nappe, South Kitakami Nappe, and Kurosegawa Nappe (Figure 7(4)). The single tectonic unit, therefore, was then separated into three belts, the Hida Gaien Belt in the Inner Zone (continental side), and the South Kitakami and Kurosegawa Belts in the Outer Zone (oceanic side). Then a rapid northward motion of the Izanagi Plate (*ca.* 20–30 cm year) caused a large-scale sinistral displacement (1500–2000 km) of the Outer Zone, including the South Kitakami and

Figure 8 Reconstruction of Early Cretaceous Japan. Solid arrow shows the direction of strike-slip motion in that time. Shading as for **Figure 6**. (After Tazawa (2004).)

Kurosegawa Belts along the Tanakura Tectonic Line, Median Tectonic Line and some associated faults in the Early Cretaceous to Palaeogene, mainly in the Early Cretaceous (**Figures 7(5) and 8**). The pre-Cretaceous arrangement of the Joetsu, Hida Gaien, South Kitakami, and Kurosegawa regions in this order (from north to south) is supported by the biogeographical evidence from the Middle Permian brachiopod faunas, as well as the Late Jurassic to Early Cretaceous floras. This spatial arrangement of the pre-Cretaceous Japan was greatly changed by the sinistral strike-slip faulting in the Early Cretaceous to Palaeogene. From the Early Jurassic to Palaeogene, the Mino and Shimanto Terranes accreted on to the subduction zone at the front of the Proto-Japan. Finally, the opening of the Japan

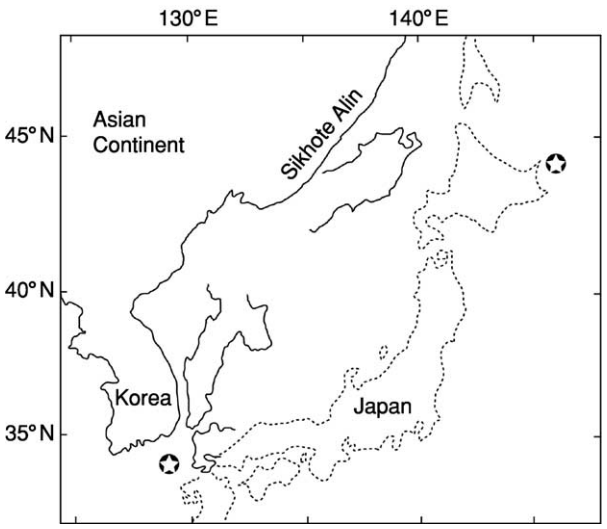


Figure 9 Reconstruction of Early Neogene Japan, showing the 'Double door opening model' of NE Japan and SW Japan. Stars show the rotation pivots for NE Japan and SW Japan, respectively. (Modified from Otofuji *et al.* (1985).)

Sea caused a clockwise rotation of South-west Japan and a counterclockwise rotation of North-east Japan in the Early Neogene (Miocene) (**Figure 7(6) and 9**).

See Also

China and Mongolia. **Russia**. **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges; Mountain Building and Orogeny.

Further Reading

Editorial Committee for the Geology of the Japanese Islands (ed.) (2002) *Computer Graphics: Geology of the Japanese Islands, CD-ROM Version*, Tokyo: Maruzen.

Hashimoto M (ed.) (1991) *Geology of Japan.* Tokyo: Terra Scientific Publishing Company.

Ichikawa K, Mizutani S, Hara I, Hada S, and Yao A (eds.) (1990) *Pre-Cretaceous Terranes of Japan*, Publication of IGCP Project No. 224. Osaka: Nippon Insatsu Shuppan.

Isozaki Y (1996) Anatomy and genesis of a subduction-related orogen: A new view of geotectonic subdivision and evolution of the Japanese Islands. *The Island Arc* 5: 289–320.

Isozaki Y and Maruyama S (1991) Studies on orogeny based on plate tectonics in Japan and new geotectonic subdivision of the Japanese Islands. *Journal of Geography* 100: 697–761. (In Japanese).

Kanmera K, Hashimoto M, and Matsuda T (eds.) (1980) *Geology of Japan, Earth Sciences* 15. Tokyo: Iwanami Shoten. (In Japanese).

Kimura T, Hayami I, and Yoshida S (1991) *Geology of Japan.* Tokyo: University of Tokyo Press.

Otofuji Y, Matsuda T, and Nohda S (1985) Palaeomagnetic evidence for the Miocene counter-clockwise rotation of Northeast Japan – rifting process of the Japan Arc. *Earth and Planetary Science Letters* 75: 265–277.

Taira A and Tashiro M (eds.) (1987) *Historical Biogeography and Plate Tectonic Evolution of Japan and Eastern Asia*. Tokyo: Terra Scientific Publishing Company.

Tazawa J (1993) Pre-Neogene tectonics of the Japanese Islands from the viewpoint of palaeobiogeography. *Journal of the Geological Society of Japan* 99: 525–543. (In Japanese).

Tazawa J (2002) Late Paleozoic brachiopod faunas of the South Kitakami Belt, northeast Japan, and their paleobiogeographic and tectonic implications. *The Island Arc* 11: 287–301.

Tazawa J (2004) The strike-slip model: A concept on the tectonic evolution of the Japanese Islands. *Journal of the Geological Society of Japan* 110: 503–517. (In Japanese).

JUPITER

See **SOLAR SYSTEM: Jupiter, Saturn and Their Moons**

LAGERSTÄTTEN

S E Gabbott, University of Leicester, Leicester, UK

Introduction

This article will describe in detail Fossil Lagerstätten, with a special emphasis on conservation Lagerstätten – deposits containing exceptionally preserved fossils. Concentration Lagerstätten are deposits where fossils are especially abundant and the processes that operate to produce these types of fossil horizons will be described. Fully articulated skeletons and the preservation of the soft parts of organisms, which are usually rapidly lost to decay after death, both constitute examples of exceptional preservation. In order to create these extraordinary fossil deposits, scavengers must be prohibited from disarticulating and consuming the carcass and bacterial putrefaction of soft tissues must be circumvented. Authigenic mineralization of soft tissues produces fossils with the highest degree of fidelity known, where even subcellular features can be recognised. Mineralization processes, including replacement by apatite, pyrite, clays, and silica are documented using examples from conservation Lagerstätten. Even exceptionally preserved fossils show some degree of decay-induced morphological change, so that decay experiments have proved crucial in their accurate interpretation. In addition, decay experiments have determined some of the physical and chemical factors, such as degree of oxygenation, pH, temperature, and soft tissue composition which all affect preservation. Exceptionally preserved faunas not only produce some of the world's most spectacular fossils, but they are also critical in the reconstruction and understanding of ancient life and evolutionary processes. Such deposits may be temporally constrained because some geological periods are replete in examples compared with others; the putative explanations for such temporal trends are explored.

Lagerstätten, a German word, was originally used to describe mineral and ore deposits of economic worth. Adolf Seilacher first coined the term Fossil Lagerstätten (singular Lagerstätte) to describe a body of rock that is unusually rich in palaeontological information because, either the fossils are so well preserved and/or the fossils are so abundant that they warrant exploitation and scientific attention. The distinction between exceptionally preserved fossils and exceptionally numerous fossils has led to the terms conservation Lagerstätten and concentration Lagerstätten being applied, respectively. It is the conservation Lagerstätten that have received the most attention because they contain fossils that are extraordinary in the quality of their preservation. They may contain completely articulated skeletons or soft-bodied animals composed of nonmineralized, so-called soft-parts and soft-tissues. Fossil conservation Lagerstätten include some of the world's most celebrated deposits, such as the Cambrian Burgess Shale of Canada and the Jurassic Solnhofen Lithographic Limestone of Germany. Exceptionally preserved faunas are confined to particular environmental settings, where a number of processes must occur in order to fossilise the nonmineralised tissues. In recent years, some of these processes have become well documented and will be described below; however, there are still many unanswered questions relating to exceptional preservation.

It should be appreciated that there is no absolute Criterion to discriminate a Lagerstätte from a fossiliferous horizon. However, this is consistent with the notion that Lagerstätten are not distinct rock units, but are end-members of a range of sedimentary facies in which fossils are preserved.

Concentration Deposits

In these deposits it is the sheer abundance of fossils, although not often very well preserved, that is important and it is vitally important to reveal how such profusion occurred. Concentration may be a real reflection of community ecology or behaviour, or concentration may be a sedimentological artefact produced by time-averaging and physical processes, such as winnowing, which act to distillate skeletal remains. Two types of concentration deposits are considered in more detail below.

Stratiform Concentration Deposits

Where shelly fossils are locally abundant and form dense concentrations they may be referred to as 'shell beds' or 'coquinas'. They are formed in a variety of ways, reflecting a range of processes and time-scales. For example, they may form from a sudden, rapid influx of shells from a mass mortality, or from the slow accumulation of shells over many years during times of low deposition rates. An example to illustrate concentration by multiple, physical processes is the famous ammonite coquina of the Middle Jurassic (Normandy Coast, France) where beds contain different ammonites representing about 2 million years. The ammonites clearly do not reflect a single mass

mortality event, rather sedimentation to preserve the shells (so that they were not corroded) and probably several mixing events during storms that concentrated and yet did not destroy the shells.

Bone bed coquinas comprise an unusually rich concentration of phosphatic vertebrate remains, including bones, teeth, scales, and coprolites. Remains are not articulated and commonly include terrestrial and marine components. There are several models proposed to explain bone bed genesis. Winnowing involves episodes of erosion of the seafloor to rework and concentrate phosphatic material, which was previously more widely dispersed. Condensation involves extremely low sedimentation rates so that the input of phosphatic material is not diluted by sediment, but builds up on the seafloor over long time periods. Another model suggests that other biogenic material, composed of carbonate, may be dissolved, resulting in the apparent concentration of phosphatic material. Finally, in the transgressive lag model, a transgressive sea picks up phosphatic clasts from previously formed sediments and concentrates them into a basal bone bed. Examples of bone beds include the Westbury Formation of the Rhaetian Penarth Group, Wales and the Silurian of Ludlow, England.

Concentration Traps

In terrestrial and marine environments, fossils may be concentrated in protected environments such as cavities. Terrestrial cave and fissures constitute the most spectacular of these deposits where animals are concentrated either because they used the cavity as a dwelling, or because the cavity formed a death trap, successively killing and preserving animals over time. The Early Holocene Shield Trap Cave (Montana, USA) records many disarticulated bison bones and was formed in a bell-shaped limestone cave after roof collapse.

Conservation Deposits

The Importance of Exceptional Preservation

Entirely soft-bodied organisms or the soft parts of organisms usually rapidly disappear after death, due to decay processes and, consequently, most fossils comprise the remaining hard, mineralized parts of animals. This has two consequences: firstly, entirely soft-bodied animals are usually not preserved (except in conservation Lagerstätten); and secondly, if the animal has hard parts, these become disarticulated as the soft tissues no longer support them. Disarticulation may pose significant problems when trying to understand ancient life. Plants in particular undergo disarticulation when transported and because their component parts behave differently hydrodynamically their leaves, pollen, seed, fruit, bark, stems, roots, etc. often become widely separated, making it difficult to reconstruct the whole plant. Animals also suffer disarticulation with some, such as echinoderms, being more susceptible than others.

Conservation Lagerstätten (exceptionally preserved faunas) provide palaeontologists with a unique and unrivalled window through which to view the biology, ecology, and evolution of life. Up to two-thirds of modern shallow water marine communities are composed of entirely soft-bodied animals (*see* **Palaeoecology**). So conservation Lagerstätten are vital because they preserve a more complete picture of community diversity. However, it is important to determine how much of the original biota is represented in a fossil assemblage; in other words, is a particular taxon absent because it was not part of the community, or was it just not preserved? Understanding the taphonomy (processes of fossilization) of different conservation Lagerstätte, individual fossil specimens, and a range of different tissues types, is crucial in gauging preservational bias.

The preservation of soft tissues is also extremely helpful when trying to interpret animal affinities. Many animals have hard parts that are uninformative as to the underlying soft part anatomy. For example, conodonts (*see* **Microfossils:** Conodonts) (microscopic, phosphatic, tooth-like elements) were known for over 100 years before their affinities were resolved. This is because, although conodont elements are made of apatite and therefore have a good fossil record, the rest of the animal, to which these conodonts belonged, had never been found. Fortunately, in 1982, a set of conodont elements was reported with associated soft tissues from the Carboniferous Granton Shrimp Bed near Edinburgh. The analyses of these fossils led to the interpretation that conodonts were in fact vertebrates with an eel-like form, a notochord, V-shaped muscle blocks, a caudal fin, and that the conodont elements were located in the animal's oral cavity and functioned as a feeding apparatus.

Finally, the preservation of soft-bodied organisms provides palaeontologists with the evidence to explore biological phenomenon such as the Cambrian explosion, the nature of early metazoans, extinctions, and radiations.

Death, Decay, and Destruction

After death, several processes usually occur which would normally inhibit the potential of an animal or plant from becoming a fossil, and this is particularly the case for nonmineralised tissues. Necrolysis is the organic modification to carcasses, and occurs through autolysis, which is the breakdown of cells by their

own enzymes, scavenging which is the consumption of carcasses by macroscavengers, decomposition which is degradation by fungi, and putrefaction which is the destruction by bacteria. Biostratinomy describes the physical alterations that affect a carcass prior to burial, and after burial, diagenesis includes the chemical and mechanical changes that take place to fossils. Taphonomy is the study of all of these processes of fossilization, from death of the animal to its discovery as a fossil. In order for exceptional preservation to occur, the taphonomy must be exceptional to bypass decay and ultimately destruction – the normal fate of nonmineralized tissues.

Decay Experiments

Decay experiments have proved to be an extremely valuable tool to document and understand decay-induced morphological changes and the processes of fossilisation. Decay affects morphology and it is useful to compare the morphological features of a fossil with similar morphological features that have undergone decay, rather than the same features from an extant, living animal. Decay experiments have been undertaken on a range of taxa including vertebrates, from entirely soft-bodied hagfish (Figure 1) through to deer carcasses, invertebrates such as cnidarians, various worms, various echinoderms, cephalopods, crustaceans, insects, cephalochordates, and prokaryotes. In addition to assessing decay-induced morphological changes, these experiments have investigated the role of different factors (pH, oxygen concentration, temperature, salinity, and even microbe type) in the process of decay and fossilization. Decay experiments have provided important insights in to why some organic tissues survive decay and how other tissues become replaced very rapidly by authigenic minerals. In the future, understanding decay and fossilization processes may be used to predict the environmental conditions, and thus the type and location of sediments that may favour exceptional preservation.

Conservation Traps

There are two broad groups of conservation deposits: conservation traps and stratiform conservation deposits.

Conservation traps are deposits that are generally very geographically localised and are usually temporally restricted. They occur where animals and plants are trapped in a setting where macroscavengers are unable to access the remains, and the destruction by bacteria and enzymes is inhibited. Often the processes responsible for preservation in conservation traps are akin to those we employ to preserve our food, such as freezing, drying, and pickling. Examples include deep frozen mammoths from the Pleistocene found in clefts in the glacial permafrost in Siberia, mummified, desiccated ground sloths in caves in South America, and insects trapped in amber resin. Perhaps the most famous conservation trap is the Pleistocene Rancho La Brea tar pits of California, USA, which have preserved three million fossils including mammals, birds, fish, reptiles, invertebrates, and plants. In this respect, Rancho La Brea tar pits are a concentration, conservation trap. Conservation traps are temporally restricted because if conditions change, for example, if a carcass is re-hydrated or thawed, bacteria will resume their relentless destruction of the organic remains.

Stratiform Conservation Deposits (Table 1)

Stratiform conservation Lagerstätten are deposits where incomplete necrolysis occurs throughout the

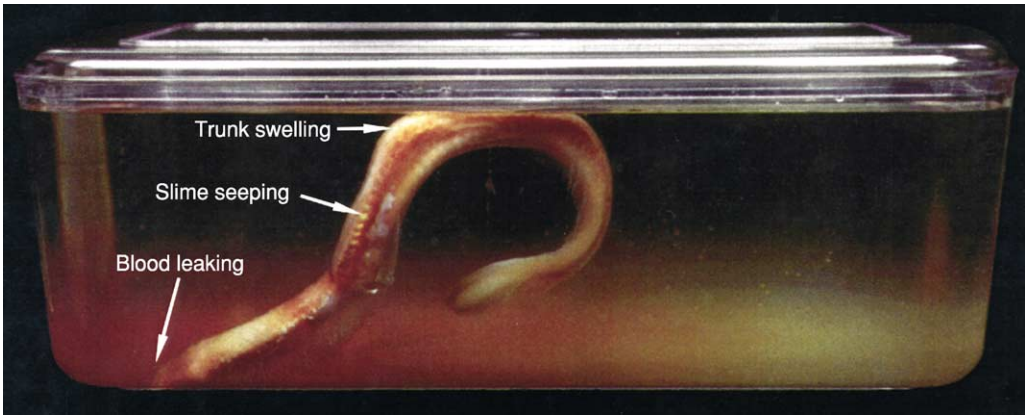


Figure 1 Decay experiment showing a hagfish in normal marine water just 6 days after death. Decay gases in the body cavity have resulted in the carcass floating. On day 7 of this experiment, the skin tore away and began to sink; disintegration of the carcass occurred shortly afterwards. This experiment demonstrates that exceptional preservation requires that a carcass be rapidly buried, or kept on the substrate by other means, such as by microbial mats, and that exceptional preservation must occur very rapidly after death. (Experiment and photographs by Dr K Freedman.)

Table 1 A small selection of conservation Lagerstätten with age, location, and brief comments on the environmental conditions and preservation of soft parts (sp.)

Lagerstätten	*Age*	*Location*	*Brief comments*
Doushantuo Formation	Latest Proterozoic	South China	One deposit in carbonaceous shales, another in phosphorites phosphatisation of sp.
Chengjiang	Early Cambrian	Yunnan Province, China	Burgess Shale-type fauna, microturbidites various mineral replacement of sp. and possible organic remains
Burgess Shale	Middle Cambrian	British Columbia, Canada	Burgess Shale-type fauna, obrution and anoxia, kerogenised organic films and clay mineral replacement of sp.
Orsten	Late Cambrian	Southern Sweden, Germany and Poland	Anthraconite limestone nodules in bituminous shale, apatite (phosphatic) coating and replacement of sp.
Soom Shale	Late Ordovician	Cape Province, South Africa	Euxinic and anoxic bottom and porewaters, clay mineral replacement of sp.
Herefordshire	Silurian	Herefordshire, UK	Calcite nodules in volcanic ash deposit after decay three-dimensional animal void infilled by calcite
Hunsrück Slate	Early Devonian	Western Germany	Obrution by microturbidites pyrite replacement of sp.
Rhynie Chert	Early Devonian	Eastern Scotland	Plants and arthropods engulfed in Si-rich hotspring water silicification of sp.
Mazon Creek	Pennsylvanian Carboniferous	Illinois, USA	Rapid burial in oxygen depleted sediment and siderite nodules sp. represented as highly compressed, light-on-dark impressions
Monte San Giorgio	Middle Triassic	Southern Switzerland	Bottom waters anoxic, but surface waters normal marine rare phosphatised sp., mostly articulated skeletons
Posidonia Shale	Lower Jurassic	Holzmaden region, Germany	Anoxic bottom waters, soupy sediment and occasional sediment blanketing, sp. of belemnites, cephalopods
Solnhofen Limestone	Upper Jurassic	Southern Germany	Hypersaline, oxygen-depleted bottom waters, soupy sediment obrution and microbial mats. Soft parts mostly impressions rare phosphatisation of sp. and organic residues
Santana Formation	Lower Cretaceous	NE Brazil	Hypersaline waters, soupy sediment, and mass mortalities many carbonate concretions and sp. phosphatised
Grube Messel	Mid Eocene	West Germany	Quiet, anoxic lacustrine waters silhouettes of sp. made of autolithified, sideritic bacterial films
Baltic Amber	Late Eocene-Early Oligocene	Baltic, NW Europe	Amber is fossilized resin from trees, which traps arthropods; resin stops bacterial and fungal decay and acts as desiccant and antibiotic

sediment layer(s) rather than being constrained to special, localized settings, as is the case for conservation traps.

Scavenging by macro-organisms is a ubiquitous event in most environments and leads to disarticulation, and sometimes wide dispersal of the animal hard parts, and the removal of soft parts. Terrestrial and aquatic environments abound with macroscavengers and they may devour soft tissues in surprisingly little time. For example, a fish carcass introduced to a cold (4°C), dysaerobic seabed off the coast of California was stripped of soft parts and disarticulated in 2–3 days by brittle stars. Clearly, one of the most important prerequisites for exceptional preservation is the precludement of macroscavengers from carcasses and plants. This may be accomplished in a variety of ways, some of which are illustrated below, using examples from fossil conservation Lagerstätten.

Obrution Catastrophic, rapid burial, by turbidites or tempestites, may help to preserve nonmineralised tissues because it protects the animal (which may be buried dead or alive) from macroscavengers and bioturbators, (*see* **Trace Fossils**) and it tends to induce anoxia (see below). However, whilst there are fewer bacteria with depth, and anaerobic bacteria are less efficient at breaking down organic carbon, a rapidly buried carcass still requires other processes to occur to inhibit or circumvent the decay of organic remains. Sediment smothering often produces an ecological bias because it will affect bottom-living organisms more than nektonic ones. In the Phyllopod Bed of the Middle Cambrian Burgess Shale (British Columbia, Canada) there is evidence to suggest that much of the biota was catastrophically buried, such as graded beds and variable orientations of specimens relative to bedding (Figure 2). Furthermore, Phyllopod Bed sediments are parallel laminated (i.e., undisturbed) suggesting that, owing to anoxia, conditions in the sediment porewaters were inimical to bioturbators and scavengers.

Soupy substrates Burial owing to high rates of sedimentation is not the only way of rapidly covering a

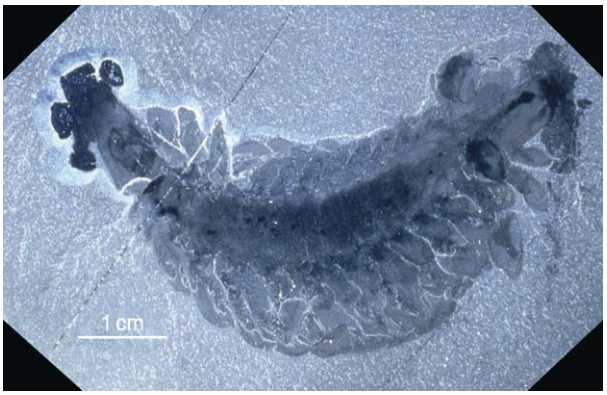


Figure 2 *Opabinia* from the Middle Cambrian Stephan Formation, Burgess Shale, Canada. Only three of the five eyes are evident and the flexible proboscis is still buried, and may be curved back beneath the head. The trunk consists of 15 segments, each bearing a pair of lateral lobes.

carcass and thus prohibiting many macroscavengers. If the sediment is soft, or even soupy, carcasses can sink into the sediment and thus be buried. Articulated vertebrate skeletons and the preservation of the soft parts of ichthyosaurs of the Jurassic Oxford Clay (UK) are suggested to have occurred because carcasses sunk into a soupy seafloor. This example is fascinating because areas of vertebrate carcasses that protruded above the soupy sediment were often encrusted and bored, suggesting that vertebrate skeletons provided 'benthic islands' of hard substrate, amid the otherwise inhospitable, soupy seafloor.

Stagnation, anoxia Anoxia is the absence of dissolved oxygen and results if the rate of carbon deposition exceeds that of oxygen supply. Euxinic conditions describe waters that are anoxic and sulphidic. Benthic organisms are prohibited from living in such environments as are bioturbators, resulting in sediments with parallel (i.e., undisturbed) laminations. Carcasses falling into such bottom water conditions have an enhanced potential for exceptional preservation for several reasons. Firstly, macroscavengers would not be able to devour and disarticulate the carcass. Furthermore, anaerobic bacteria putrefy soft tissues more slowly and so increase the time for rapid mineralization (see below) to replace the tissues. In addition, anaerobic bacteria may not be able to utilise very recalcitrant organic matter, which then has the potential to survive into the fossil record as an organic residue. A classic example of a deposit where anoxia and stagnation led to exceptional preservation is the Jurassic Posidonia Shales (Holzmaden, Germany). Here black, bituminous shales containing extremely rare benthonic organisms were overlain by an upper water body rich in oxygen and life. The fauna includes plants and many invertebrates, but the shale is most famous for its complete vertebrate skeletons such as sharks, ichthyosaurs, crocodiles, and pterosaurs. However, anoxia cannot account for the occurrence of the articulated crinoids, because if they lay unprotected on the seafloor, albeit an anoxic one, they would disarticulate in a matter of days. It has been suggested, therefore, that rapid sediment smothering also occurred in the Posidonia Shales.

It should be appreciated that exceptional preservation in several conservation Lagerstätten can be attributable to more than one environmental circumstance. The Jurassic Solnhofen Limestone (Bavaria, Germany) was deposited in a basin where the stratified bottom waters were inhospitable to life owing to high salinity and low oxygen content. Periodic storm events washed animals into the basin, which were then rapidly buried by carbonate ooze. So rapid burial, hypersalinity, and anoxic stratified waters may all have contributed to exceptional preservation. Moreover, on the basis of the anatomy of the commonly preserved *Saccocoma*, a crinoid, it has been suggested that periodically the bottom waters were not saline, but that the sediment was soupy, allowing carcasses to sink into the limey muds.

Bacteria

Even where organisms are protected from macroscavengers, their soft parts are still vunerable, owing to the action of autolytic enzymes and bacteria. In the presence of oxygen, bacteria decompose soft tissues rapidly. Once oxygen has been used up, decay proceeds anaerobically. In freshwater environments, anaerobic decay is dominated by nitrate reduction and methanogenesis, whereas in the marine environment sulphate reduction and methanogenesis dominate. These reactions all lead to the decomposition of soft tissues, although there is evidence to suggest that anaerobic decay proceeds more slowly and is less efficient than aerobic decay. Decay may be inhibited by microbial mats (*see* **Biosediments and Biofilms**) because these mats effectively limit the diffusion of toxic bacterial metabolic by-products away from the site of decomposition, and the influx of oxygen and sulphate (marine environments) needed to fuel the bacteria. Mats may also aid soft part preservation by trapping organisms on the seafloor and creating a 'closed' microenvironment that may well favour rapid mineralization processes (see below). Under certain conditions, bacterial decomposition may also be retarded in fine-grained mudrocks by the prevalence of clay minerals. It has been suggested that the digestive enzymes used by bacteria in decomposition may become adsorbed onto specific clay minerals and thus rendered inert. Such a model has been proposed

to explain the preservation of organic films in the Burgess Shale (Figure 2). However, some of the highest fidelity of soft-part preservation comes about when the soft tissues are mineralized before they are decomposed.

Soft Tissue Mineralization

Nonmineralized tissues display a spectrum of resistance to decay depending on their physical and chemical makeup. Decay-resistant (sometimes called recalcitrant) tissues, such as chitin of arthropod cuticles and woody plant tissues, may retain their organic composition as fossils. Decay-susceptible (sometimes called labile) tissues cannot survive unless authigenic minerals rapidly replace them, a process sometimes referred to as permineralisation. Mineralization can be extremely rapid, occurring within days, weeks, and months of death. Authigenic mineralization is controlled by a number of factors, including the organic makeup of the soft tissues, the geochemistry of the sediment, the pH and Eh of the porewaters and the concentrations of mineral forming ions. In addition, bacteria may mediate mineralization by breaking bonds, speeding up reactions, and concentrating authigenic mineral ions on to the tissue surfaces – the exact nature of their participation is not yet understood. In some deposits it is the bacteria themselves that autolithify, and they can pseudomorph details of the underlying animal morphology. For example, in the Eocene oil shale of Grube Messel (Germany) exquisite details, such as single hairs or the barbules of feathers, have been preserved as layers of lithified bacteria.

Only a few minerals are known to replace soft tissues, the most important of which are apatite, pyrite, clay minerals, and silica.

Apatite

Apatite is the mineral most often involved in replacing soft tissues and it does so with the highest degree of fidelity. Details, down to the subcellular level, such as Cretaceous fish cell nuclei sitting on muscle fibres, have been recorded and such fine detail is possible because the apatite crystallites can be very small (commonly <30 nm). Apatite mineralization requires a sufficient concentration of phosphate, which may be released from the decaying animal, or concentrated in the surrounding sediment. Decay experiments have also shown that slightly acidic pH favours apatite authigenesis over carbonate authigenesis, which does not have the same potential to replace soft tissues. Experimental phosphatisation (the replacement of soft tissues by apatite) and the analyses of experimentally produced and fossil apatite

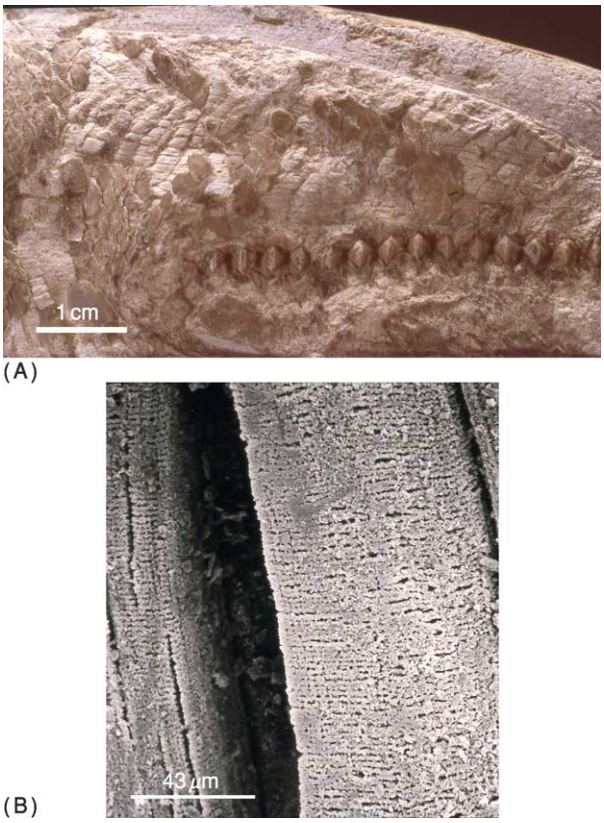


Figure 3 Colour photograph shows the elopomorph fish *Notelops* sp. with phosphatised striated muscle tissue (white fibrous material). Black and white photograph shows phosphatised muscle fibres from another *Notelops* sp, obtained using a Scanning Electron Microscope. Both specimens are from the Early Cretaceous Romualdo Member, Santana Formation, NE Brazil. (Photographs by Dr D Martill.)

textures has greatly improved our understanding of apatite authigenesis. Some of the more spectacular examples of soft tissues replaced by apatite include: delicately preserved tiny arthropods in nodules from the Upper Cambrian 'Orsten' (northern and central Europe) and fish muscles (Figure 3), gill filaments, and eggs from the Cretaceous Santana Formation (Brazil).

Pyrite

Pyrite is commonly formed in the sediment of anoxic black shales but it is relatively rarely involved in mineralizing animal soft tissues. Although it is quite commonly associated with the preservation of plant remains where it may preserve cellular details (e.g., Eocene London Clay of England). Pyrite crystals are generally larger than apatite and clay minerals, and this constrains the fidelity of pyritised soft tissues. Pyrite commonly fills voids or coats the gross morphology or outline of tissues and so delicate, rapidly decayed tissues, such as muscle are not known to be

Figure 4 A specimen of the arthropod *Leancholia* from the Early Cambrian Chengjiang Formation, China, 2.3 cm in length. (Photograph by Dr Derek Siveter.)

pyritised. Replacing animal tissues by pyrite requires that pyrite must grow on the carcass rather than throughout the sediment. In order to do this the sediment must have high concentrations of iron in porewater solutions and low total organic matter. This allows sulphide generated from seawater sulphate by the decay of soft tissues to be trapped instantly by formation of iron sulphide minerals within, or very close to the carcass. An example of soft part pyritisation in seen in the Early Devonian Hunsrück Slate (Germany) where several taxa (including echinoderms, molluscs, arthropods, vertebrates, and cnidarians) are beautifully preserved with some relief owing to early pyritisation. In addition, the soft parts of several taxa from the Early Cambrian Chengjiang biota of China were preserved by pyrite, which later oxidised to create the characteristic pink and orange fossils (Figure 4).

Clay Minerals

Clay minerals (*see* **Clay Minerals**) are capable of replacing soft tissues with fidelity slightly less than that of apatite but considerably better than that of pyrite. Clays have an affinity for organic matter when certain geochemical conditions are met. It has been demonstrated that a particularly low pH may be important in this respect, and may explain why, although clay minerals are ubiquitous in fine-grained sediments, clay mineral replacement of soft tissues is not. It is possible that detrital, colloidal clays which are extremely small (1 nm to 1 μm across) may become attracted towards, and template on to decaying soft tissues. Or clay authigenesis may be responsible for soft tissue preservation. In either case, it has been suggested that bacteria play an important role in somehow controlling and mediating the process. Several taxa from the Upper Ordovician Soom Shale of South Africa show soft tissue preservation by clay minerals, including the muscle fibres and fibrils of a conodont animal. Whilst it has been shown that the Middle Cambrian Burgess Shale fossils of Canada are composed of kerogen (or another graphite-like structure), clay minerals are also involved in their preservation and are seen to replace and coat organic surfaces.

Other Minerals

Silica and calcium carbonate are also involved in mineralising soft tissues. The Earth's oldest Lagerstätte, the Apex Cherts (*see* **Precambrian:** Prokaryote Fossils) (Warrawoona Group, Western Australia) contain silicified microfossils 3450–3470 Ma in age. The Early Devonian Rhynie Chert of Scotland preserves terrestrial arthropods and plants in superb detail. Here chert beds contain beautiful, three-dimensional plant material with preserved cell structures which became silicified when plants were engulfed in Si-rich hotspring waters (*see* **Sedimentary Rocks:** Chert).

The Role of Nodules

In a number of conservation Lagerstätten, early diagnetic concretions (nodules) formed around the fossil and protected it from later compaction. The Silurian Herefordshire fauna (UK) has produced spectacular three-dimensionally preserved fossils, owing to early stiffening of decaying animals by volcanic ash, followed by calcite infilling of the void left by the decayed tissues, and precipitation of calcite nodules around the fossil. A novel technique was developed to extract the maximum amount of information from these three-dimensional fossils. The nodules containing the fossils are serially ground at 30 μm intervals, and digitally photographed so that a set of 'slice images' of the whole fossil is created. These images are then used to create impressive 3D computerised reconstructions of the fossils (Figure 5). Other deposits where 3D preservation occurs owing to concretions include a diverse biota from the Carboniferous Mazon Creek (Illinois, USA) in siderite concretions, and vertebrates and invertebrates from the Lower Cretaceous Santana Formation (NE Brazil) in carbonate concretions (Figure 6).

Temporal Trends in Exceptional Preservation?

It has been suggested that the conditions required to create fossil Lagerstätten have not been uniformly present through the time of metazoans on Earth. In fact, taphonomic windows have appeared, leading to unique exceptionally preserved faunas, and then been 'closed'. The preservation potential of Cambrian, Ordovician, and Silurian organisms is not greatly different, and yet the Cambrian is relatively replete in fossil Lagerstätten compared with the other two periods. This may reflect a temporal difference in the processes of preservation. For example, it has been suggested that during the Cambrian, deep bioturbators had yet to invade deeper water settings,

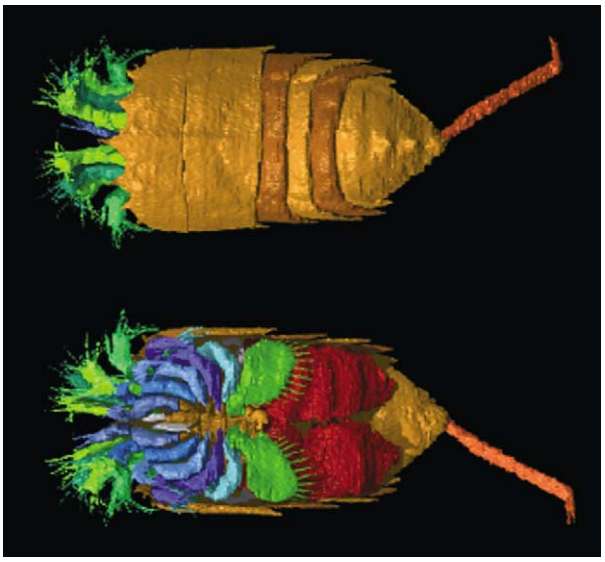


Figure 5 A computer based reconstruction of the three-dimensionally preserved arthropod *Offacolus kingi* from the Silurian Herefordshire (UK) Lagerstätte. Top is dorsal view, bottom is ventral view. Carapace *ca* 2 mm in width. (Image by Dr Mark Sutton.)

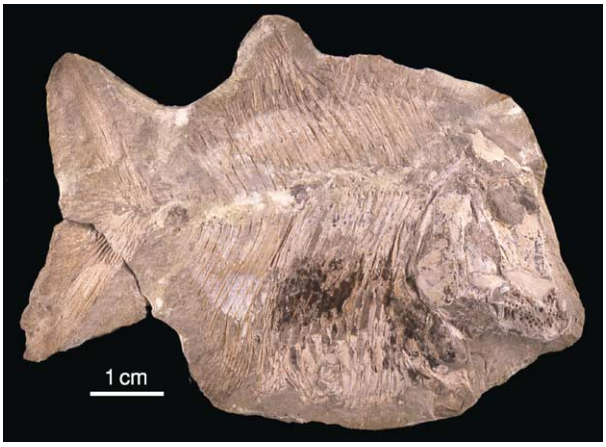


Figure 6 A pycnodont fish preserved in an early diagenetic carbonate nodule showing fully articulated skeleton. Specimen from the Early Cretaceous Romualdo Member, Santana Formation, NE Brazil. (Photograph by Dr Dave Martill.)

but after the Cambrian they radiated from onshore to offshore, decreasing the likelihood of organic preservation after the Cambrian. There are several other examples where exceptional preservation appears to be favoured during particular time periods.

See Also

Biosediments and Biofilms. **Clay Minerals**. **Fossil Invertebrates:** Arthropods. **Microfossils:** Conodonts. **Palaeoecology**. **Precambrian:** Prokaryote Fossils. **Sedimentary Rocks:** Chert; Phosphates. **Trace Fossils**.

Further Reading

Aldridge RJ, Gabbott SE, and Theron JN (2002) The Soom Shale. In: Briggs DEG and Crowther PR (eds.) *Palaeobiology II*, pp. 340–342. Oxford: Blackwell Science.

Allison PA and Briggs DEG (1991) Taphonomy of non-mineralized tissues. In: Allison PA and Briggs DEG (eds.) *Taphonomy: Releasing the Data Locked in the Fossil Record*, pp. 26–71. New York and London: Plenum Press.

Bartels C, Briggs DEG, and Brassel G (1998) *The Fossils of the Hunsrück Slate: Marine Life in the Devonian*. Cambridge: Cambridge University Press.

Barthel KW, Swinburne NHM, and Conway Morris S (1990) *Solnhofen: A Study in Mesozoic Palaeontology*. Cambridge: Cambridge University Press.

Bottjer DJ, Etter W, Hagadorn JW, and Tang C (2002) *Exceptional Fossil Preservation – A Unique View on the Evolution of Marine Life*. New York: Columbia University Press.

Briggs DEG (2002) Exceptionally preserved fossils. In: Briggs DEG and Crowther PR (eds.) *Palaeobiology II*, pp. 328–332. Oxford: Blackwell Science.

Briggs DEG (2003) Annual Review of Earth and Planetary Science. *Earth and Planetary Science* 31: 275–301.

Briggs DEG, Erwin DH, and Collier FJ (1994) *Fossils of the Burgess Shale*. Washington and London: Smithsonian Institution Press.

Butterfield NJ (1990) Organic preservation of non-mineralizing organisms and the taphonomy of the Burges Schale: *Palaeobiology*, v. 16, pp. 272–286.

Donovan SK (1991) *Processes of fossilization*. London: Belhaven Press.

Hou X-G, Aldridge RJ, Bergström J, Siveter DJ, and Feng X-H (2004) *The Cambrian Fossils of Chengjiang, China; The Flowering of Early Animal Life*. Oxford: Blackwell Science Ltd.

Kidwell SM (1991) The stratigraphy of shell concentrations. In: Allison PA and Briggs DEG (eds.) *Taphonomy: Releasing the Data Locked in the Fossil Record*, pp. 212–279. New York and London: Plenum Press.

Knoll AH and Shuhai X (2002) Precambrian Lagerstätten. In: Briggs DEG and Crowther PR (eds.) *Palaeobiology II*, pp. 332–337. Oxford: Blackwell Science.

Martill DM (1993) *Fossils of the Santana and Crato Formations, Brazil*. Dorcherter: Henry Ling Ltd.

Müller KJ (1990) Taphonomy of Fossil-Lagerstätten, Upper Cambrian 'Orsten.' In: Briggs DEG and Crowther PR (eds.) *Palaeobiology I*, pp. 274–277. Oxford: Blackwell Scientific Publications.

Selden PA and Nudds JR (2004) *Evolution of Fossil Ecosystems*. London: Manson Publishing Ltd.

Seilacher A (1990) Taphonomy of Fossil-Lagerstätten, Overview. In: Briggs DEG and Crowther PR (eds.) *Palaeobiology I*, pp. 266–270. Oxford: Blackwell Scientific Publications.

Spicer RA (1991) Plant taphonomic processes. In: Allison PA and Briggs DEG (eds.) *Taphonomy: Releasing the Data Locked in the Fossil Record*, pp 72–115. New York and London: Plenum Press.

Sutton MD, Briggs DG, Siveter D, and Siveter D (2001) Methodologies for the visualization and reconstruction of three-dimensional fossils from the Silurian Herefordshire Lagerstätte. *Palaeontologica Electronica* 4(1). http://laeoelectronica.org/2001_1/ s2/issue1_01.htm

Sutton MD, Briggs DEG, Siveter DJ, Siveter DJ, and Orr PJ (2002) The arthropod *Offacolus kingi* (Chelicerata) from the Silurian of Herefordshire, England: computer based morphological reconstructions and phylogenetic affinities. *Proceedings of the Royal Society, London B 269*: 1195–1203.

Trewin RH (2002) The Rhynie Chert. In: Briggs DEG and Crowther PR (eds.) *Palaeobiology II*, pp. 342–346. Oxford: Blackwell Science Ltd.

LARGE IGNEOUS PROVINCES

M F Coffin, University of Tokyo, Tokyo, Japan
O Eldholm, University of Bergen, Bergen, Norway

Introduction

Large igneous provinces (LIPs) are massive crustal emplacements of predominantly iron- and magnesium-rich (mafic) rock that form by processes other than normal seafloor spreading; they are the dominant form of near-surface magmatism on the terrestrial planets and moons of our solar system. On the Earth's surface, LIP rocks are readily distinguishable from the products of the two other major types of magmatism – mid-ocean ridge magmatism and arc magmatism – on the basis of petrologic, geochemical, geochronological, geophysical, and physical volcanological data. LIPs occur both on the continents and in the oceans, and include continental flood basalts, volcanic passive margins, oceanic plateaus, submarine ridges, seamount chains, and ocean-basin flood basalts (Figure 1 and Table 1). LIPs and their contemporary small-scale analogues, hotspot volcanoes, are commonly attributed to decompression melting of hot low-density mantle material ascending from the Earth's interior in mantle plumes, and thus provide a window onto mantle processes. This type of magmatism currently accounts for about 10% of the mass and energy flux from the Earth's deep interior to its crust. The flux may have been higher in the past, but is episodic over geological time, in contrast to the relatively steady-state activity at seafloor spreading centres. Such episodicity reveals dynamic non-steady-state circulation within the Earth's mantle, perhaps

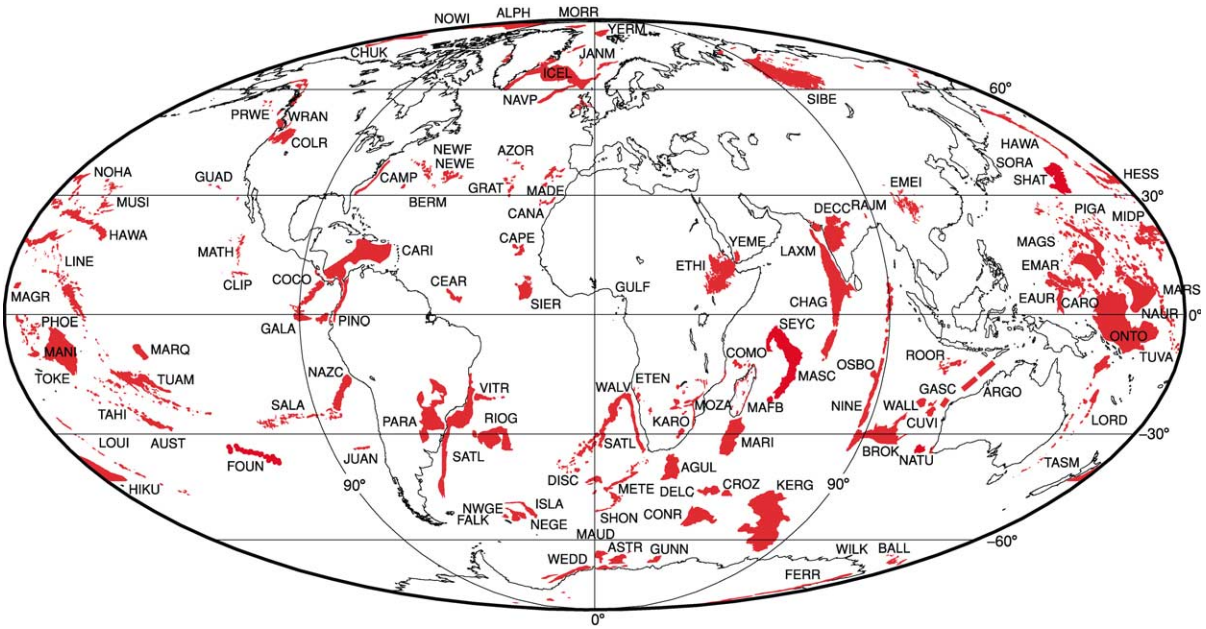


Figure 1 Phanerozoic global LIP distribution (red), with LIPs labelled (*see* Table 1).

Table 1 Large igneous provinces

Large igneous province	*Abbreviation (see Figure 1)*	*Type*
Agulhas Ridge	AGUL	SR
Alpha-Mendeleyev Ridge	ALPH	SR/OP
Argo Basin	ARGO	VM
Astrid Ridge	ASTR	VM
Austral Seamounts	AUST	SMT
Azores	AZOR	SMT
Balleny Islands	BALL	SMT
Bermuda Rise	BERM	OP
Broken Ridge	BROK	OP
Canary Islands	CANA	SMT
Cape Verde Rise	CAPE	OP
Caribbean Flood Basalt	CARI	OBFB (partly accreted)
Caroline Seamounts	CARO	SMT
Ceara Rise	CEAR	OP
Central Atlantic Magmatic Province (VM only)	CAMP	CFB/VM
Chagos–Laccadive Ridge	CHAG	SR
Chukchi Plateau	CHUK	VM
Clipperton Seamounts	CLIP	SMT
Cocos Ridge	COCO	SR
Columbia River Basalt	COLR	CFB
Comores Archipelago	COMO	SMT
Conrad Rise	CONR	OP
Crozet Plateau	CROZ	OP
Cuvier (Wallaby) Plateau	CUVI	VM
Deccan Traps	DECC	CFB/VM
Del Caño Rise	DELC	OP
Discovery Seamounts	DISC	SMT
East Mariana Basin	EMAR	OBFB
Eauripik Rise	EAUR	OP
Emeishan Basalts	EMEI	CFB
Etendeka	ETEN	CFB
Ethiopian Flood Basalt	ETHI	CFB
Falkland Plateau	FALK	VM
Ferrar Basalts	FERR	CFB
Foundation Seamounts	FOUN	SMT
Galapagos–Carnegie Ridge	GALA	SMT/SR
Gascoyne Margin	GASC	VM
Great Meteor–Atlantis Seamounts	GRAT	SMT
Guadelupe Seamount Chain	GUAD	SMT
Gulf of Guinea	GULF	VM
Gunnerus Ridge	GUNN	VM
Hawaiian–Emperor Seamounts	HAWA	SMT
Hess Rise	HESS	OP
Hikurangi Plateau	HIKU	OP
Iceland–Greenland–Scotland Ridge	ICEL	OP/SR
Islas Orcadas Rise	ISLA	SR
Jan Mayen Ridge	JANM	VM
Juan Fernandez Archipelago	JUAN	SMT
Karoo	KARO	CFB
Kerguelen Plateau	KERG	OP/VM
Laxmi Ridge	LAXM	VM
Line Islands	LINE	SMT
Lord Howe Rise Seamounts	LORD	SMT
Louisville Ridge	LOUI	SMT
Madagascar Flood Basalts	MAFB	CFB
Madagascar Ridge	MARI	SR/VM?
Madeira Rise	MADE	OP
Magellan Rise	MAGR	OP
Magellan Seamounts	MAGS	SMT
Manihiki Plateau	MANI	OP
Marquesas Islands	MARQ	SMT
Marshall Gilbert Seamounts	MARS	SMT
Mascarene Plateau	MASC	OP
Mathematicians Seamounts	MATH	SMT
Maud Rise	MAUD	OP
Meteor Rise	METE	SR
Mid-Pacific Mountains	MIDP	SMT
Morris Jesup Rise	MORR	VM
Mozambique Basin	MOZA	VM
Musicians Seamounts	MUSI	SMT
Naturaliste Plateau	NATU	VM
Nauru Basin	NAUR	OBFB
Nazca Ridge	NAZC	SR
New England Seamounts	NEWE	SMT
Newfoundland Ridge	NEWF	VM
Ninetyeast Ridge	NINE	SR
North Atlantic Volcanic Province	NAVP	CFB
North-east Georgia Rise	NEGE	OP
North-west Georgia Rise	NWGE	OP
North-west Hawaiian Ridge	NOHA	SR/SMT
Northwind Ridge	NOWI	SR
Ontong Java Plateau	ONTO	OP (partly accreted)
Osborn Knoll	OSBO	OP
Paraná	PARA	CFB
Phoenix Seamounts	PHOE	SMT
Pigafetta Basin	PIGA	OBFB
Piñón Formation (Ecuador)	PINO	OP (accreted)
Pratt–Welker Seamounts	PRWE	SMT
Rajmahal Traps	RAJM	CFB
Rio Grande Rise	RIOG	OP
Roo Rise	ROOR	OP
Sala y Gomez Ridge	SALA	SR
Seychelles Bank	SEYC	VM
Shatsky Rise	SHAT	OP
Shona Ridge	SHON	SR
Siberian Traps	SIBE	CFB
Sierra Leone Rise	SIER	OP
Sorachi Plateau (Japan)	SORA	OP (accreted)
South Atlantic Margins	SATL	VM
Tahiti	TAHI	SMT
Tasmantid Seamounts	TASM	SMT
Tokelau Seamounts	TOKE	SMT
Tuamotu Archipelago	TUAM	SMT
Tuvalu Seamounts	TUVA	SMT
Vitória–Trindade Ridge	VITR	SR/SMT
Wallaby Plateau (Zenith Seamount)	WALL	OP
Walvis Ridge	WALV	SR
Weddell Sea	WEDD	VM
Wilkes Land Margin	WILK	VM
Wrangellia	WRAN	OP (accreted)
Yemen Plateau Basalts	YEME	CFB
Yermak Plateau	YERM	VM

CFB, continental flood basalt; OBFB, ocean-basin flood basalt; OP, oceanic plateau; SMT, seamount; SR, submarine ridge; VM, volcanic margin.

extending far back into Earth history, and suggests a strong potential for LIP emplacements to contribute to, if not instigate, major environmental changes.

Composition, Physical Volcanology, Crustal Structure, and Mantle Roots

LIPs are defined by the characteristics of their dominantly iron- and magnesium-rich (mafic) extrusive rocks; these typically consist of subhorizontal subaerial basalt flows. Individual flows can extend for hundreds of kilometres, be tens to hundreds of metres thick, and have volumes as great as 10^4–10^5 km^3. Silica-rich rocks also occur as lavas and intrusive rocks and are usually associated with the initial and late stages of LIP magmatic activity. Relative to mid-ocean-ridge basalts, LIPs include higher MgO lavas, basalts with more diverse major-element compositions, rocks with more common fractionated components, both alkalic and tholeiitic differentiates, basalts with predominantly flat light-rare-earth-element patterns, and lavas erupted in both subaerial and submarine settings.

As the extrusive component of LIPs is the most accessible for study, nearly all of our knowledge of LIPs is derived from the lavas forming their uppermost crusts. The extrusive layer may exceed 10 km in thickness. On the basis of geophysical, predominantly seismic, data from LIPs and from comparisons with normal oceanic crust, LIP crust beneath the extrusive layer is believed to consist of an intrusive layer and a lower crustal body, characterized by P-wave velocities of 7.0–7.6 km s^{-1}, at the base of the crust (Figure 2). Beneath continental crust this body may be considered as a magmatically underplated layer. Seismic-wave velocities suggest an intrusive layer that is probably gabbroic and a lower crust that is ultramafic. If the LIP forms on pre-existing continental or oceanic crust or along a divergent plate boundary, dikes and sills are probably common in the middle and upper crust. The maximum crustal thickness, including extrusive and intrusive layers and the lower crustal body, of an oceanic LIP is about 35 km, as determined from seismic and gravity studies of the Ontong Java Plateau (Figure 1 and Table 1).

Low-velocity zones have recently been observed in the mantle beneath the oceanic Ontong Java Plateau and the continental Deccan Traps and Paraná flood basalts (Figure 1 and Table 1). Interpreted as lithospheric roots or keels, these zones can extend at least 500–600 km into the mantle. In contrast to the high-velocity roots beneath most continental areas and the absence of lithospheric keels in most oceanic areas, the low-velocity zones beneath LIPs apparently reflect primarily residual chemical, and perhaps some thermal, effects of mantle-plume activity. High-buoyancy roots extending well into the mantle beneath oceanic LIPs would suggest a significant role for LIPs in continental growth via accretion of oceanic LIPs to the edges of continents.

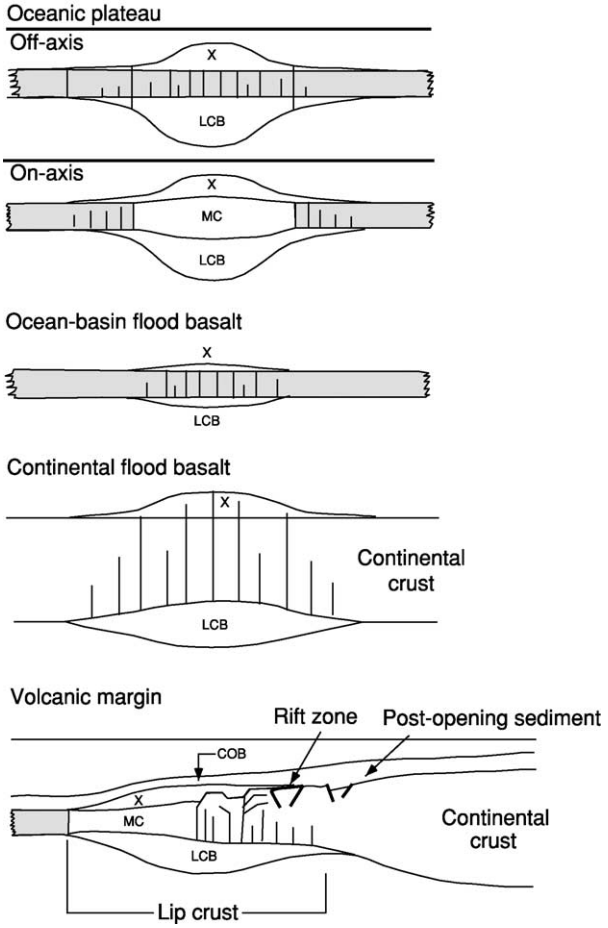


Figure 2 Schematic LIP plate-tectonic settings and gross crustal structure. The LIP crustal components are extrusive cover (X), middle crust (MC), and lower crustal body (LCB). Normal oceanic crust is shown in grey, and intrusives are denoted by vertical lines. The continent-ocean boundary (COB) is indicated for the volcanic margin.

Distribution, Tectonic Setting, and Types

LIPs occur worldwide, in both continental and oceanic crust, in purely intraplate settings and along present and former plate boundaries (Figure 1 and Table 1), although the tectonic setting at the time of formation is unknown for many features. If a LIP forms at a plate boundary, the entire crustal section is LIP crust (Figure 2). Conversely, if one forms in an intraplate setting, the pre-existing crust must be intruded and sandwiched by LIP magmas, albeit to an

extent that is not resolvable by current geological or geophysical techniques.

Continental flood basalts, which are the most intensively studied LIPs owing to their exposure, are erupted from fissures in continental crust (Figure 1 and Table 1). Most continental flood basalts overlie sedimentary basins that formed via extension, but it is not clear what happened first, the magmatism or the extension. Volcanic passive margins form as a result of excessive magmatism during continental breakup along the trailing rifted edges of continents. In the deep ocean basins, four types of LIPs are found. Oceanic plateaus, commonly isolated from the major continents, are broad typically flat-topped features generally lying 2000 m or more above the surrounding seafloor. They can form at triple junctions (e.g. the Shatsky Rise), at mid-ocean ridges (e.g. Iceland), or in intraplate settings (e.g. the northern Kerguelen Plateau). Submarine ridges are elongated steep-sided elevations of the seafloor. Some form along transform plate boundaries (e.g. the Ninetyeast Ridge). In the oceanic realm, oceanic plateaus and submarine ridges are the most enigmatic LIPs with respect to the tectonic setting in which they formed. Seamounts, which are closely related to submarine ridges, are local elevations of the seafloor; they may be discrete, form linear or random groups, or be connected along their bases and aligned along a ridge or rise (*see* **Seamounts**). They commonly form in intraplate regions (e.g. Hawaii). Ocean-basin flood basalts (e.g. the Nauru Basin and the Caribbean province) are the least-studied type of LIP and consist of extensive submarine flows and sills lying above and postdating the normal oceanic crust.

Ages

Age control for all LIPs apart from continental flood basalts is poor owing to their relative inaccessibility, but the $^{40}Ar/^{39}Ar$ dating technique is having a particularly strong impact on studies of LIP volcanism. Geochronological studies of continental flood basalts (e.g. the Siberian Traps, the Karoo, the Ferrar Basalts, the Deccan Traps, and the Columbia River Basalt; Figure 1) suggest that most LIPs result from mantle plumes, which initially transfer huge volumes (*ca.* 10^5–10^7 km^3) of mafic rock into localized regions of the crust over short intervals (*ca.* 10^5–10^6 years) but which subsequently transfer mass at a far lesser rate, albeit over significantly longer intervals (10^7–10^8 years). Transient magmatism during LIP formation is commonly attributed to mantle-plume 'heads' reaching the crust following transit through all or part of the Earth's mantle, whereas persistent magmatism is considered to result from steady-state mantle-plume 'tails' penetrating the lithosphere, which is moving relative to the plume (Figure 3). However, not all LIPs have obvious connections with mantle plumes or even hotspot tracks, suggesting that more than one source model may be required to explain all LIPs.

LIPs are not distributed uniformly over time. For example, many LIPs formed between 50 Ma and 150 Ma, whereas few have formed during the past 50 Myr (Figure 4). Such episodicity probably reflects variations in rates of mantle circulation, and this is supported by high rates of seafloor spreading during a portion of the 50–150 Ma interval, specifically during the long Cretaceous Normal Superchron

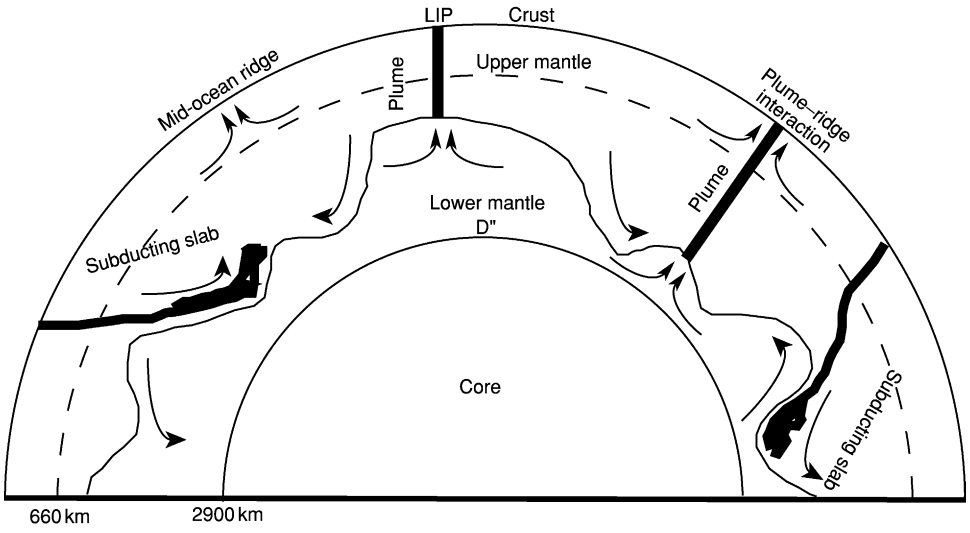


Figure 3 Diagram of the Earth's interior, showing plumes (tails), subducting slabs, and two mantle layers that move in complex patterns but never mix.

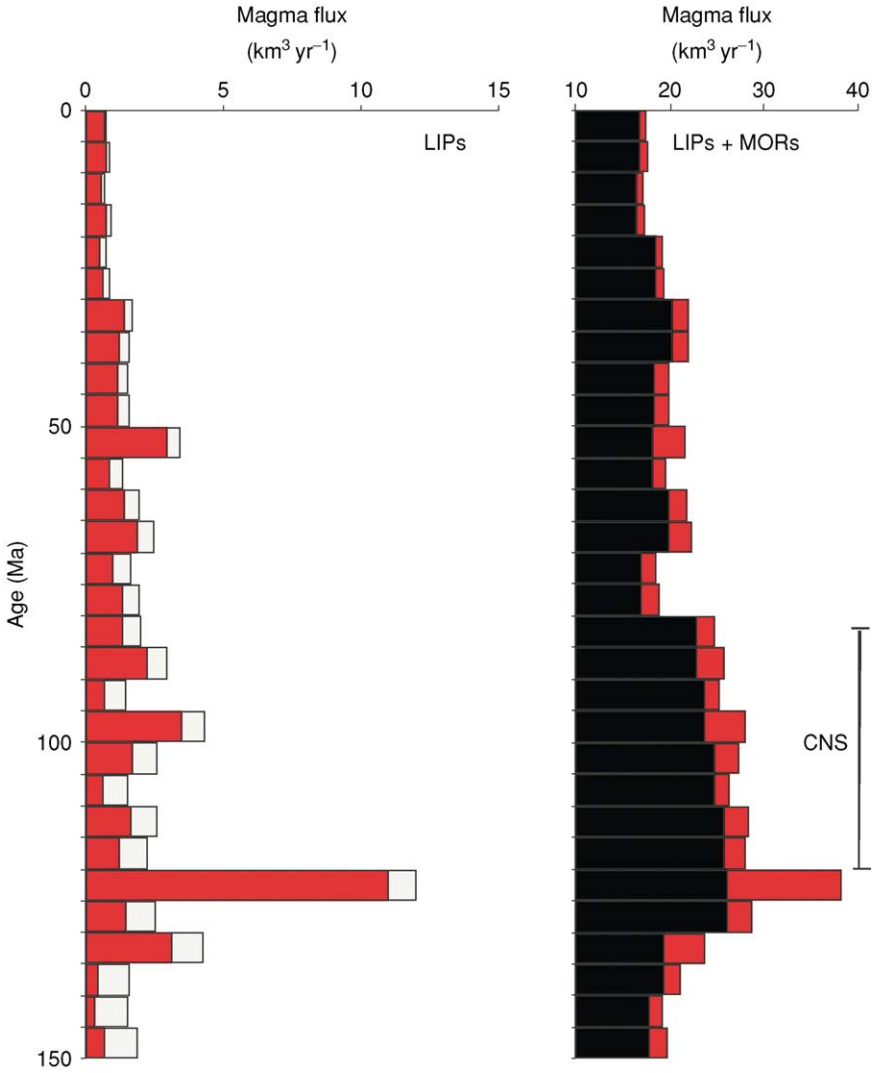


Figure 4 LIP, corrected for subduction (left; preserved in red and subduction correction in white), and summed LIP (red) and mid-ocean ridge (MOR; black) (right) magma production since 150 Ma. Note the difference in the *x*-axis scales. CNS, Cretaceous Normal Superchron.

(*ca*. 120–80 Ma), a time during which the Earth's magnetic field was of normal polarity. Thus, although LIPs manifest types of mantle processes distinct from those resulting in seafloor spreading, the waxing and waning rates of overall mantle circulation probably affect both sets of processes. A major question that emerges from studies of the global LIP production rate is whether the mantle is circulating less vigorously as the Earth ages.

LIPs and Mantle Dynamics

The formation of various sizes of LIP in a variety of tectonic settings on both continental and oceanic lithosphere suggests that a variety of thermal anomalies in the mantle give rise to LIPs and that the lithosphere strongly controls their formation. Equivalent thermal anomalies beneath continental and oceanic lithosphere should produce more magmatism in the latter setting, as oceanic lithosphere is thinner, allowing more decompression melting. Similarly, equivalent thermal anomalies beneath an intraplate region (e.g. Hawaii) and a divergent plate boundary (e.g. Iceland) (Figure 1 and Table 1) will produce more magmatism in the latter setting, again because decompression melting is enhanced. Recent seismic tomographic images of mantle-velocity (a proxy for temperature) structure beneath Iceland and Hawaii show significant differences between the two.

Only recently, seismic tomography has revealed that slabs of subducting lithosphere can penetrate the entire mantle to the D″ layer just above the

boundary between the mantle and the core, at a depth of approximately 2900 km (Figure 3). If we assume that the volume of the Earth's mantle has remained roughly constant throughout geological time, then the mass of crustal material fluxing into the mantle must be balanced by an equivalent mass of material fluxing from the mantle to the crust. Most, if not all, of the magmatism associated with the plate-tectonic processes of seafloor spreading and subduction is believed, on the basis of geochemistry and seismic tomography, to be derived from the upper mantle (above *ca*. 660 km depth). It is reasonable to assume that the lithospheric material that enters the lower mantle is eventually recycled, in some part contributing to the emplacement of LIPs at the Earth's surface.

Although LIPs are commonly believed to have originated from mantle plumes generated solely by solid Earth processes, alternative mechanisms have also been proposed. The spatial, if not temporal, association of flood basalts and impact craters on the Moon, as well as limited evidence on Earth, suggests that massive decompression melting of the mantle or at least significant crustal thinning and fracturing forming conduits for mantle material to reach the surface of a terrestrial planet could account for the emplacement of some LIPs. Such a mechanism has been proposed as an alternative to the plume hypothesis for the Siberian Traps, the Ontong Java Plateau, and the Deccan Traps. Other LIPs may originate as a result of a combination of plate divergence or fracturing and co-located underlying thermally anomalous mantle. Thus, multiple mechanisms may be required to explain all LIPs, both on Earth and elsewhere in our solar system.

LIPs and the Environment

The formation of LIPs has had documented environmental effects both locally and regionally. The global effects are less well understood, but the formation of some LIPs may have affected the global environment, particularly when conditions were at or near a threshold state. Investigations of volcanic passive margins and oceanic plateaus have demonstrated widespread and voluminous subaerial basaltic eruptions. The eruption of enormous volumes of basaltic magma during LIP formation releases volatiles such as carbon dioxide, sulphur, chlorine, and fluorine (Figure 5). A key factor affecting the magnitude of volatile release is whether the eruptions are subaerial or submarine; hydrostatic pressure inhibits vesiculation and degassing of relatively soluble volatile components (water, sulphur, chlorine, and fluorine)

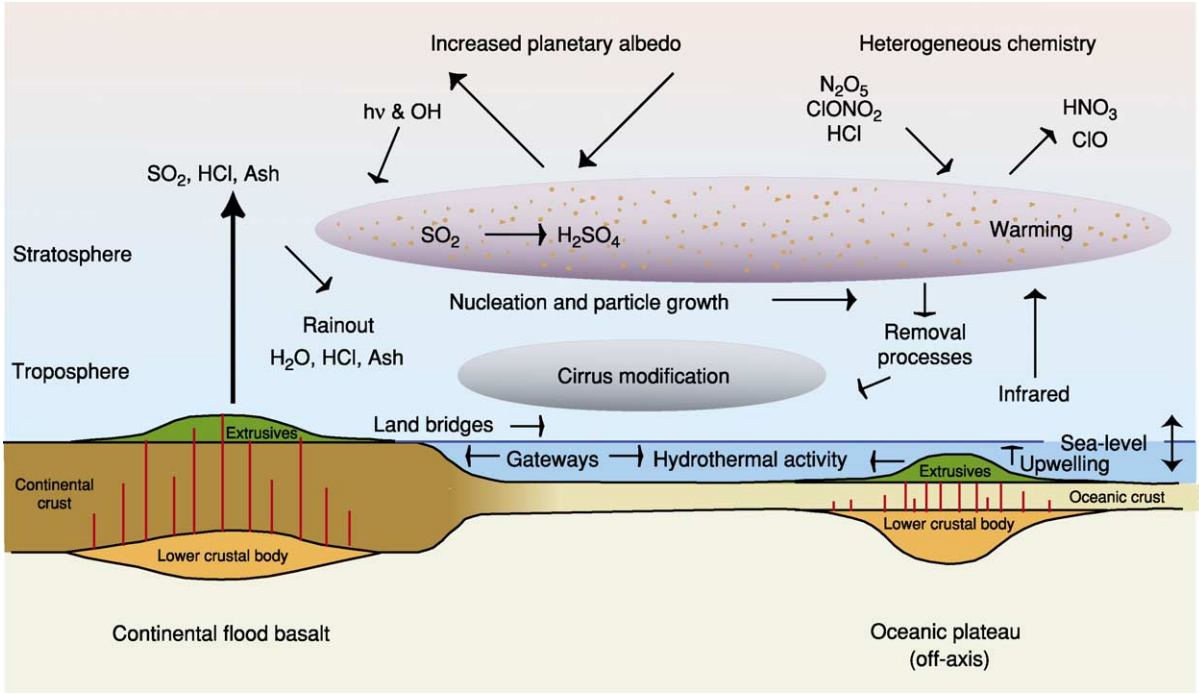


Figure 5 Environmental effects of LIP formation. LIP eruptions can perturb the Earth–ocean–atmosphere system significantly. Note that many oceanic plateaus form, at least in part, subaerially. Energy from solar radiation is hv, where h = Planck's constant, and v = frequency of electromagnetic wave of solar radiation.

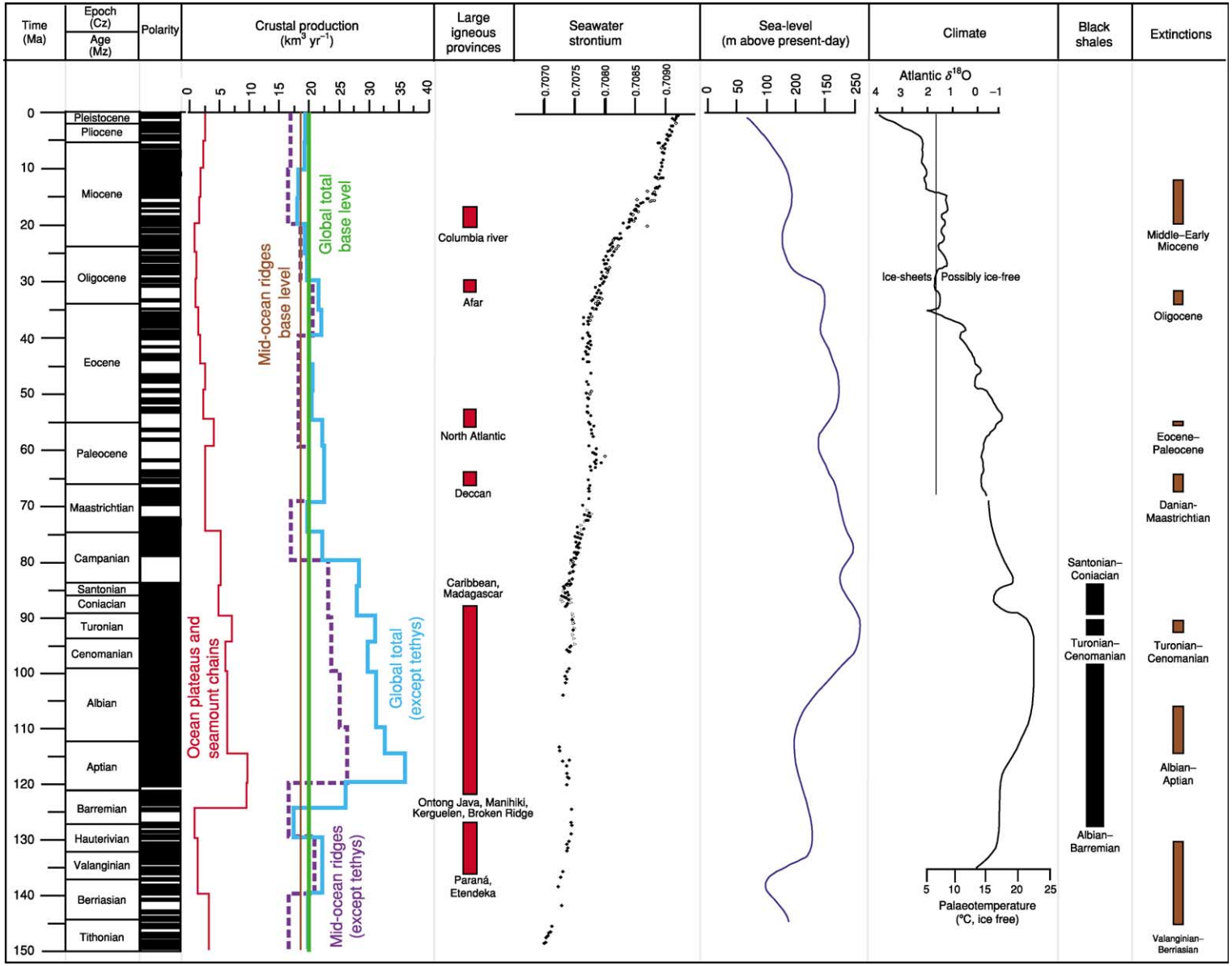


Figure 6 Temporal correlations among geomagnetic polarity, crustal production rates, LIPs, seawater strontium, sea-level, climate, black shales, and extinctions.

during deep-water submarine eruptions, although low-solubility components (carbon dioxide and noble gases) are mostly degassed even at abyssal depths.

Another important factor affecting the environmental impact of LIP volcanism is the latitude at which the LIP forms. In most basaltic eruptions, released volatiles remain in the troposphere. However, at high latitudes, the tropopause is relatively low, allowing large-mass-flux basaltic fissure eruption plumes to transport SO_2 and other volatiles into the stratosphere. Sulphuric acid aerosol particles that form in the stratosphere after such eruptions have a longer residence time and greater global dispersal than if the SO_2 remains in the troposphere; therefore, the effects on climate and atmospheric chemistry are greater. The large volume of volatiles released, over relatively brief geological intervals, by the subaerial flood basalts of high-latitude LIPs would contribute to potential global environmental effects.

Highly explosive felsic eruptions, such as those documented from volcanic passive margins, an oceanic plateau (Kerguelen; Figure 1 and Table 1), and continental flood basalt provinces, can also inject both particulate material and volatiles (SO_2 and CO_2) directly into the stratosphere. The total volume of felsic volcanic rocks in LIPs is poorly constrained, but they may account for a small, but not negligible, fraction of the volcanic deposits in LIPs. Significant volumes of explosive felsic volcanism would further contribute to the effects of predominantly mafic LIP volcanism on the global environment.

Between about 145 Ma and 50 Ma, the global oceans were characterized by variations in chemistry, relatively high temperatures, high relative sea-level, episodic deposition of black shales, high production of hydrocarbons, mass extinctions of marine organisms, and radiations of marine flora and fauna (Figure 6). Temporal correlations between the intense pulses of igneous activity associated with LIP formation and environmental changes suggest a causal relationship. Perhaps the most dramatic example is the eruption of the Siberian Traps (Figure 1 and Table 1) at approximately 250 Ma, coinciding with the largest extinction of plants and animals in the geological record. It is estimated that 90% of all species became extinct at that time (*see* **Palaeozoic:** End Permian Extinctions). On Iceland, the 1783–1784 eruption of Laki provides the only human experience of the type of volcanism that constructs LIPs. Although Laki produced a basaltic lava flow representing approximately 1% of the volume of a typical (10^3 km^3) LIP flow, the eruption's environmental impact resulted in the deaths of 75% of Iceland's livestock and 25% of its population from starvation.

Conclusions

Oceanic plateaus, volcanic passive margins, submarine ridges, seamount chains, ocean-basin flood basalts, and continental flood basalts share geological and geophysical characteristics that indicate an origin distinct from that of igneous rocks formed at mid-ocean ridges and arcs. These characteristics include

- a broad areal extent (in excess of 10^4 km^2) of iron- and magnesium-rich lavas;
- massive transient basaltic volcanism occurring over 10^5–10^6 years;
- persistent basaltic volcanism from the same source lasting 10^7–10^8 years;
- lower crustal bodies characterized by P-wave velocities of 7.0–7.6 km s^{-1};
- a component of more silica-rich volcanic rocks;
- higher MgO lavas, basalts with more diverse major-element compositions, rocks with more common fractionated components, both alkalic and tholeiitic differentiates, and basalts with predominantly flat light-rare-earth-element patterns, relative to mid-ocean-ridge basalts;
- thick (tens to hundreds of metres) individual basalt flows;
- long (up to 750 km) single basalt flows; and
- lavas erupted in both subaerial and submarine settings.

There is strong evidence that many LIPs both manifest a fundamental mode of mantle circulation, commonly distinct from that which characterizes plate tectonics, and contribute episodically, at times catastrophically, to global environmental change. Nevertheless, it is important to bear in mind that we have literally only scratched the surface of oceanic and continental LIPs, and that LIPs on other terrestrial planets await investigation.

See Also

Earth: Mantle. **Igneous Processes**. **Lava**. **Mantle Plumes and Hot Spots**. **Palaeozoic:** End Permian Extinctions. **Plate Tectonics**. **Seamounts**. **Tectonics:** Mid-Ocean Ridges.

Further Reading

Campbell IH and Griffiths RW (1990) Implications of mantle plume structure for the evolution of flood basalts. *Earth and Planetary Science Letters* 99: 79–93.

Coffin MF and Eldholm O (1993) Large igneous provinces. *Scientific American* 269: 42–49.

Coffin MF and Eldholm O (1994) Large igneous provinces: crustal structure, dimensions, and external consequences. *Reviews of Geophysics* 32: 1–36.

Condie KC (2001) *Mantle Plumes and Their Record in Earth History*. Cambridge: Cambridge University Press.
Davies GF (2000) *Dynamic Earth: Plates, Plumes and Mantle Convection*. New York: Cambridge University Press.
Duncan RA and Richards MA (1991) Hotspots, mantle plumes, flood basalts, and true polar wander. *Reviews of Geophysics* 29: 31–50.
Eldholm O and Coffin MF (2001) Large igneous provinces and plate tectonics. In: Richards MA, Gordon RG, and van der Hilst RD (eds.) *The History and Dynamics of Global Plate Motions*. Geophysical Monograph 121, pp. 309–326. Washington: American Geophysical Union.
Ernst RE and Buchan KL (2001) *Mantle Plumes: Their Identification through Time*. Special Paper 352. Boulder: Geological Society of America.
Ernst RE and Buchan KL (2003) Recognizing mantle plumes in the geological record. *Annual Review of Earth and Planetary Sciences* 31: 469–523.
Macdougall JD (ed.) (1989) *Continental Flood Basalts*. Dordrecht: Kluwer Academic Publishers.
Mahoney JJ and Coffin MF (eds.) (1997) *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism*. Geophysical Monograph 100. Washington: American Geophysical Union.
Morgan WJ (1981) Hotspot tracks and the opening of the Atlantic and Indian oceans. In: Emiliani C (ed.) *The Oceanic Lithosphere, The Sea, Volume 7*, pp. 443–487. New York: John Wiley & Sons.
Richards MA, Duncan RA, and Courtillot VE (1989) Flood basalts and hot-spot tracks: plume heads and tails. *Science* 246: 103–107.
Sleep NH (1992) Hotspot volcanism and mantle plumes. *Annual Review of Earth and Planetary Sciences* 20: 19–43.
White RS and McKenzie D (1995) Mantle plumes and flood basalts. *Journal of Geophysical Research* 100: 17 543–17 585.

LAVA

N Geshi, Geological Survey of Japan, Ibaraki, Japan

Introduction

Red-hot lava is characteristic of an active volcano, and the spectacle of flowing lava gives us an impression that the Earth is really living. Lava flows are the most common volcanic feature on Earth, and large volcanic structures, such as shield volcanoes and stratovolcanoes, are formed mainly from a pile of lava flows. Furthermore, the upper part of the ocean floor consists of submarine lava flows produced at the mid-ocean ridges, which form the largest volcanic zone on Earth. The structural characters of a lava flow, such as its size, composition, and viscosity, tell us much about its nature and emplacement mechanism. This article describes the primary features of lava flows and cites some examples of disasters caused by them.

Eruption of Lava

Lava is the term for molten magma erupted onto the Earth's surface as a continuous melt and also for the rock that solidifies from it. A flow of hot molten rock over the ground surface is called a lava flow, as is the solidified rock that forms from it.

Lava filling a broad depression or a vent is often called a 'lava lake'. A jet of fluid lava sprayed 10–100 m into the air is called a 'lava fountain'. Lava fountains are typical of extensive basaltic eruptions. Lava fountains may occasionally reach more than 500 m into the air (Figure 1).

Lava is generally a product of non-explosive moderate eruption. By contrast, fragments of molten lava produced in an explosive eruption, such as volcanic ash, pumice, scoria, and volcanic bombs, are called 'pyroclastics' or 'tephra' (*see* **Pyroclastics**). Some basaltic and andesitic lava flows are generated during explosive eruptions with high effusion rates by refusion and reactivation of an accumulation of still-molten pyroclastics in the vicinity of the vent. This type of lava is often called 'clastogenic lava' or 'rootless lava'.

The explosivity of a magma is controlled by the behaviour of bubbles that are present in the magma while it is ascending. When magma containing bubbles ascends rapidly, decompression causes expansion and rupturing of the bubbles and an explosion will occur. In contrast, effective separation of the bubbles from the magma will result in non-explosive eruption of lava flows. Bubble separation is controlled by many factors, such as viscosity, decompression speed, and gas content, and its mechanism is one of the current topics in volcanology.

Temperature and Viscosity of Lava

The highest measured temperature of the lava tends to correspond to the liquidus of a rock of the same composition. The liquidus is the temperature at which magma begins to crystallize, and the liquidus of dry magma at a pressure of one atmosphere ranges

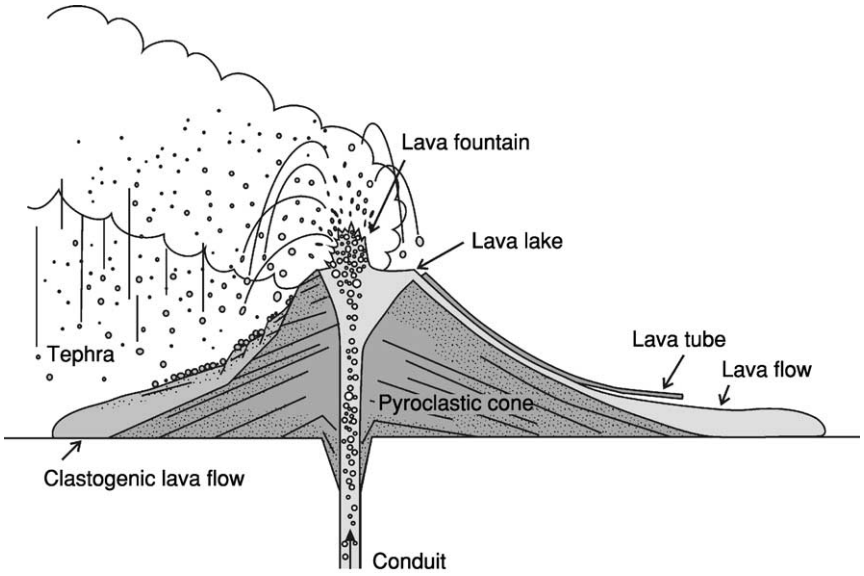


Figure 1 The eruption of quite fluid magma. Bubbles rise from conduit ruptures in the vent and produce pyroclastic particles. Degassed magma flows down the side of the cone and forms a lava flow. Lava travels through a lava tube crusted with solidified lava. Rapid sedimentation of still-molten pyroclastic materials causes welding and remobilization of the pyroclastics, forming a clastogenic lava flow.

from 1150°C to 1250°C for basalt and from 900°C to 1000°C for dacite and rhyolite. However, the temperature of flowing lava is usually less than the liquidus temperature because it is cooled quickly by radiation from its surface and conduction to the atmosphere. Flowing basaltic lavas from the Hawaiian volcanoes, Etna, and Izu-Oshima have temperatures of between 1000°C and 1100°C, and andesitic lavas from the 1946 eruption of Sakurajima had temperatures of 850–1000°C.

Viscosity is one of the important parameters controlling the movement and emplacement of lava flows. The viscosity of the melt increases as its silica content increases because silica particles form chain structures in silicate-rich melts. Elements such as magnesium can sever these chain structures, so the viscosities of magmas rich in these elements are low. Water also disrupts the silica chains, and the viscosities of magmas containing water are lowered dramatically. Viscosity is also strongly controlled by the temperature of the lava. Basaltic lava is usually of low viscosity (less than 10^4 Pa), and dacite and rhyolite lavas display very high viscosities (more than 10^{10} Pa).

Natural magma usually contains crystals that have crystallized from the cooling melt. The abundance of these crystals also controls the viscosity of the magma because of the way that the crystal grains interact. Collision of the crystals may check the flow of melt and increase the bulk viscosity of the magma, so the magma becomes more viscous with increasing crystal content. When the quantity of crystals exceeds 30%, this effect becomes noticeable, and magmas that are more than 60–70% crystals are unable to flow. This effect also depends strongly on the shape and size distribution of the crystals. Interaction among bubbles in the magma has a similar effect on bulk viscosity.

Volume, Effusion Rate, and Speed

The volume of a lava flow depends on the scale of the eruption, viscosity, and various other factors. The largest lava flow in the historic record was issued during the 1783 eruption in Iceland and is called the Laki flow. The lava was erupted from a 25 km long fissure over a period of 6 months and the total volume reached 12 km^3. The longest branch of the lava flow extended for more than 60 km and covered an area of more than 500 km^2. Flows of lava more than 10 km long generally have a basaltic composition (although the Yatta Plateau in Kenya, which is an intermediate flow of phonolite, extended for 300 km), and the maximum volume of lava flows decreases as their silica content increases. The effusion rate of lava is usually less than several cubic metres per second for basaltic lava flows and decreases as silica content increases because its high viscosity prevents flow through the conduit. This tendency may be controlled by many factors such as eruption style, the size of the magma batch, and the viscosity of the magma. However, this does not mean that a huge silicic eruption

Figure 2 (A) An advancing basaltic pahoehoe lava; Kilauea, Hawaii (photograph supplied by K Niihori). (B) The surface of an a'a' lava of basaltic andesite composition that was issued during the 1983 eruption of Miyakejima, Japan. Clinkers several centimetres across cover the surface of the flow. (C) The flow front of an andesitic block lava issued in the 1783 eruption of Asama, Japan. The surface of the lava flow is covered with angular blocks several metres in diameter. (D) Dacite lava dome of Unzen, Japan, formed during the 1991–1995 eruption.

cannot occur: silicic magmas sometimes cause Plinian eruptions – the most violent eruption style – and issue ignimbrites covering several square kilometres within a few hours or days.

Many scientists have tried to model the eruption mechanism that forms lava flows. Recent studies using numerical simulation have shown that eruption rate loosely controls the width of a single lobe of lava. In contrast, these studies have also shown that the length of a lava flow is affected by many complex parameters, such as slope angle, viscosity, and effusion time. In fact, a real lava flow is more complex. A large lava field usually consists of a complex of many small lava flows fed for a long period by tube systems, and the final shape is controlled by many other factors, such as the topography of the surface.

The maximum speed of a lava flow is usually less than several tens of metres per second. Basaltic lava, with its low viscosity, sometimes flows very fast. Hawaiian basaltic lava often flows at more than 15 m s^{-1} through lava tubes. During the 1950 eruption on Mauna Loa, Hawaii, lava flows travelled from the vent to the ocean, 10 km away, in as little as 3 h. In contrast, silicic lava flows very slowly. During the 1991–1995 eruption on Unzen, Japan, the maximum flow speed of the dacitic lava was less than 50 m day^{-1}.

Structure of Lava Flow

Subaerial Lava

A subaerial flow of lava is largely controlled by its viscosity. Lavas with a mafic composition – basalt and andesite – are normally fluid enough to flow away from the vent under gravity and form relatively thin flow lobes. In contrast, silicic magma cannot flow far because of its high viscosity and forms short thick flows (Figure 2).

Basaltic lava with a smooth or ropy surface is called 'pahoehoe lava'. Pahoehoe is a native Hawaiian term for lava with a smooth surface. During the advance of pahoehoe lava, small lobes and toes continuously break out from a cooled crust. When a red-hot molten lobe breaks out of the front of a previous lobe, a thin glassy crust is quickly formed on its surface by rapid cooling. Injection of molten lava into the lobe stretches the crust at its tip. As cooling progresses, the crust becomes rigid, and, rather than stretching further, it breaks to allow issue of the next new lobe. Sometimes flexible cooled crust is wrinkled by the motion of the molten lava below, forming a remarkable ropey structure on its surface. If the effusion rate

is high enough, the lava will form a large flat sheet-like lobe with a smooth surface. Pahoehoe lava is typical of basaltic lava with low viscosity and is rarely found in andesitic or rhyolitic lava. The thickness of a pahoehoe flow is usually less than 1 m, although lobes may often inflate to more than several metres thick as a result of the additional injection of molten lava.

Another typical type of basaltic lava is called 'a'a' lava'. The term a'a' also originates from the native Hawaiian term for lava with a rough surface. The surface of a'a' lava is completely covered by well-vesiculated rugged blocks called 'clinker'. Flow lobes of a'a' lava are, in many cases, several metres thick, much thicker than pahoehoe lava. During the advance of a'a' lava, flow cooling and vesiculation form a brittle crust at the top of the flow. Motion of the inner molten part breaks the crust and forms clinker. Since the velocity of the flow is at its maximum at the surface, clinkers on the top of the flow move forwards and tumble down the steep front of the flow. The advancing a'a' flow buries the fallen clinker, which forms a basal clinker layer. Some pahoehoe lavas change to a'a' lavas during their travel, but the reverse case is rare. During the transition from pahoehoe to a'a', the less-flexible cooled crust at the surface of the pahoehoe flow is fragmented by the motion of the inner molten part and forms a'a' clinker. This process indicates that the yield strength of the cooling crust controls the transition from pahoehoe to a'a' (Figure 3).

Viscous andesitic lava forms 'block lava'. The surface of block lava is covered by angular dense fragments with smooth faces. The angular blocks were formed by the break-up of the partly or wholly congealed upper part of the flow as the still-mobile magma moved beneath the crust. Block lava is usually thicker than a'a' lava: sometimes more than 100 m of fragmented material may constitute the entire thickness, making up a greater proportion of the flow than in a'a' lava.

Highly viscous lava, such as rhyolite and trachyte, cannot flow any distance and will form a thick short lobe of block lava around the extrusive point. Such a rise of lava is called a 'lava dome'. Usually the surface of a lava dome is covered with a block of lava produced by the internal motion of the viscous core and so resembles blocky lava.

Underwater Eruption of Lava

When magma is extruded at the bottom of a sea, lake, or glacier, rapid cooling by the surrounding water

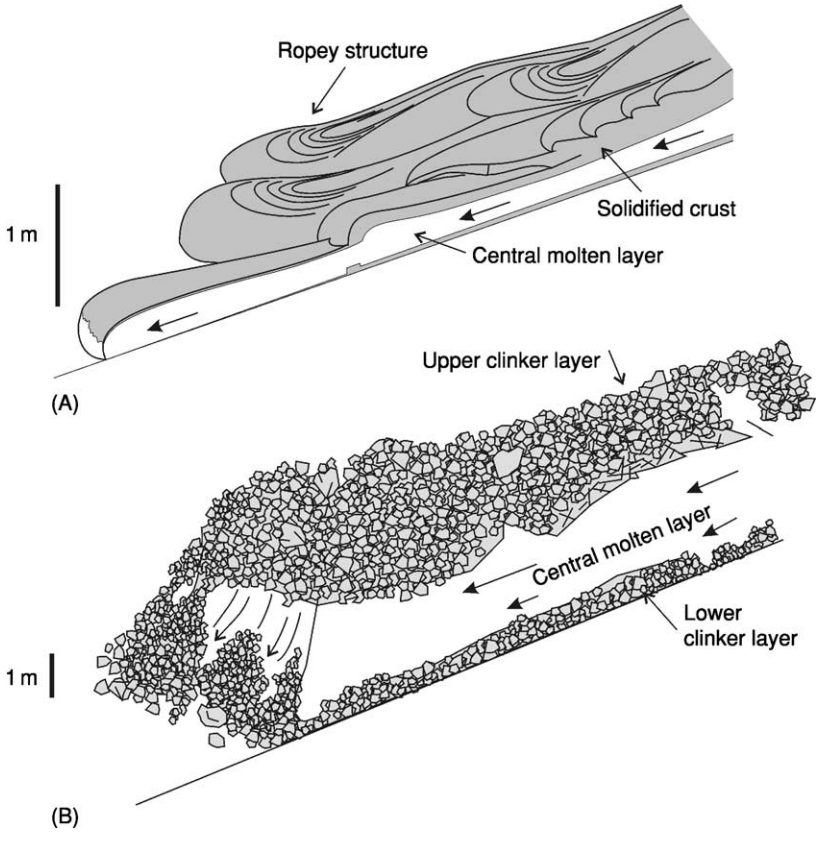


Figure 3 The structure of (A) pahoehoe lava and (B) a'a' lava. Note that the scales of the two diagrams are different.

forms a unique lava structure. Because of the large specific heat, thermal diffusivity, and evaporation heat of water, lava will be cooled more effectively in water than in air. When lava with this unique structure is exposed on land, a geologist can tell that the rock was formed underwater because of its structure.

One of the typical structures of lava so erupted is 'pillow lava', which forms mounds of elongated 'sacks' with quenched glassy rims. Repeated oozing and quenching of hot basaltic magma produces the pillow structure. A newly extruded lobe of lava is quickly surrounded by a flexible glassy skin due to rapid cooling by the surrounding water. Continuous injection of lava into the lobe expands it and forms a mass of basalt with a pillow shape. Finally, the skin breaks and new basalt extrudes into another lobe. Repetition of this sequence forms a thick deposit of pillow-like lobes of basalt (Figure 4). Pillow lava is a characteristic structure of fluid basaltic lava, corresponding to pahoehoe lava on land. When the effusion rate is high, basaltic lava spreads out on the seafloor to form 'sheet lava'. Rare pillows are known from more siliceous lavas.

More viscous lava such as andesite forms a volcanic breccia (hyaloclastite) by brittle fracturing due to quenching on contact with water. The surface of the flow lobe is chilled and forms a brittle crust while the viscous lava is still travelling. Motion of the molten interior fractures the crust and, because the surrounding water invades the inside of the flow, brittle fracturing advances deeply into the flow. As a result, a highly brecciated lobe consisting of angular glassy and massive blocks is formed.

Figure 4 A pile of pillow lavas produced during the Miocene by the Ogi, Japan, submarine volcano. Each pillow lobe has a glassy skin and a massive interior with radial cooling joints. (Photograph supplied by T Oikawa.)

Lava Tubes

Lava tubes are natural tunnels through which lava travels beneath the surface of a lava flow. Flowing lava is cooled from its surface by radiation and thermal convection of air, and a rigid crust is formed. Once the rigid crust is formed, it provides insulation because of its small thermal conductivity, and the inner molten lava can flow without cooling. In a broad lava-flow field, lava tube systems with a main tube from the vent and a series of smaller branches develop, which can supply lava to the front of the flow without it cooling. When the supply of lava ceases at the end of an eruption or the lava is diverted elsewhere, lava in the tube flows out and leaves a partially empty tunnel beneath the ground. Lava can also erode downwards, deepening the tube and leaving empty space above the flowing lava, because the walls and floor of the tube consist of lava of the same composition as the hot lava flowing in the tube, so the flowing lava can melt the wall rock. Lava tubes often develop in basaltic lava flows with low viscosity and are rare in highly viscous felsic lava.

Cooling Joints

One of the remarkable internal structures of lava is the systematic cooling joint. As the temperature of lava drops, its volume decreases and strain within the lava causes it to fracture (Figure 5). A typical cooling-joint system is 'columnar jointing', which forms prismatic columns of rock with polygonal cross sections. Columnar jointing is formed as follows: the spread out lava cools from its upper and lower surfaces, and the volume of lava decreases as its temperature falls. Since the surface area of the lava is fixed, tension from the contraction of the main part of the lava will form a polygonal – in many cases pentagonal or hexagonal – fracture system. Shrinkage fractures parallel to the surface rarely develop because tension in the vertical dimension is accommodated by a decrease in the thickness of the flow. A similar phenomenon, called 'sun cracks', is often observed in mud when a puddle dries up. In this case, the mud shrinks as water in the mud evaporates. As in the progress of cooling inside the lava, the fractures propagate inwards and polygonal pillars surrounded by platy fractures are formed. The axis of the pillar is normal to the isothermal plane. Columnar jointing often develops not only in lava but also in tabular intrusions such as dykes and sills.

Another type of cooling joint is the 'platy jointing'. This joint system consists of subparallel fractures forming thin plates. Tabular joints are typically observed at the bottom of a lava flow and at the wall of a dyke, where the shear strain acts during

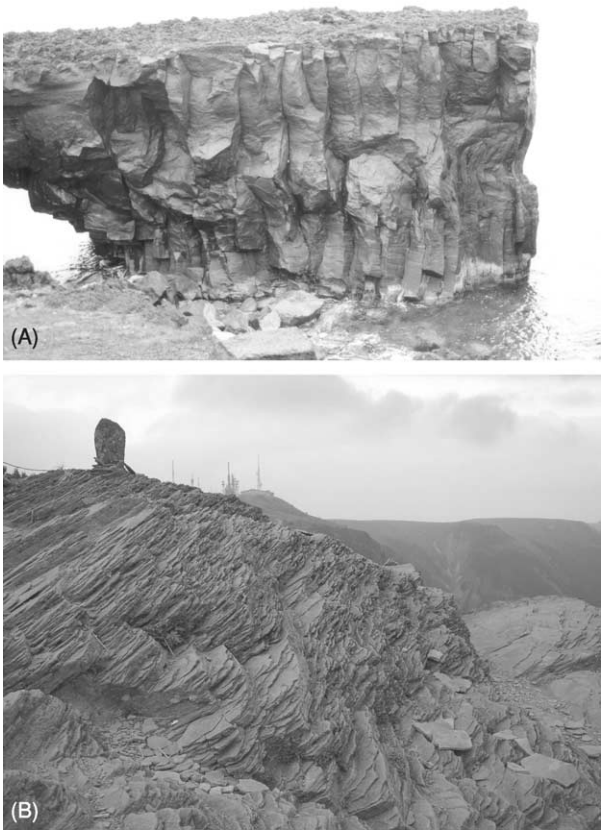


Figure 5 (A) Columnar jointing in thick basaltic lava from Miyakejima, Japan. (B) Tabular jointing in the middle Pleistocene andesitic lava of Utsukushigahara, Japan (Photographs supplied by T Oikawa).

flow. The brittle cooling margin at the base of the lava flow is dragged by the still-mobile magma inside the flow, and fractures parallel to the flow are formed. Shear deformation occurs at the bottom of the flow. Platy jointing is commonly observed in silicic lava flows such as those composed of andesite or dacite.

Flood Basalt

At certain times in the geological record, prolonged volcanic activity of immense scale resulted in thick piles of numerous basaltic lava flows that covered thousands of square kilometres (similar flows of the intermediate, more siliceous, lava type phonolite are also known in East Africa). The large-volume basaltic lava flows are called 'flood basalts' or 'plateau basalts'. The total volume of a flood-basalt eruption may sometimes exceed 100 000 km^3, and this huge volume is erupted over a short time interval, usually less than 1 Ma. Flood-basalt sequences consist of a pile of thousands of flows, and the individual flows may be more than several tens of metres thick. The individual flows may extend for hundreds of kilometres. How can such huge lava flows be emplaced? The old idea for the emplacement of flood basalts was that these flows were emitted at incredible velocities. However, based on the detailed survey of the structure of flood basalts, geologists arrived at an alternative idea that these flows are emplaced by slow movement, with most of the great thickness being achieved by injecting lava into the interior of an initially thin flow (*see* **Large Igneous Provinces**).

The most famous example of a flood basalt is the Deccan Traps of central India, where about 1 000 000 km^3 of basaltic lavas were erupted within a half million years at the end of the Cretaceous. The Siberian Traps, which erupted in Permian–Triassic times, are much larger than the Deccan Traps but less well understood. Many scientists consider that the volcanism that produces flood basalts is related to the rise of hot mantle plumes. The origin of the flood basalts is closely associated with entire-mantle dynamics, and their magma genesis is one of the currently popular topics in Earth science.

Lava Flow Hazards

Lava flows destroy or ignite all things in their path. When lava invades developed areas, it can be disastrous. In the cases of many volcanoes, the people living nearby make great efforts to stop the advance of lava flows, but much land and many buildings are destroyed and buried by lava flows. Fortunately, the speed of advance of a lava flow is, usually, very slow, and in many cases people can escape from it.

Kilauea (Hawaii, USA)

Kilauea is one of the most active volcanoes in the world. More than 90% of the land surface of the Kilauea volcano has been re-covered by new lava within the last 1500 years. In particular, the southeastern flank of the east rift zone of Kilauea has been threatened frequently by lava flows, and more than 30% of the land surface between the east rift zone and the coast has been covered by lavas since 1955, less than half a century ago (Figure 6).

The Puu Oo–Kupaianaha eruption of Kilauea from 1983 to the present (2004) is the longest duration of a single eruption in the historical record. Lava flows from the Puu Oo and Kupaianaha craters spread widely on the southern flank of the east rift zone and poured down to the ocean. Lava frequently changed direction, and one of the flows invaded Kalapana village in 1990. By the end of October that year, lava had destroyed all of Kalapana and

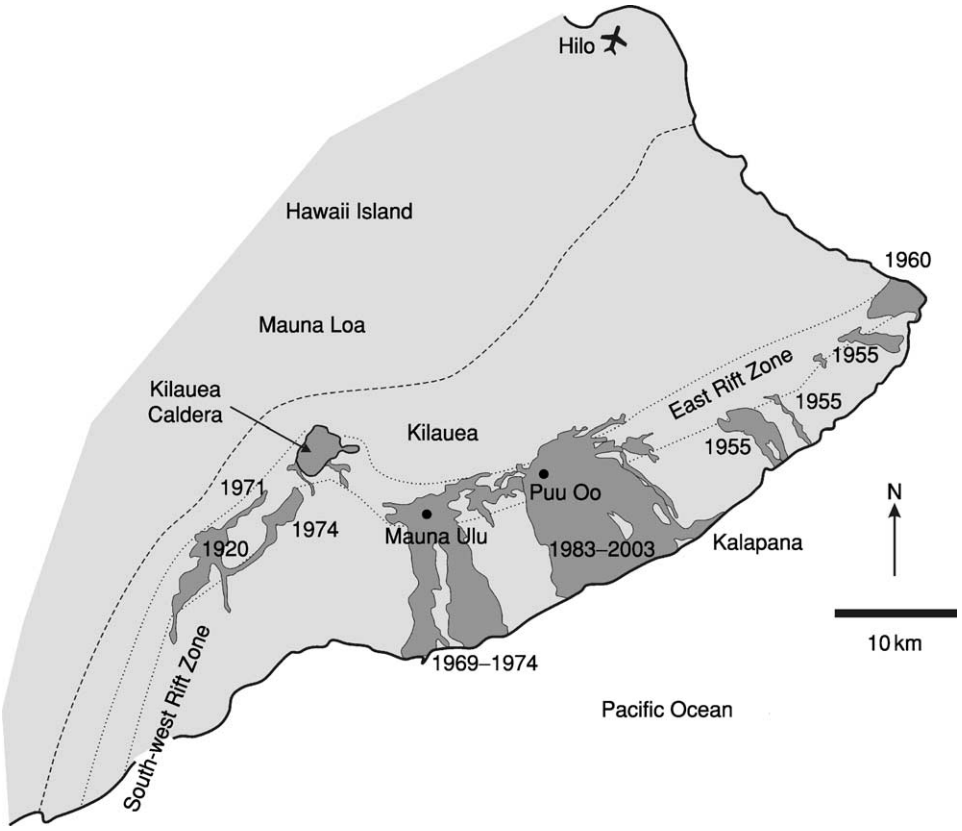


Figure 6 Distribution of the lava flows of Kilauea in the last 100 years. Presently, eruption is occurring at the Puu Oo vent in the east rift zone. From 1983 to 2003, about 2.3 km^3 of lava erupted and covered an area of more than 110 km^2.

filled up Kamimu bay. The lava spread slowly and people could evacuate safety, but more than 170 structures were destroyed. The lavas have added more than a square kilometre of new land to Hawaii Island, the largest of the Hawaiian island chain.

Nyiragongo (Congo)

Lava flowing from the Nyiragongo volcano in eastern Congo has also been disastrous. Nyiragongo is a stratovolcano that erupts predominantly nephelinitic lavas, with very low viscosity, and it is one of most active volcanoes in the East African rift zone. This volcano is famous for the active lava lake in the summit crater. A major fissure eruption occurred in 1977 on the south flank, which drained the lava lake in the summit caldera: very fluid lava rapidly covered several square kilometres and destroyed some villages. Because the lava was of very low viscosity and the eruption rate was very high, lava quickly ran down the steep slope of the volcano. The lava travelled about 10 km within 20 min and reached Goma, one of the major cities of Congo. The maximum speed of the lava was more than 50 $km\,h^{-1}$ in the vicinity of the fissure, which is extraordinarily fast for a lava flow. Over 70 people were killed by the lava flow. In January 2002, another major flank eruption occurred, and streams of lava spread over the centre of the city of Goma, dividing it in two. More than 40 people were killed by this eruption (*see* **Engineering Geology:** Natural and Anthropogenic Geohazards).

Etna (Sicily)

People living on the flanks of Etna are forever fighting against lava. Mount Etna is the largest stratovolcano in Europe and has erupted more than 150 times in the last 2700 years. The flank areas of Etna have frequently been invaded by lava flows during its long history. In 1966, a major fissure eruption occurred on the south-eastern flank. Lava erupted from a 12 km long fissure towards the city of Catania. Some people dug into the levee at the side of the lava flow to try to change the flow direction, but their effort ended in failure. In March 1983, a fissure on the southern flank began to produce a'a' lava, which destroyed several buildings. The lava flow advanced more than 6 km, and three towns were threatened by the lava by the end of April. To change the direction of the lava flow, people tried to breach the side levee of lava and

lead the flow into a diversion channel. The levee was reduced to only a few metres thick, and 400 kg of dynamite were set in the hot wall of the levee. Because of numerous technical difficulties, the breach was smaller than originally planned, and only 20% of the flow was diverted out of the main channel. Next, people tried to construct a rubble barrier about 10 m high, 30 m wide, and 400 m long along the western margin of the flow. Even though lava overtopped the first barrier, more barriers were formed and finally they succeeded in preventing lateral spreading of the flow field into developed areas.

Heimaey (Iceland)

The 1973 eruption on the island of Heimaey is a famous example of fighting a lava flow by cooling it with water. A fissure eruption began in January from a 2 km long fissure across the island in the vicinity of the centre of the town of Vestmannaeyjar, one of Iceland's major fishing ports. Lava flow from the vent began to threaten the town and the port. As the flow advanced to the north and east, the mouth of the port began to be buried by lava. A second large lava flow moved north-west on the west side of the main flow and had covered many houses by the end of March. By early May, some 300 buildings had been engulfed by lava flows or gutted by fire. To prevent the advance of the lava and save the town and port, people sprayed seawater onto the moving lava during the eruption. More than 30 km of water pipes and 43 pumps were used to deliver seawater at up to $1\,m^3\,s^{-1}$. A total of 6 000 000 m^3 of water was poured onto the lava. The front of the lava flow was solidified by the cooling effect of the water and it stopped moving. The eruption ended in July that year and the port was saved. After the eruption, people made a great effort to remove the lava and tephra from the centre of the city.

See Also

Engineering Geology: Natural and Anthropogenic Geohazards. **Igneous Processes**. **Large Igneous Provinces**. **Mantle Plumes and Hot Spots**. **Plate Tectonics**. **Pyroclastics**. **Tectonics:** Mid-Ocean Ridges. **Volcanoes**.

Further Reading

Bardintzeff JM and McBirney AR (2000) *Volcanology,* 2nd edn. Sudbury, MA: Jones & Bartlett Publishers.

Cas RAF and Wright JV (1987) *Volcanic Successions: Modern and Ancient*. London: Chapman & Hall.

Decker R and Decker B (1989) *Volcanoes: Revised and Updated Edition*. New York: WH Freeman and Company.

Decker RW, Wright TL, and Stauffer PH (eds.) (1987) *Volcanism in Hawaii*. Professional Paper 1350. US Geological Survey.

Fink JH (ed.) (1990) *Lava Flows and Domes: Emplacement Mechanisms and Hazard Implications*. IAVCEI Proceedings in Volcanology 2. New York: Springer Verlag.

Green J and Short NM (1971) *Volcanic Landforms and Surface Features – A Photographic Atlas and Glossary.* New York: Springer-Verlag.

Hall A (1996) *Igneous Petrology,* 2nd edn. London: Longman Group Limited.

Macdonald GA, Abbot AT, and Peterson FL (1990) *Volcanoes in the Sea: The Geology of Hawaii,* 2nd edn. Honolulu: University of Hawaii Press.

Schminke HU (2003) *Volcanism*. New York: Springer Verlag.

Sigurdsson H, Houghton B, McNutt ST, Rymer H, and Stix J (eds.) (2000) *Encyclopedia of Volcanology*. San Diego: Academic Press.

Wright T, Takahashi TJ, and Griggs JD (1992) *Hawaii Volcano Watch: A Pictorial History, 1779–1991*. Honolulu: University of Hawaii Press and Hawaii National History Association.

MAGNETOSTRATIGRAPHY

S G Lucas, New Mexico Museum of Natural History, Albuquerque, NM, USA

Introduction

Most of the dense core of the Earth is iron. The outer portion of the core is liquid, and the motion of this liquid produces a magnetic field, so that the Earth behaves like a giant bar magnet. This dynamo, however, changes easily, and for unknown reasons the magnetic field periodically reverses itself – the north and south magnetic poles switch positions.

On average, the magnetic field reverses itself about every 500 000 years, though the pattern of reversals is erratic. Flip-flops of the Earth's magnetic field, when recorded in a stratigraphical succession of rocks, are the basis of magnetostratigraphy (a contraction of 'magnetic-polarity stratigraphy').

Magnetostratigraphy correlates rocks on the basis of similarities in their magnetic-reversal patterns and is generally used to correlate surface exposures of rocks, though it can also be applied to subsurface cores. As explained below, magnetostratigraphy is not an independent method of correlating rocks. Nevertheless, it is a powerful tool because magnetostratigraphical correlation is based on matching magnetic reversals, which are geologically simultaneous events worldwide.

The Geomagnetic Polarity Time-Scale

Magnetic reversals have occurred frequently but irregularly during Earth history. The process of reversal seems to take about 4000–5000 years. The current state of the magnetic field (in which a compass needle points towards the north magnetic pole) has persisted for the last 700 000 years and is referred to as an interval of normal polarity. Geologists refer to periods when the poles had switched positions (so that a compass needle would have pointed to the south magnetic pole) as intervals of reversed polarity.

The first attempts at magnetostratigraphy were made in the 1950s, especially by the Russian scientist A N Khramov. Since the 1960s, geologists have made a concentrated effort to decipher the history of the Earth's magnetic field, and this research is ongoing.

During much of Earth history, the magnetic field reversed frequently (Figure 1). This has been the case throughout most of the Mesozoic and Cenozoic, but during the Late Carboniferous and most of the Permian, an interval of about 70 Ma, the magnetic field was stable (reversed). The pre-Carboniferous nature of the magnetic field is still not as well understood as its later history.

Because of plate tectonics and the subduction of oceanic crust, the oldest seafloor preserved on Earth dates from the beginning of the Late Jurassic, about 160 Ma ago. Geologists have determined the magnetic polarities of rocks from the seafloor, which are lavas for which some numerical ages have been calculated. Bands of cooled lava on the seafloor adjacent to spreading ridges preserve magnetic stripes that are symmetrical about the ridge. This seafloor magnetization provides a template that geologists have used

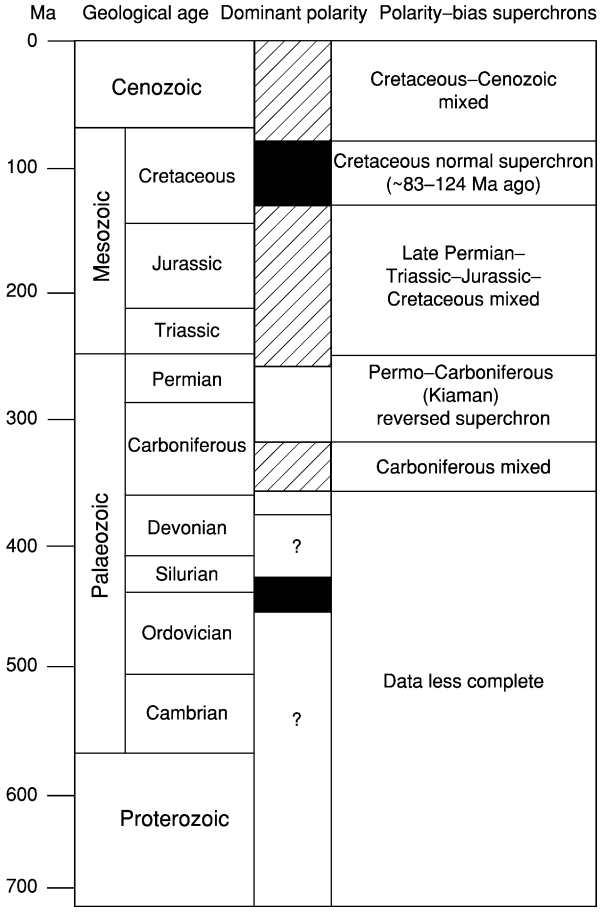


Figure 1 The polarity-bias superchrons during the last 700 Ma. The polarity history is not well understood before about 350 Ma, and since then it has been mixed (many magnetic reversals) except for two long intervals of polarity stability: the Permo-Carboniferous reversed and the Cretaceous normal superchrons.

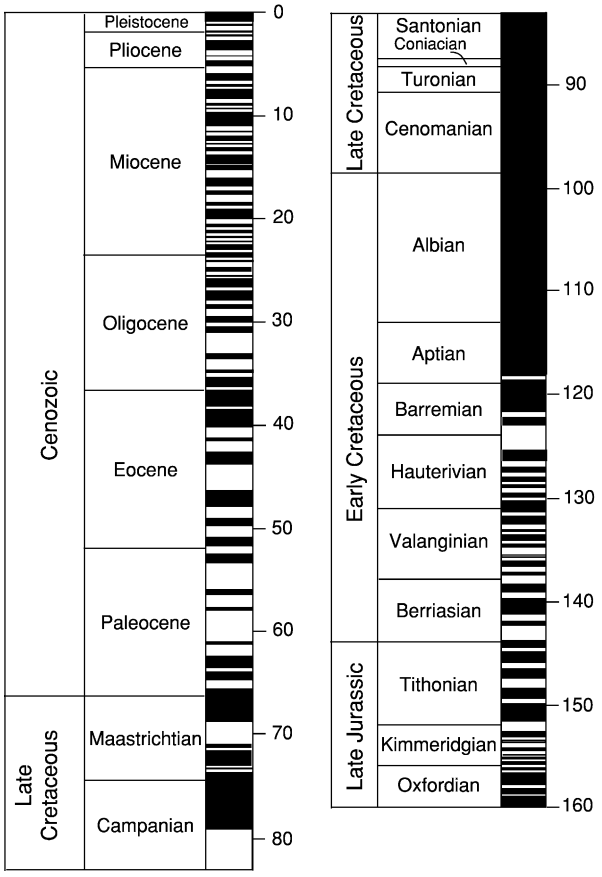


Figure 2 The global polarity time-scale for the last 160 Ma is a well-established and widely accepted record of magnetic-polarity reversals since the beginning of the Late Jurassic.

to plot the magnetic-polarity history of the Earth back to the beginning of the Late Jurassic (Figure 2). This polarity history is referred to as the geomagnetic-polarity time-scale (GPTS). It provides a globally consistent pattern of normal- and reversed-polarity intervals that can be used to estimate the ages of rocks and the events that they record during the last 160 Ma of Earth history.

In the history of the Earth's magnetic field, a superchron is an interval of tens of millions of years during which the polarity remains constant. There are two well-established superchrons: the Cretaceous normal superchron, from about 118 Ma to 83 Ma ago, and the Permo-Carboniferous (also called Kiaman, after a place in Australia) reversed superchron, from about 316 Ma to 262 Ma ago (Figure 1).

Early studies of magnetostratigraphy named the magnetic-polarity intervals (then called 'epochs') after scientists and mathematicians (Brunhes, Matuyama, Gauss, Gilbert) and the polarity events after the places where they were first identified (Jaramillo, Mammoth, Olduvai). However, it was subsequently realized that there are so many intervals (now called 'chrons') and events (now called 'subchrons') that numbering them is simpler than assigning them names. Chron numbers are followed by an 'n' or an 'r' to indicate whether they are normal or reversed.

Remnant Magnetization

Most rocks contain minerals that are naturally magnetic, such as the iron oxide minerals haematite (Fe_2O_3) and magnetite (Fe_3O_4). In the crystals in which they are bound, minute grains of magnetic minerals act like tiny bar magnets, and these mineral grains can record the direction of the Earth's magnetic field. Thus, when lava flows cool, these magnetic minerals align themselves with the Earth's magnetic field. Magnetic mineral grains also align themselves with the magnetic field when they are deposited in sediments. These processes provide a record of the state of the magnetic field (normal or reversed polarity) when a rock is formed. This record is called the remnant magnetization.

Heat destroys the magnetization of a rock. Indeed, magnetic minerals lose their magnetization at a certain temperature (usually above 500°C), called the Curie point (after the chemist Pierre Curie). The natural remnant magnetization is locked into a rock when it cools below its Curie point, which is approximately 650°C for haematite and 580°C for pure magnetite.

There are three kinds of remnant magnetization. Thermal remnant magnetization is the result of a molten rock cooling below its Curie point, at which magnetic minerals align with the current magnetic field and become locked into the crystal with that alignment.

Detrital remnant magnetization occurs when an igneous rock erodes, and its magnetic minerals become loose sedimentary particles. These tiny magnetic grains (which are only a few micrometres in diameter) act as bar magnets and align with the magnetic field as they settle through the water column and are deposited as sedimentary particles (detritus).

Chemical remnant magnetization takes place when iron weathers out of a rock, moves through groundwater, and precipitates elsewhere. Usually it precipitates as some form of haematite. During the precipitation, the magnetic minerals, which were initially aligned when the original rock was formed, realign themselves with the magnetic field at the time of precipitation. Thus, the new alignment is a younger magnetization than the original magnetization that was acquired when the rock formed.

Field Sampling and Laboratory Analysis

To determine the magnetic polarity a rock, sampling begins in the field, either by drilling small cores out of well-indurated rocks or by cutting small blocks (cubes) out of less resistant rocks. The orientations of these samples with respect to the current magnetic field and their precise geographical locations are recorded. Orientated samples are taken to the laboratory to have their magnetization analysed.

The intensity of the magnetic field is measured using a unit called a gauss (named after the German

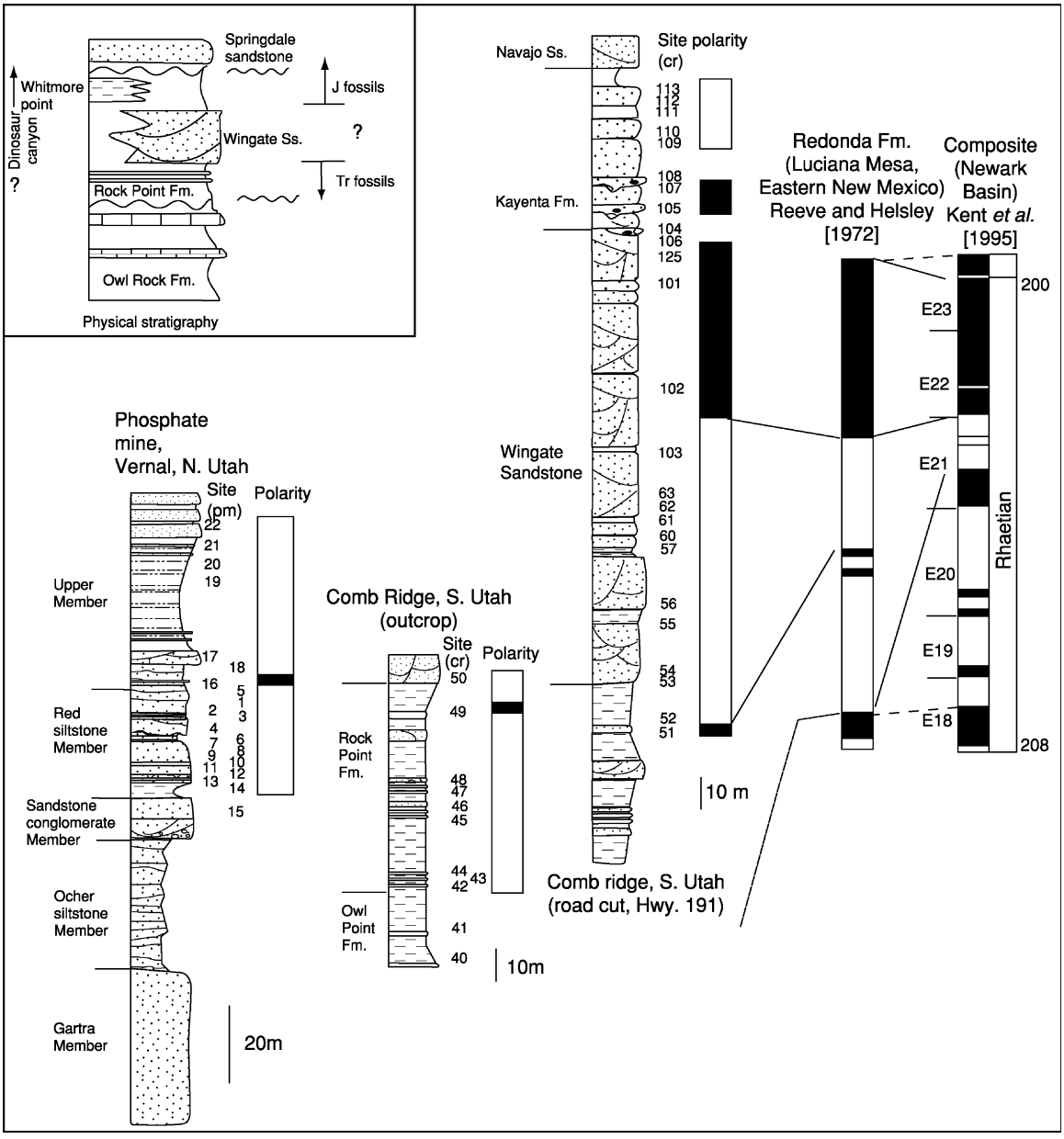


Figure 3 An example of how rock sections can be correlated with each other based on their magnetic-polarity zonation. The three columns on the left show the magnetostratigraphy of three stratigraphical sections in Utah, with black indicating normal-polarity magnetozones and white indicating reversed-polarity magnetozones. The short interval of normal polarity between two long intervals of reversed polarity allows the three Utah columns to be correlated with each other. On the right, an attempt is made to correlate the Utah magnetostratigraphies with magnetozones in New Mexico and New Jersey (Newark). The match between the Utah and New Mexico magnetostratigraphies is fairly straightforward, but matching these with New Jersey data is more complex, because the Newark column has many more magnetic reversals. Molin-Garza RS, Geissman JW and Lucas SG (2003) Paleomagnetism and magnetostratigraphy of the lower Glen Canyon and upper Chinle groups, Jurassic-Triassic of northern Arizona and northeast Utah. *Journal of Geophysical Research*, v. 108, no. B4, 2181.

mathematician Karl Gauss). In the laboratory, a magnetometer is used to measure the intensity and direction of the magnetic vector of the rock sample. At this point, all the magnetizations acquired by the rock since it was formed are present. The goal is to identify the original magnetization at the time of rock formation (i.e. the natural remnant magnetization). Subsequent (younger) magnetizations are referred to as overprints. These younger magnetizations need to be removed; in a sense we can think of the process as cleaning up the magnetization history of the rock so that only the natural remnant magnetization remains (or at least can be identified). This is done in one of two ways, depending on which kind of magnetometer is used. Alternating-field demagnetization subjects the rock samples to a strong alternating magnetic field that destroys the weaker magnetizations in the rock, so that only the stronger natural remnant magnetization remains. Thermal demagnetization heats the rock; during this process weaker magnetizations tend to disappear first, at lower temperatures than the stronger natural remnant magnetization. Neither method is usually strong enough to realign the primary magnetic mineral grains in the rock, and thus remagnetize it.

Magnetostratigraphical Correlation

Once the magnetic polarity of a succession of rocks has been determined, a magnetostratigraphy can be established (Figure 3). The basic unit of such a magnetostratigraphy is the magnetostratigraphical polarity zone (magnetozone for short), a body of rock with normal or reversed polarity. Now, the problem is to establish whether the observed succession of reversed and normal intervals has a pattern that can be correlated with that of another succession and/or with one or more segments of the global polarity time-scale. In other words, this piece (preserved by a local rock succession) needs to be matched with another piece or with a piece of the geomagnetic-polarity time-scale. This matching is termed 'correlation'. In correlation, a signature – a distinctive pattern of magnetic-polarity reversals – is looked for in order to establish a match (Figure 3).

The global pattern of magnetic reversals is irregular and nonperiodic, so that distinctive long intervals of magnetic reversals can be recognized. Reversals took place wordwide, independently of rock types and environments, and were geologically instantaneous (in rocks millions of years old, the 5000 or so years it takes for the reversal to happen is insignificant). However, polarity events are not unique, so that only a long succession of polarity intervals of distinctive lengths can be correlated. Furthermore, sedimentary hiatuses (unconformities) and changes in sedimentation rate can confuse the picture. Because of these problems, an independent method of correlation is often needed to provide a tie point (datum) against which to correlate the magnetostratigraphy. In other words, some other idea of the general age of the local rock succession – either derived from an index fossil or a numerical age – is usually needed to help narrow the possible correlation of magnetic-reversal histories. This means that magnetic-polarity-based correlations are typically not an independent means of correlating strata, although, once an index fossil or numerical age places the local slice of magnetic-polarity history 'in the ballpark', the matching of magnetic signatures often provides a more exact correlation than can be obtained from fossils or numerical ages alone.

Secular Variation

The Earth's magnetic north pole is close to, but not the same as, the geographical north pole. This means that, in most places, there is a small east-west difference between true north and magnetic north. The angle of this east-west deviation, measured from anywhere on Earth, is called the declination. For example, in California the declination is about 20° to the east, whereas in New York it is about 10° to the west of true north. The angle that the magnetic field makes with the Earth's surface is called the inclination. At the equator, the inclination is nearly horizontal, whereas at the magnetic pole it is vertical.

The magnetic field varies globally on geologically short time-scales of a few hundred years. These variations in declination, inclination, and field intensity are called secular variation. Secular variations are not magnetic reversals, but they are well documented over at least the last 10 000 years, and such palaeosecular variation can provide a succession of magnetic events that may be useful in correlation, particularly in archaeological research.

See Also

Analytical Methods: Geochronological Techniques. **Lava**. **Palaeomagnetism**. **Sedimentary Rocks:** Ironstones. **Stratigraphical Principles**. **Tectonics:** Mid-Ocean Ridges.

Further Reading

Butler RF (1992) *Paleomagnetism: Magnetic Domains to Geologic Terranes*. Boston: Blackwell.

Kennett JP (ed.) (1966) *Magnetic Stratigraphy of Sediments*. Stroudsburg, PA: Dowden, Hutchinson and Ross.

Khramov AN (1958) *Paleomagnetism and Stratigraphic Correlation*. Canberra: Australian National University [English translation, published 1960].

Khramov AN (1987) *Paleomagnetology*. Berlin: Springer-Verlag.
McElhinny MW (1973) *Paleomagnetization and Plate Tectonics*. Cambridge: Cambridge University Press.
McElhinny MW and McFadden PL (2000) *Paleomagnetization: Continents and Oceans*. San Diego: Academic Press.
Opdyke ND and Channell JET (1996) *Magnetic Stratigraphy*. San Diego: Academic Press.
Tarling DH (1983) *Paleomagnetization*. London: Geological Society.
Tauxe L (1998) *Paleomagnetic Principles and Practice*. Dordrecht: Kluwer Academic Publishers.

MANTLE PLUMES AND HOT SPOTS

D Suetsugu, B Steinberger, and T Kogiso, Japan Marine Science and Technology Center, Yokosuka, Japan

Introduction

Hotspots are defined as anomalous volcanism that cannot be attributed to plate tectonics, unlike that associated with island arcs and spreading ridges. Mantle plumes, which are upwelling instabilities from deep in Earth's mantle, are thought to be responsible for hotspots that are relatively stationary, resulting in chains of islands and seamounts on moving oceanic plates. The volcanic rocks associated with hotspots have signatures in trace elements and isotopes distinct from those observed at mid-oceanic ridges and island arcs. Seismic imaging has revealed low-velocity anomalies associated with some deep-rooted hot mantle plumes, but images of their full-depth extent are of limited resolution, thus evidence for plumes and hotspots is primarily circumstantial. Commonly, it is not even clear which areas of intraplate volcanism are underlain by a mantle plume and should be counted as a hotspot.

Surface Expression of Hotspots

The primary surface expression of mantle plumes consists of hotspot tracks. These are particularly evident in the oceans as narrow ($\approx$100 km) chains of islands and seamounts, such as the Hawaiian–Emperor chain, or as continuous aseismic ridges, such as the Walvis Ridge, up to several kilometres high. These tracks are thought to form as lithospheric plates move over plumes. The active hotspot is at one end of the chain; radiometric dating has determined that the ages of the volcanics along the chain tend to increase with distance from the active hotspot. Interpretation of age data is complicated, because volcanics do not necessarily erupt directly above a plume. Late-stage volcanism may occur several million years (My) after passage over a plume. Many hotspot tracks begin with a flood basalt or large igneous province. Volcanic volumes and age data indicate that these form during short time-spans with much higher eruption rates than are found at present-day hotspots. Examples of continental flood basalts (CFBs) include the Deccan Traps (associated with the Reunion hotspot) and the Parana basalts (associated with the Tristan hotspot). The Deccan Traps have erupted a volume of $\approx 1.5 \times 10^6\,\mathrm{km}^3$ within less than 1 My, whereas the present-day eruption volume at the Reunion hotspot is $\approx 0.02\,\mathrm{km}^3\,\mathrm{year}^{-1}$. For other tracks, older parts have been subducted, and yet others, particularly shorter ones, begin with no apparent flood basalt. The length of tracks shows that hotspots may remain active for more than 100 My. For example, the Tristan hotspot track indicates continuous eruption for 120 My. Numerous shorter tracks exist as well, particularly in the south central Pacific, commonly without clear age progression. This may indicate either that the region is underlain by a broad upwelling or that widespread flow from a plume is occurring beneath the lithosphere, with locations of volcanism controlled by lithospheric stresses. Geometry and radiometric age data of hotspot tracks indicate that the relative motion of hotspots is typically slow compared to plate motions. However, for the Hawaiian hotspot between 80 and 47 million years ago (Ma), inclination of the magnetization of volcanics indicates formation at a palaeolatitude further north than Hawaii, with hotspot motion southward of several centimetres per year. The Hawaiian–Emperor bend may therefore represent more than a change in Pacific plate motion. In most other cases in which palaeolatitude data are available, inferred hotspot motion is slow or below detection limit. Associated with many tracks is a hotspot swell ($\approx$1000 km wide, with up to 3 km anomalous elevation). Swells are associated with a geoid anomaly. Swell height slowly decreases along the track away from the active hotspot, and the swell also extends a few 100 km 'upstream' from the hotspot. The geoid-to-topography ratio remains approximately constant along swells, and this value indicates isostatic compensation at depths $\approx$100 km. From the

product of swell cross-sectional area and plate speed relative to the hotspot, plume anomalous mass flux (volume flux times density anomaly of plume relative to ambient mantle - this is the quantity that is directly estimated from observations), volume flux (volume flowing through plume conduit per time unit), and heat flux, (volume flux density times heat capacity times temperature anomaly of plume relative to ambient mantle), can be estimated. For Hawaii, the hotspot track and swell are both evident (Figure 1). The anomalous mass flux determined for Hawaii is the largest of any plume. Its estimated heat flux corresponds to ≈16% of global hotspot heat transport; global hotspot heat transport in turn is estimated to be ≈5% of total global heat flux.

Direct measurements of heat flow above hotspot swells yield only small anomalies (≈5–10 m W m^{-2}) or no anomalies at all. This may indicate that the swell is caused by buoyant uplift due to hot plume material spreading beneath the lithosphere, rather than by thermal erosion and heating of the lithosphere. The amount of volcanics produced at hotspots is estimated to be $<1\,km^3\,year^{-1}$, ≲5% of volcanics produced at mid-ocean ridges. Magma production at mid-plate hotspots thus appears less efficient than at mid-ocean ridges. However, during formation of large igneous provinces, magma production rate and heat flow were higher than at present.

Hotspots cluster in two antipodal regions around the Pacific and Africa. They are mostly absent from regions where subduction has occurred in the past 100 My, and tend be in highs of the 'residual' geoid (actual geoid minus contribution of subducted slabs). Probably the hotspot plumes are not the primary cause of these geoid highs. Rather, both may be due to the same cause: a less dense and presumably hotter lower mantle in these regions. Plate boundaries, in particular mid ocean ridges, may move across hotspots; hotspots may hence, successively or simultaneously, leave tracks on different plates. Examples are the Reunion hotspot, with tracks on the Indian and African plates, and the Tristan hotspot, with tracks on the South American and African plates. If a plume is located close to a spreading ridge, eruption of plume material may occur not only directly above the plume, but also at the section of the ridge

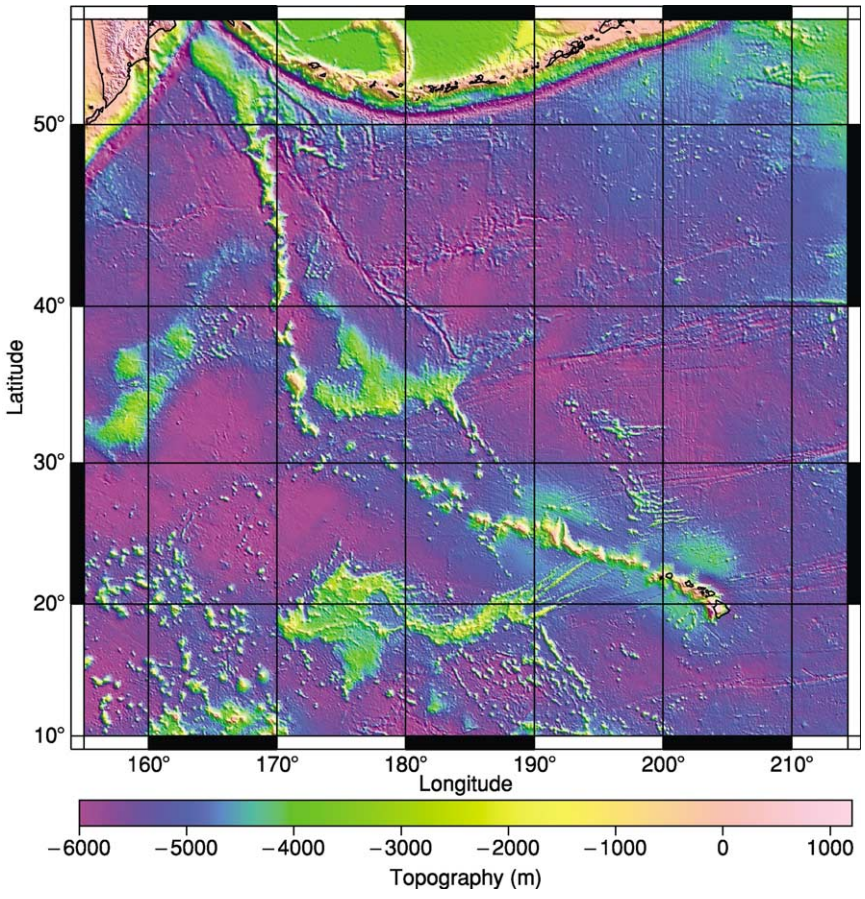


Figure 1 Topographic map of the North Pacific, showing both the swell and the track associated with the Hawaiian plume.

close to the hotspot. Elongated aseismic ridges may be formed by volcanics above channels along which plume material flowed to the spreading ridge (for example, in the Musicians Seamounts near 200° E, 25° N) (Figure 1). Hotspots and their tracks are less obvious on continents. The Yellowstone National Park area in the United States is frequently regarded as a hotspot plume, with the Snake River Plain being its hotspot track.

Seismic Images of the Mantle Plumes

Column-like anomalies of low seismic velocity associated with high temperatures are expected under hotspots, if hotspots are the surface expression of mantle plumes. Seismic imaging of hotspots has advanced in the past decade, and seismic images beneath some hotspots have been obtained. Commonly, these are imaged only at specific depths. For some hotspots, no low-velocity anomalies have been found.

Upper Mantle

Global mapping of the upper mantle by long-period (50–300 s) surface waves has revealed low seismic velocities associated with hotspots and spreading ridges. There is a distinct difference between hotspots and ridges, concerning the depth extent of these low-velocity regions : low seismic velocities beneath hotspots extend to a depth of 200 km, whereas low seismic velocities beneath ridges are confined to the upper 100 km. This suggests that hotspots are caused by active upwellings (mantle plumes) with deeper sources, compared to ridges, which may be caused by passive upwelling. The 200-km depths of the slow velocities under hotspots do not necessarily correspond to the actual source depths of mantle plumes, but rather to the depths to which surface waves can resolve. Seismic array observations have been carried out in hotspot regions to resolve fine structures such as plume conduits. A recent example of an S-velocity model beneath the Icelandic hotspot was obtained from body and surface wave data recorded by a temporary seismic array; a low-velocity plume can be seen beneath the hotspot (Figure 2). A 200-km-thick low-velocity zone extends laterally beneath Iceland; a vertical column of low velocities under central Iceland extends to a depth of at least 400 km. Similar array observations carried out at other hotspots (e.g., Hawaii, Yellowstone, and Massif Central) detected low-velocity anomalies extending to sublithospheric depths in the upper mantle.

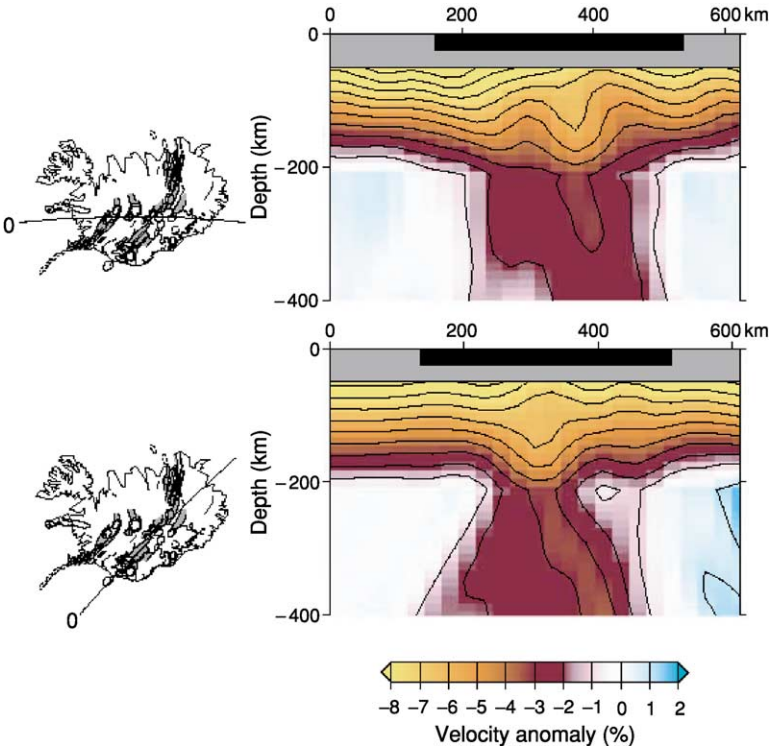


Figure 2 The S-velocity profile beneath the Icelandic hotspot. Reproduced with permission from Allen RM, Nolet G, Morgan WJ, *et al.* (2002) Imaging the mantle beneath Iceland using integrated seismological techniques. *Journal of Geophysical Research—Solid Earth* 107(B12): 2325, doi:10.1019/2001 JB000595.

The Transition Zone

The mantle transition zone is bounded by seismic discontinuities at depths around 410 and 660 km (referred to hereafter simply as '410' and '660'); the discontinuities at these two depths are interpreted to be due to olivine–β-olivine and postspinel phase changes, respectively. Topography of the 410 and 660 provides important data for understanding thermal structure and the ascending process of hot plumes; the 410 and 660 are expected to be depressed and elevated, respectively, in and near a hot plume, because the olivine–β-olivine and the postspinel phase changes have positive and negative Clapeyron slopes, respectively. The thickness of the transition zone (defined as the interval between the 410 and 660) is also used as a measure for thermal anomalies, because estimates are less affected by upper mantle velocity structure and therefore can be more reliably determined. Discontinuity depths are determined by detecting waves reflected or converted at the discontinuities, and measuring their timings. Using P- to S-converted waves recorded by regional seismograph arrays, the transition zone beneath Iceland and Hawaii has been estimated to be thinner than in the surrounding mantle by ≈20 and ≈40 km, respectively, for areas 400 km in diameter. Amounts of thinning can be converted to temperature anomalies, with the Clapeyron slope obtained experimentally, suggesting temperatures of ≈150 K and ≈300 K beneath Iceland and Hawaii, respectively. The transition zone beneath the Society hotspot in the South Pacific Superswell was estimated, using S waves reflected beneath the discontinuities, to be 20–30 km thinner than it is in the surrounding area (an area 500 km in diameter), suggesting temperatures that are 150–250 K hotter (Figure 3). These results suggest that mantle plumes beneath these hotspots originate at least from transition zone depths, and possibly from the lower mantle.

Lower Mantle and D″ Layer

To explore the deeper structure beneath hotspots requires analysis of body waves, which penetrate much deeper in the mantle. Global tomographic studies using body waves in the lower mantle have identified possible 'plume conduit signatures' beneath some hotspots, but not all. At present, different studies using different data and methods give somewhat inconsistent results for the same hotspot, making it difficult to derive definitive conclusions about the geometry of plume conduits and the source depths of hotspots. Different depth extents may indicate various source depths for hotspots. Two robust features – broad, low-velocity anomalies – have been identified beneath the South Pacific and Africa (Figure 4); often referred to as 'superplumes', these features have diameters of 1000–2000 km. The latter is tilted from south-west to the north-east, suggesting that the plume conduit may be advected by large-scale mantle flow. These superplumes may be related to the hotspots in the South Pacific Superswell and Africa, although robust seismic images of the plume conduits to each of the hotspots have not been obtained.

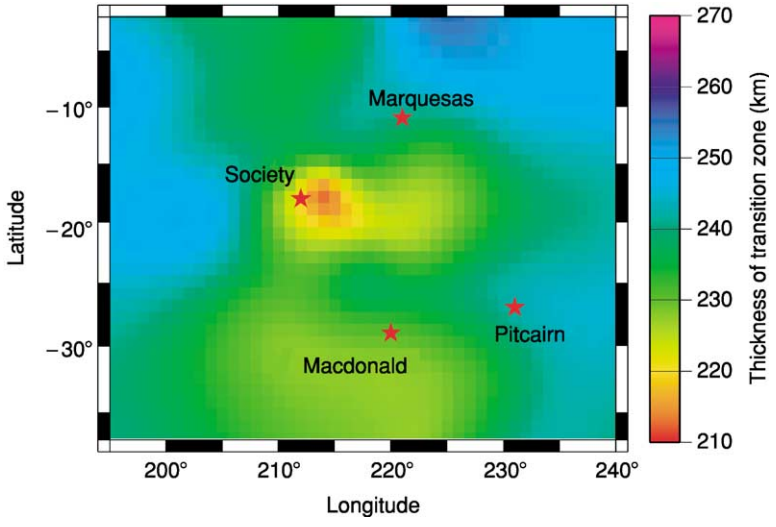


Figure 3 Thickness of the transition zone beneath the South Pacific Superswell. Deviation from the global average is shown (colour bar, kilometres). Stars represent hotspots. Reproduced with permission from Niu F, Solomon SC, Silver PG, Suetsugu D, and Inoue H (2002) Mantle transition-zone structure beneath the South Pacific Superswell and evidence for a mantle plume underlying the Society hotspot. *Earth and Planetary Science Letters* 198: 371–380.

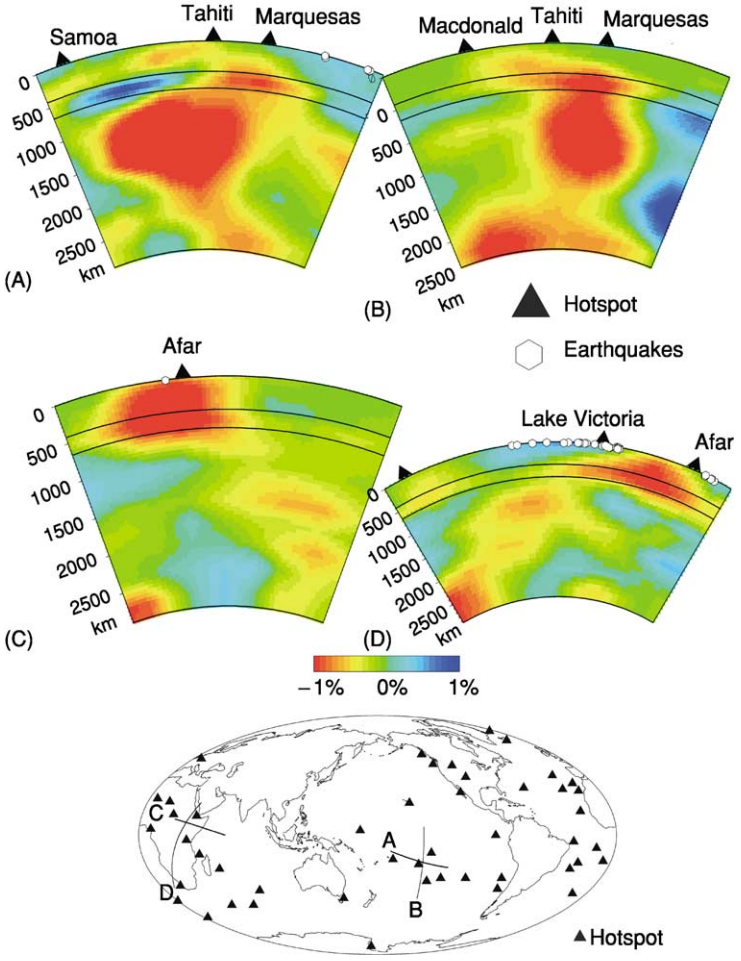


Figure 4 The P-velocity profile of the whole mantle beneath the South Pacific Superswell (top) and Africa (middle). (Bottom) Locations of profiled areas. Reproduced with permission from Zhao D (2001) Seismic structure and origin of hotspots and mantle plumes. *Earth and Planetary Science Letters* 192: 251–265.

Over most of the lower mantle depth range, lateral variation of seismic velocities is less than it is in the upper mantle, but pronounced lateral variation does occur at depths below 2500 km. Broad, low-velocity anomalies located beneath the South Pacific Superswell and Africa seem to be linked to the superplumes; high-velocity anomalies beneath circum-Pacific regions may be associated with subducted slabs. At the bottom of the D″ layer (the lowermost part of the mantle), above the core and mantle boundary (CMB), 5- to 50-km-thick zones of very low velocities have been found; these ultra-low-velocity zones (ULVZs) have low P and S velocity anomalies (10% or greater) and may represent a layer of partial melting and/or chemical heterogeneities. There is a spatial correlation between the ULVZs and hotspots, suggesting that the ULVZs may be a possible source of hotspots. Many plume conduits may be too narrow to be detected by present seismic data and techniques; dense seismic arrays on the land and seafloor and new analysis techniques may be required to resolve mantle plumes at full depth.

Petrological and Geochemical Signatures of Hotspot Rocks

Volcanic rocks associated with plume-related (hotspot) volcanism are quite variable in rock type and chemical composition and show significant correlation with the tectonic environment in which they erupt. Within the oceanic environment, hotspot volcanism produces mainly basaltic lavas, often called ocean island basalts (OIBs); these have chemical characteristics distinct from those of mid-ocean ridge basalts (MORBs), of which most of the ocean floor is composed. Hotspot volcanism in the oceanic environment sometimes produces voluminous oceanic plateaus or rises (e.g., the Ontong Java Plateau). In

continental environments, mantle plume activity is commonly manifested by eruption of huge floods of basaltic lavas, commonly called continental flood basalts (CFBs). Less voluminous magmatic activities also occur in continents with extensional tectonics such as intracontinental rifts and grabens. Volcanics in continental rift systems are thought to be products of mantle plume activity, although some continental rift volcanism is not clearly related to mantle plumes.

OIBs are commonly olivine-bearing lavas with subalkalic (tholeiitic) and alkalic compositions. Some large ocean islands, such as Hawaii and Iceland, are dominantly composed of tholeiitic basalts and their derivatives, with small amounts of alkalic rocks at later stages of eruptive sequences. A vast majority of ocean islands are composed entirely of alkalic basalts and their derivatives. The major element composition of tholeiitic OIBs is generally similar to that of MORBs, although lavas with higher MgO content are more abundant than at mid-oceanic ridges. Alkali olivine basalt is the most abundant type in alkalic OIB suites, but more alkalic lavas, such as basanite and nephelinite, also commonly erupt in many ocean islands. Compared with MORBs, both tholeiitic and alkalic OIBs have higher concentrations of incompatible trace elements, and their chondrite-normalized rare earth element (REE) patterns show variable but strong enrichment in light REEs relative to heavy REEs. Alkalic OIBs are typically more enriched in incompatible elements, including light REEs, compared to tholeiitic OIBs. Sr–Nd–Pb–Os isotopic ratios of OIBs show quite large variations relative to MORBs, which extend from nearly MORB-like values towards one component with highly radiogenic Pb isotopes (designated by the term 'HIMU', referring to the high U/Pb source ratios) and other components with nonradiogenic Pb isotopes (designated as 'enriched mantle 1', or EM1) and with high $^{87}Sr/^{86}Sr$ ratios (EM2). Variation in helium isotopic ratios of OIBs extends considerably towards the primordial value with high $^{3}He/^{4}He$ ratios, although some OIB suites have slightly lower $^{3}He/^{4}He$ ratios than MORBs have. Volcanic rocks of oceanic plateaus and rises are commonly aphyric lavas with tholeiitic composition. Their major element composition is relatively homogeneous compared to tholeiitic OIBs and more akin to MORBs. Oceanic plateau basalts have trace element characteristics similar to those of MORBs, but they are slightly more enriched in incompatible elements. Chondrite-normalized REE patterns are typically flat or slightly enriched in light REEs. Radiogenic isotope ratios of oceanic plateau basalts are quite variable, but their entire range is smaller than that of OIBs.

Petrological features of CFBs are similar to those of oceanic plateau basalts. Most CFB suites are composed of voluminous aphyric lavas with tholeiitic composition, and their major element compositions cover similar ranges as those covered by MORBs. However, compared to MORBs, CFBs generally have higher FeO/MgO ratios, and SiO_2-rich (andesitic) lavas are more dominant in CFBs. Most CFB suites are associated with minor amounts of alkalic rocks. In contrast to major elements, trace element features of CFBs are more like those in OIBs than in oceanic plateau basalts and MORBs. CFB suites are highly enriched in incompatible elements, sometimes even more than OIBs are, and have REE patterns with strong enrichment in light REEs. Sr–Nd–Pb–Os isotopic ratios of CFBs are much more variable than those of OIBs are, extending beyond the highest $^{87}Sr/^{86}Sr$ and the least radiogenic Pb isotopes of OIBs. Volcanic rocks associated with continental rift systems show considerable diversity in rock type, from transitional olivine basalt-rhyolite to strongly alkalic nephelinite–phonolite series, melilitites, and even carbonatite lavas (Ol Doinyo Lengai, Tanzania). A noticeable feature of continental rift volcanics is that acidic rocks, such as rhyolite, trachyte, and phonolite, are more abundant than in other environments. Enrichment of incompatible elements in continental rift volcanics is as great as or sometimes much greater than in alkalic OIBs. Variations of radiogenic isotope ratios are as wide as those in CFBs.

Geochemical diversity in plume-related volcanic rocks is generally attributed to chemical heterogeneity of mantle plume material. In particular, isotopic variations of OIBs, which are least affected by crustal material, clearly indicate that mantle plumes contain several geochemical components. The origin of such geochemical components in the mantle is still in dispute, but the most commonly accepted model is that subducted crustal materials (oceanic basaltic crust plus silicic sediments) contribute to formation of distinct geochemical reservoirs in the mantle, such as HIMU, EM1, and EM2. Chemical interaction between core and mantle has also been proposed to explain isotopic diversity of OIBs, especially for Pb and Os. High $^{3}He/^{4}He$ ratios in some OIB suites require the existence of some primordial volatile component in mantle plumes. Extremely radiogenic Sr isotopic ratios observed in some CFBs and continental rift volcanics have been explained by contamination or assimilation of continental crust. Major element diversity of plume-related volcanic rocks has been interpreted in terms of potential temperature of mantle plumes. A conventional view is that tholeiitic basalts in relatively large ocean islands and CFB suites are produced at higher degrees of melting in hot

mantle plumes, and alkalic rocks are produced at lower degrees of melting in colder plumes. On the other hand, the major element diversity of plume rocks can be explained by heterogeneity in major element composition of mantle plumes, because subducted crustal material, if contained in mantle plumes, may produce both tholeiitic and alkalic magmas under mantle conditions. In this case, no significant excess temperature is required for the plume, because the melting temperature of crustal materials is much lower than that of mantle peridotite.

Dynamics of Mantle Plumes

In Earth's mantle, the Rayleigh number, a measure of the vigour of convection, is estimated to be 5×10^6 to 5×10^7. For this range, time-dependent convection, with instabilities originating at thermal boundary layers, is expected. Plume formation may, however, be suppressed by large-scale flow related to plate motions, which, if sufficiently fast, advects growing instabilities towards large-scale upwellings before individual plumes can form. For somewhat slower large-scale flow, individual plumes may exist, but are advected towards, and hence cluster in, regions of large-scale upwellings. A thermal boundary layer exists above the core–mantle boundary, which is therefore a likely plume source depth. Another thermal boundary and possible source depth is at the 660. Plumes may also originate at chemical boundaries within the mantle. No conclusive evidence exists for those, but anticorrelation of bulk sound and shear wave anomalies in the lowermost mantle indicates that the D″ layer at the base of the mantle may be chemically distinct from the overlying mantle. Topography of the chemical boundary could be up to a few 100 km, and plumes may rise from the high points (cusps) on the boundary. Plumes may entrain material from the underlying layer; the amount of entrained material depends on the density difference between the two layers, due to chemical and temperature differences. Both laboratory and numerical experiments show that a plume that is less viscous than the surrounding mantle tends to have a large, roughly spherical or mushroom-shaped head, followed by a narrow tail (conduit) connecting the head and the source region (Figure 5). Scaled to Earth dimensions, a head diameter of several 100 to 1000 km is expected, if it rises from the lowermost mantle. Such a size is also required to explain the volume of many large igneous provinces, thus adding support to the notion that these originate from upwellings from the lowermost mantle, whereas hotspots not associated with flood basalts may be caused by plumes from shallower depth. Conduits should be only ≈100 km wide. Thinner conduits may occur for strongly temperature-dependent viscosity. If mantle viscosity increases strongly with depth, conduit diameter could be up to a few hundred kilometres in the lower mantle. Ascent times of plume heads through the whole

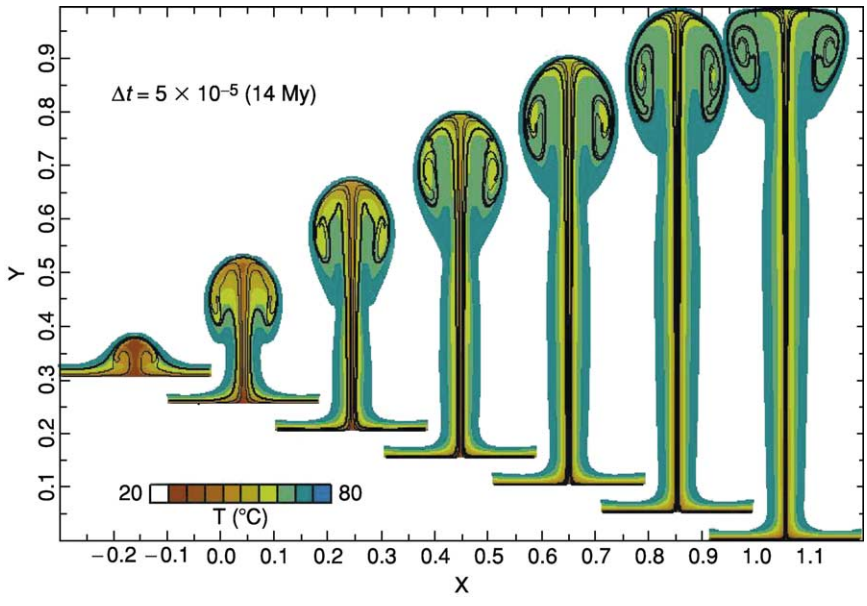


Figure 5 Numerical model of an axisymmetric mantle plume with strongly temperature-dependent viscosity. Colours indicate temperature (red = hottest, blue = ambient mantle); black lines indicate marker chains. Plots are at 14-My time intervals. Reproduced with permission from van Keken P (1997) Evolution of starting mantle plumes: a comparison between numerical and laboratory models. *Earth and Planetary Science Letters* 148: 1–11.

mantle are estimated to be ≈10 My to several tens of millions of years. For nonlinear mantle rheology, decrease of effective viscosity due to larger stresses around plume heads may cause rise times shorter than for Newtonian viscosity. As a plume head rises, the surrounding mantle heats up and becomes buoyant and less viscous. Mantle material is entrained into the rising plume, which therefore contains a mixture of materials from the source region and ambient mantle. Formation of a flood basalt is thought to occur when a plume head reaches the base of the lithosphere. Subsequently, the conduit may remain in existence as long as hot material is flowing in at its base – for 100 My or longer. The conduit consists of a narrow core, where most of the material transport occurs, and a thermal halo. Due to thermal diffusion, heat is lost from plume conduits as they traverse the mantle. Though strong plumes, such as the Hawaii plume, are not significantly affected, heat loss significantly reduces the temperature anomaly expected for weaker plumes. For weak plumes from the core–mantle boundary, with anomalous mass flux of $\lesssim$500–1000 $m^3 s^{-1}$, the sublithospheric temperature anomaly is low and no melting is expected. Weak hotspots may therefore have shallower origins. On the other hand, the temperature anomaly of strong plumes, inferred from observations, is much less than expected from the temperature drop across a thermal boundary layer between core and mantle. This may indicate a chemically distinct layer at the base of the mantle, with plumes rising from its top. Mantle plumes may coexist with superplumes, and conduits are expected to be tilted and distorted in large-scale mantle flow. The rising of a tilted conduit may cause further entrainment of ambient mantle material. If the tilt exceeds ≈60°, the conduit may break into separate diapirs, which may lead to extinction of the plume. As a consequence of conduit distortion, overlying hotspots are expected to move. Hence, mantle plumes probably do not provide a fixed reference frame. However, if plumes arise from a high-viscosity lower mantle, hotspots should move much more slowly than lithospheric plates move. Conduits are likely to be time variable, with disturbances traveling along them; these may be wave-like or may take the shape of secondary plume heads. Waves are associated with increased conduit flux, which may explain flux variations in mantle plumes. Ascending plumes interact with mantle phase transitions. The 660 somewhat inhibits flow across but is unlikely to block penetration of plumes. Experiments involving high pressure suggest phase relations of a pyrolite mantle such that, at the high temperatures of mantle plumes, this phase boundary does not hinder flow across. Beneath the lithosphere, buoyant plume material flows out of the conduit, spreads horizontally in a low-viscosity asthenosphere, and is dragged along with moving plates. Plume material buoyantly lifts up the lithosphere and causes a hotspot swell. Partial melt extraction at the hotspot may leave behind a buoyant residue that also contributes to swell formation. Plume material does not necessarily erupt directly above the conduit. It may also flow upward along the sloping base of the lithosphere, and enhanced melting may occur at steep gradients. A sloping base exists near spreading ridges. If ridge and hotspot are less than a few hundred kilometres apart, eruption of volcanics may occur at the ridge rather than, or in addition to, directly above the plume. Also, in other cases, such as in Africa, the spatial distribution of plume-related melting and magmatism may be controlled by the lithosphere rather than by the plume position. Formation of vertical fractures and ascent of magma through the lithosphere preferably occur for tensile lithospheric stresses. Loading of the lithosphere by hotspot islands causes stresses that may influence formation of fractures and therefore determine the spacing of hotspot islands along tracks. Feeding of plume material to a nearby ridge may put the lithosphere above the plume under compression and shut off eruption directly above the plume. If a hotspot (e.g., Iceland) is located close to the ridge, the viscosity contrast between plume and ambient mantle may become a factor 1000 or more beneath thin lithosphere. Such large viscosity variations facilitate ridge-parallel flow of plume material and help to explain geochemical anomalies south of Iceland. Propagation of pulses in plume flux explains the V-shaped topography and gravity anomalies at the Reykjanes Ridge. Probably not all intraplate volcanism is caused by plumes as described. In many cases, the origin of intraplate volcanism may be shallow, due to cracks in the lithosphere caused by tensional stresses, or due to edge-driven convection at locations where lithospheric thickness varies laterally.

See Also

Igneous Processes. **Large Igneous Provinces**. **Lava**. **Plate Tectonics**. **Rocks and Their Classification**. **Seamounts**. **Tectonics:** Propagating Rifts and Microplates At Mid-Ocean Ridges; Seismic Structure At Mid-Ocean Ridges.

Further Reading

Allen RM, Nolet G, and Morgan WJ, *et al.* (2002) Imaging the mantle beneath Iceland using integrated seismological techniques. *Journal of Geophysical Research—Solid Earth* 107(B12): 2325, doi:10.1019/2001 JB000595.

Courtillot V, Davaille A, Besse J, and Stock J (2003) Three distinct types of hotspots in the Earth's mantle. *Earth and Planetary Science Letters* 205: 295–308.

Dickin AP (1995) *Radiogenic Isotope Geology*. Cambridge: Cambridge University Press.

Hofmann AW (1997) Mantle geochemistry: the message from oceanic volcanism. *Nature* 385: 219–229.

Jackson I (ed.) (1998) *The Earth's Mantle: Composition, Structure, and Evolution*. Cambridge: Cambridge University Press.

Mahoney JJ and Coffin MF (eds.) (1997) *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism*. American Geophysical Union Geophysical Monograph 100. Washington, DC: American Geophysical Union.

Montelli R, Nolet G, Dahlen FA, Masters G, Engdahl RE, and Hung S-H (2004) Finite-frequency tomography reveals a variety of plumes in the mantle. *Science* 303: 338–343.

Nataf H-C (2000) Seismic imaging of mantle plumes. *Annual Review of Earth and Planetary Sciences* 28: 391–417.

Niu F, Solomon SC, Silver PG, Suetsugu D, and Inoue H (2002) Mantle transition-zone structure beneath the South Pacific Superswell and evidence for a mantle plume underlying the Society hotspot. *Earth and Planetary Science Letters* 198: 371–380.

Ritsema J and Allen RM (2003) The elusive mantle plume. *Earth and Planetary Science Letters* 207: 1–12.

Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge: Cambridge University Press.

van Keken P (1997) Evolution of starting mantle plumes: a comparison between numerical and laboratory models. *Earth and Planetary Science Letters* 148: 1–11.

Wilson M (1989) *Igneous Petrogenesis*. London: Unwin Hyman.

Zhao D (2001) Seismic structure and origin of hotspots and mantle plumes. *Earth and Planetary Science Letters* 192: 251–265.

Zindler A and Hart S (1986) Chemical geodynamics. *Annual Reviews for Earth and Planetary Sciences* 14: 493–571.

MARS

See **SOLAR SYSTEM: Mars**

MERCURY

See **SOLAR SYSTEM: Mercury**

MESOZOIC

Contents

Triassic

S G Lucas, New Mexico Museum of Natural History, Albuquerque, NM, USA
M J Orchard, Geological Survey of Canada, Vancouver, BC, Canada

Introduction

In 1834, the German geologist Frederich August von Alberti coined the term 'Triassic' for rocks originally recognized in Germany as the Bunter, Muschelkalk, and Keuper formations. Today, the rocks of Triassic age (~200 to 251 million years ago) are recognized on all continents (Figure 1). Most of these are sedimentary rocks consisting of dominantly shallow-water carbonates of marine origin and siliciclastic red beds of non-marine origin. These rocks represent a record of sedimentation on and around the vast Pangaean supercontinent and tell the tale of its final union and the initiation of its subsequent fragmentation. In this brief overview of the global Triassic, the rock record, time-scale, palaeogeography, tectonics and sedimentation, sea-levels, climate, and biota of the time period are considered.

Triassic Rocks

Triassic rocks are exposed on all the world's continents (Figure 1). Estimates of their maximum thickness have been given as 9 km, and their total volume is ~45 million km^3. These estimates are slightly more than the values estimated for the Permian, but substantially less than estimates for the Jurassic or Cretaceous. Triassic rocks are mostly sedimentary in origin, and volcanic rocks do occur in relatively minor amounts: they have been estimated as constituting ~1–2% overall of the exposed Triassic rocks in the Americas. Triassic volcanic rocks can be substantial in some regions, however, as exemplified in the Pacific north-west of North America.

The Triassic was a time of great continental emergence due to a combination of widespread epeirogenic uplift and relatively low sea-level. Marine deposition was mostly confined to the Tethys, the circum-Pacific, and the circum-Arctic (Figure 1). A worldwide survey identifies 15 significant Triassic outcrop belts: (1) the Cordillera of the western United States and western Canada, which exposes significant accumulations of both marine and non-marine strata, as well as a substantial record of Triassic rocks in accreted terranes; (2) the Newark Supergroup non-marine rift basins of eastern North America; (3) extensive marine and non-marine deposits of eastern Greenland, Franz Josef Land, and Svalbard; (4) western Europe, from the dominantly non-marine deposits of the Germanic basin system to the dominantly marine strata of the northern Mediterranean; (5) the extensive, dominantly marine deposits of north-eastern Siberia; (6) shallow marine deposits in Israel; (7) marine deposits in the Transcaucasian region of Iran and Azerbaijan, which include some very fossiliferous sections of the Permian-Triassic boundary (PTB); (8) dominantly marine strata of the Caspian Basin and Mangyshlak Peninsula of western Kazakhstan; (9) the Himalayan belt from Afghanistan and Pakistan through Kashmir into Tibet, also the location of some very fossiliferous PTB sections; (10) extensive non-marine deposits of the Junggur and Ordos basins of northern China; (11) marine deposits of southern China, South-east Asia, and Indonesia, including the most well-studied PTB section at Meishan in China and the phenomenal ammonite-bearing beds of Timor; (12) mixed marine-non-marine deposits on the western and eastern coasts of Australia; (13) extensive marine deposits in New Zealand; (14) deep marine deposits in Japan; and (15) non-marine strata exposed in the Transantarctic Mountains of Antarctica.

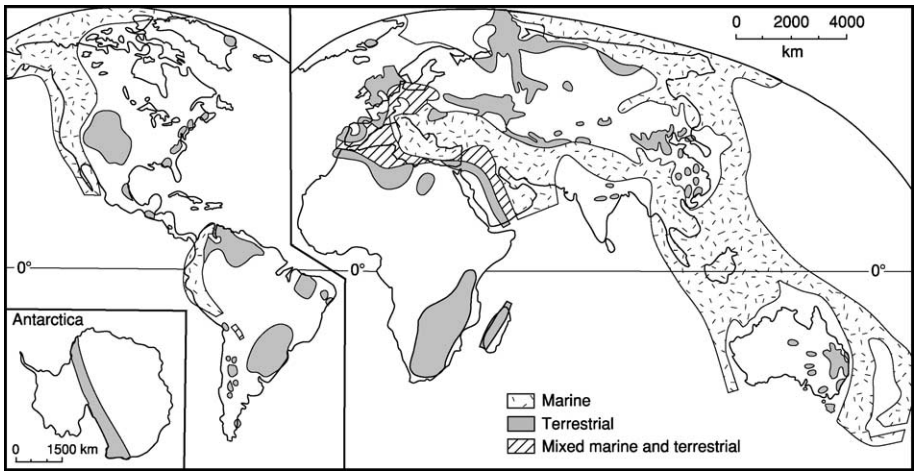


Figure 1 Map of the present world, showing the distribution of Triassic rocks. After Lucas (2000) The epicontinental Triassic, an overview. *Zentralblatt für Geologie und Paläontologie Teil I* 7-8: 475–496.

Time-Scale

The standard global chronostratigraphic scale for the Triassic is divided into seven stages. In ascending order, these are the Induan, Olenekian, Anisian, Ladinian, Carnian, Norian, and Rhaetian (Figure 2). Many other stage names of various scope and utility are employed regionally. For example, the Lower Triassic Griesbachian, Dienerian, Smithian, and Spathian stages are based on Arctic successions and have widespread currency in North America. Similarly, Tethyan substage names remain widely used in Europe. Definition and subdivision of the Triassic stages have been based largely on ammonoid biostratigraphy. However, only a single Global Stratotype Section and Point (GSSP) has been defined for the Triassic time-scale (for the Induan), and this is based on a conodont first occurrence. This datum defines the base of the Triassic system within the stratotype section at Meishan in southern China. International agreement and definition of the bases of the remaining six stages are under consideration.

The relatively low level of Triassic volcanism results in a dearth of numerical ages to provide geochronological control of the time-scale. Nevertheless, some important advances have been made in the past decade. The most intensively studied is the Permian–Triassic boundary at Meishan, South China, where U/Pb ages measured from zircons in ash beds both above and below the defined boundary provide a range of ages around 250 to 252 Ma. An age of 251 Ma for the PTB seems most consistent with ages derived from other PTB sections (Figure 2). The age of the Lower–Middle Triassic (Olenekian–Anisian) boundary is now determined from the Nanpanjiang Basin in Guizhou and Guangxi Provinces, South China. Several ash bands bracketing this interval have been dated recently, and the boundary age seems confidently set at ~247 Ma. One staggering implication of this is that the Early Triassic would appear to be on the order of only 4 million years long!

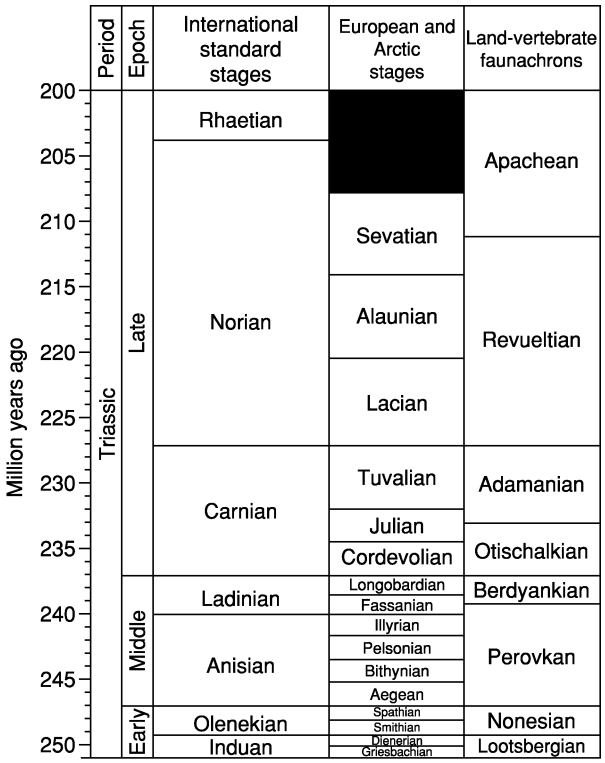


Figure 2 A simplified Triassic time-scale. After Lucas (2000) The epicontinental Triassic, an overview. *Zentralblatt für Geologie und Paläontologie Teil I* 7-8: 475–496.

Several radioisotopic ages are available from around the Anisian–Ladinian boundary. One potential position, at the base of the *Nevadites secedensis* ammonoid zone in Italy, is dated with U/Pb at

~240 Ma, whereas tuffs within the underlying *Reitziites reitzi* ammonite zone in Hungary give maximum ages closer to 241 Ma. The Upper Triassic stages are poorly constrained radioisotopically. The Ladinian–Carnian boundary approximates 237 Ma based on U/Pb ages from Italy and Hungary. Hence, the Upper Triassic is over half the length of the duration of the Triassic Period. Estimates for the Carnian–Norian boundary have been based on extrapolation from lithological cyclicity recognized in the non-marine Newark Supergroup of eastern North America. These cycles are interpreted as a response to Milankovitch orbital periodicities and can thus be measured in absolute terms. This astronomically calibrated numerical scale is set against the magnetostratigraphic polarity profile, which, in turn, can be correlated with successions elsewhere. However, there is no current agreement either on a position for a marine-based boundary or on precise correlations of the fossil scale with the Newark succession. Independent data come from an Ar/Ar age of a tuff that underlies Adamanian (=Late Carnian) tetrapods in Argentina. This gives an age of ~228 Ma. A Triassic–Jurassic boundary age of 202 Ma has been based on the ages of Newark Supergroup basalts that overlie the palynologically determined boundary. However, the palynological correlation of the boundary is debated and a more recent U/Pb age from British Columbia gives a Triassic–Jurassic boundary age very close to 200 Ma.

There is essentially no preserved Triassic seafloor, so there is no agreed geomagnetic polarity time-scale for the Triassic. However, a composite polarity timescale is now becoming available, based on successions being cobbled together from non-marine and marine sections in North America, Europe, and Asia. With continued refinement, this scale will be an important supplement to that provided by biostratigraphy in both the marine and non-marine realms.

Palaeogeography

At the onset of the Triassic, the world's continents were assembled into a single supercontinent called Pangaea (*see* **Pangaea**) (Figure 3). The rest of the globe comprised a single vast ocean called Panthalassa, with a westward-extending arm called Tethys. This followed the Late Palaeozoic assembly of the continents when Laurentia, Asia, and Gondwana collided along the Alleghanian–Variscan–Ural mountain chains. The nearly hemispheric Pangaean supercontinent was encircled by subduction zones that dipped beneath the continents while the Panthalassan and Tethyan plates carried island arcs and oceanic plateaus that were destined to become accreted to the continental margins.

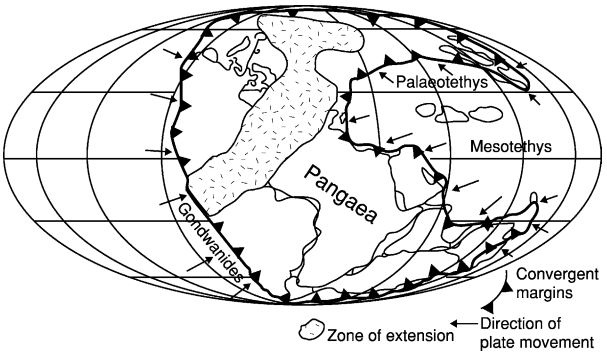


Figure 3 Triassic Pangaea, showing major tectonic elements. After Lucas (2000) The epicontinental Triassic, an overview. *Zentralblatt für Geologie und Paläontologie Teil I* 7-8: 475–496.

The supercontinent drifted northward and rotated clockwise throughout the Triassic, so there was considerable latitudinal spread to the landmass, which was nearly symmetrical about the equator (Figure 3). However, no sooner had the supercontinent been assembled than significant fragmentation began. Thus, Gondwana and Laurasia began to separate in Late Triassic time with the onset of rifting in the Gulf of Mexico basin. It was not until the Early Jurassic, though, that significant marine sedimentation took place in the nascent Atlantic Ocean basin.

Tectonics and Sedimentation

At the broadest level, the tectonics of Triassic Pangaea were simple. The accreted supercontinent was simply surrounded by convergent margins (Figure 3). However, these margins were actually complex belts of magmatic arcs and terranes moving in various directions to produce compressional and transpressional tectonics. Late Triassic Pangaea was the site of widespread extensional tectonism, especially the initial opening of the Atlantic Ocean basin by rifting of the North American and African plates. During the Late Triassic, in the Tethys, North Atlantic, and Arctic, multidirectional rift systems developed (Figure 3). Rifting also took place along a zone of transforms that extended well into the Gulf of Mexico basin and, punctuated by volcanism, dominated the northern border of western Tethys. This rifting in the North Atlantic and Tethyan regions subjected western and central Europe to progressive regional extension, culminating in the development there of complex multidirectional systems of troughs and grabens. During the Early–Middle Triassic, terminal thrusting took place along the entire Gondwanan margin of Pangaea, which was followed in the Carnian by extension in southern South America and eastern Australia.

Most Triassic sedimentation took place in one of three types of basins: foreland, fore-arc, or extensional. Perhaps the best example of a Triassic foreland basin is the Karoo Basin of South Africa, a retroarc foreland basin originally formed by the collision of the palaeo-Pacific and Gondwana plates during the Late Carboniferous. In the Karoo Basin, 12 km of Carboniferous-Jurassic red beds accumulated. Most of the Pangaean marginal basins were part of an array of arc-trench systems that surrounded much of the supercontinent. A good example is the complex Cordilleran basin of western North America, in which deposition took place between an offshore island arc and the continental margin. In the western United States portion of this basin, 1.2-km siliciclastic red beds were shed to the north-west and interfinger with marine carbonates deposited in the arc-trench system.

Of the (mostly Late Triassic) extensional basins, perhaps the best studied is the Newark basin in the eastern United States. This was a dip–slip-dominated half graben in which ~7 km of mostly lacrustine Upper Triassic–Lower Jurassic sediments accumulated. There were also other types of Triassic extensional basins more complex than the Newark half grabens, such as those of the Germanic basin system of north-western Europe.

Sea-Levels

Early Mesozoic plate reorganization was apparently associated with the development of new seafloor-spreading axes, which caused a general reduction of ocean basin volume during the Triassic. Pangaea was very emergent and, because of its high freeboard, the Triassic was a time of relatively low sea-level, which may be termed a first-order Pangaean global lowstand. After the major sea-level fall of the latest Permian, sea-level apparently rose through much of the Triassic, to peak during the Norian and then fall near the end of the period (Figure 4). There were, however, short significant falls in sea-level, especially during the Ladinian and Carnian. There are generally five second-order transgression–regression cycles recognized in the Triassic: these encompass the Lower Triassic and Middle Triassic and the Carnian, Norian, and Rhaetian.

Glacio-eustasy could not have driven Triassic sea-level change, so its underlying cause must be tectonism. Indeed, a large amount of regional Triassic tectonism has been invoked to explain sea-level changes in the western Tethys and the Arctic Sverdrup Basin. Triassic sequence boundaries caused by local tectonism or global eustasy show a remarkable degree of synchrony across Pangaea, and 12 high-order boundaries of global extent have been identified and attributed to episodic, major plate tectonic reorganizations.

Climate

Triassic climates marked the transition from the Late Palaeozoic ice house to the Mid-Late Mesozoic greenhouse (Figure 5). During the Triassic, there were no glacial ages, and there is no evidence of pack ice in the boreal or austral realms. The Triassic was thus a time of increased warmth with relatively wide subtropical dry (desert) belts at 10° to 30° latitude, as attested to by the broad latitudinal distribution of

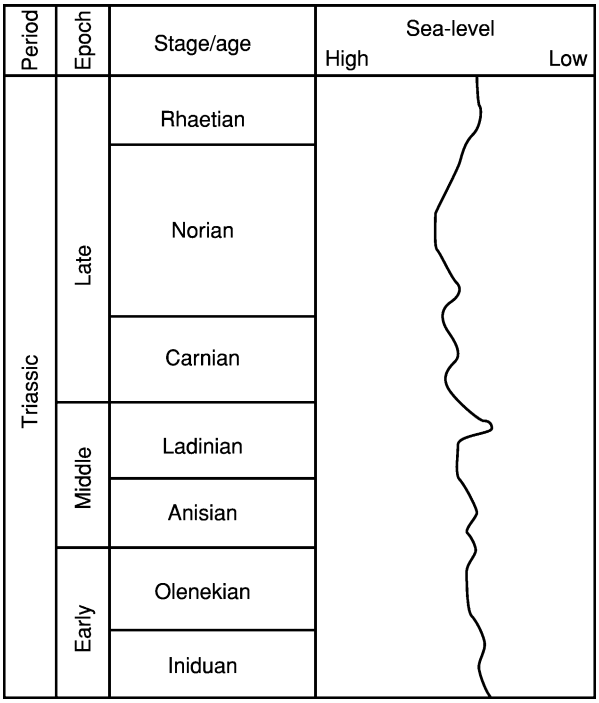


Figure 4 Triassic sea-level curve. After Lucas (2000) The epicontinental Triassic, an overview. *Zentralblatt für Geologie und Paläontologie Teil I* 7-8: 475–496.

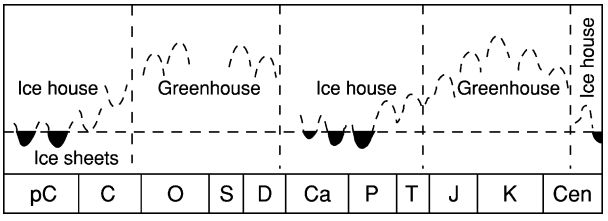


Figure 5 Triassic climate was a transition between the Late Palaeozoic ice house and the Late Mesozoic greenhouse. pC, Precambrian; C, Cambrian; O, Ordovician; S, Silurian; D, Devonian; Ca, Carboniferous; P, Permian; T, Triassic; J, Jurassic; K, Cretaceous; Cen, Cenozoic. After Lucas (2000) The epicontinental Triassic, an overview. *Zentralblatt für Geologie und Paläontologie Teil I* 7-8: 475–496.

Triassic evaporites. There was also strong east-west climatic asymmetry across Pangaea, with eastern Pangaea (at least between latitudes 40° S and 40° N) being relatively warmer and wetter because of the presence of Tethys and the absence of an Atlantic Ocean to facilitate oceanic heat exchange.

With the Pangaean landmass centered near the equator during the Triassic, and a prominent Tethyan bight, climate models suggest that seasonality was monsoonal. Hence, there were only two seasons, wet and dry. The abundant rainfall was concentrated in the summer months, and there was little annual temperature fluctuation. During the northern hemisphere summer, the northern landmass would have been relatively hot, whereas the southern land mass would have been relatively cool. Moisture from Tethys would have been pulled into the northern hemisphere low-pressure cell, producing extensive rains, whereas the southern hemisphere high-pressure cell would have remained relatively dry. During the southern hemisphere summer, this process would have occurred in reverse. Thus, seasonality across Triassic Pangaea would have been alternating hemisphere-wide wet and dry seasons. The warm and highly seasonal climates (wet–dry) of Triassic Pangaea are reflected in its biota. The Triassic saw an increase in the diversity of gymnosperms, particularly of xeromorphic scale-leaved conifers and seed ferns and cycadophytes with thick cuticles. Similarly, during the Triassic, in the evolution of reptiles, more water-efficient (putative uric-acid-excreting) diapsids diversified at the expense of less water-efficient (probably urea-excreting) synapsids.

Extinctions

The Permian ended with the greatest biotic extinction of Phanerozoic history (here termed the PTB biotic crisis) (*see* **Palaeozoic:** End Permian Extinctions). This extinction is best documented in the marine realm (Figure 6), where it is estimated that ~90% of the species, and more than half of the families of shelled marine invertebrates, became extinct. The magnitude and synchrony of the terrestrial extinction are much less clear. The Triassic records the recovery of the global biota from this massive extinction. The period also bore witness to further marine extinctions within the Late Triassic, and was terminated by a series of Late Triassic marine and non-marine extinctions.

The cause of the PTB biotic crisis remains uncertain. Some workers have identified a complex and interrelated group of terrestrial events as a possible cause: (1) major marine regression that reduced marine shelfal habitat areas and increased climatic

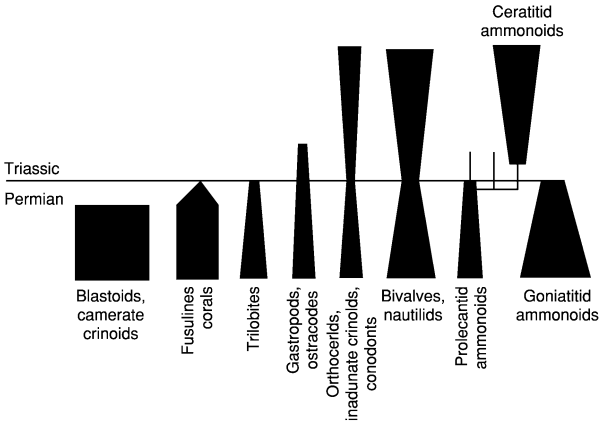


Figure 6 Characteristic extinction/diversity patterns of marine invertebrates across the Permian–Triassic boundary. After Lucas (2000) The epicontinental Triassic, an overview. *Zentralblatt für Geologie und Paläontologie Teil I* 7-8: 475–496.

variability, (2) eruption of Siberian flood basalts, (3) release of gas hydrates and erosion/oxidation of marine carbon due to the regression, and (4) elevated atmospheric CO_2 due to all of these phenomena resulting in ocean anoxia and global warming. During the Carnian, there was a substantial marine extinction of many kinds of conodonts, ammonoids, bivalves, echinoids, and reefal organisms, although the impact on land was less obvious, with evolutionary turnover occurring throughout the Late Carnian and Early Norian. A further extinction has been identified at the Norian–Rhaetian boundary with the disappearance of the ubiquitous flat clam *Monotis*. This, in fact, was part of a series of Late Triassic extinctions that included the disappearance of the conodonts, near extinction of the ammonites, decimation of about half of the marine bivalves, and collapse of the reef ecosystem. On land, there were also profound extinctions of tetrapods between the end of the Triassic and sometime in the Middle Early Jurassic (Sinemurian), but it has been difficult to establish the exact timing. A major carbon isotope anomaly has been identified in both marine and terrestrial environments at the end of the Triassic. This major perturbation in the global carbon cycle has been variously linked to a significant fall in sea-level, extraterrestrial impact, flood-basalt volcanism, and/or methane release.

Flora

During the Permian and Triassic, there was a complex and prolonged replacement of the palaeophytic flora by the mesophytic flora. This was the global change from pteriodophyte-dominated floras of the Palaeozoic to the gymnosperm-dominated floras that characterized much of the Mesozoic. Thus, the

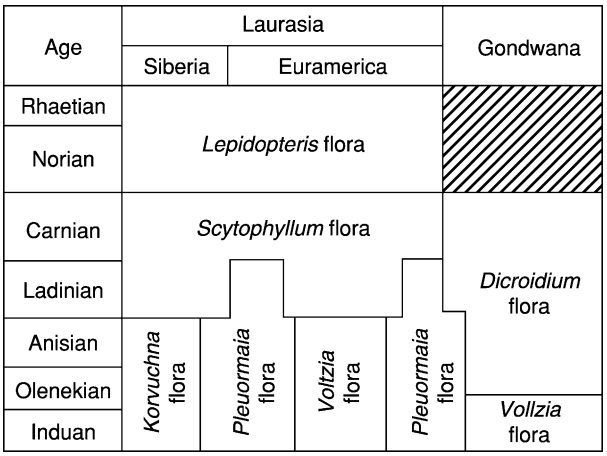


Figure 7 Triassic floral provinces and floras. After Lucas (2000) The epicontinental Triassic, an overview. *Zentralblatt für Geologie und Paläontologie Teil I* 7-8: 475–496.

arborescent lycopods and sphenopsids gave way to Triassic floras dominated by seed ferns, ginkgophytes, cycads, cycadeoids, and conifers. Distinct Gondwanan and Laurasian floras can be recognized, and within Laurasia two or three provinces are recognized – more boreal Siberian and more equatorial Euramerican provinces (Figure 7). However, the endemism of these floral provinces was not great. Triassic Laurasian floras were dominated by primitive conifers, ferns, cycads, bennettitaleans, and sphenopsids. Conifers were the dominant large trees, whereas the other plant types formed the understory. In coastal settings, stands of the lycopsid *Pleuromeia* were dominant.

Gondwanan floras of the Triassic were dominated by a wide range of seed ferns, especially the genus *Dicroidium*. These floras were generally composed of only a few (no more than 10) genera. *Dicroidium* was dominant in a variety of vegetation types, from heath to broad-leaved forest to dry woodland. Other important elements of Gondwanan floras were conifers and some Laurasian groups of cycadaleans and ginkgoes. Near the end of the Triassic, the *Dicroidium* flora declined and was replaced by a cosmopolitan conifer–benettitalean flora. From the end of the Permian until the Middle Triassic, there is a global coal discontinuity – there are no Early Triassic coal beds. This has been attributed to either an extinction of peat-forming plants at the PTB or to unfavourable tectonic conditions for coal preservation, though unfavorable climatic conditions for coal formation and preservation may also have been a factor.

Shelled Marine Invertebrates

Late Palaeozoic seas were dominated by pelmatozoans, brachiopods, and bryozoans, but molluscs

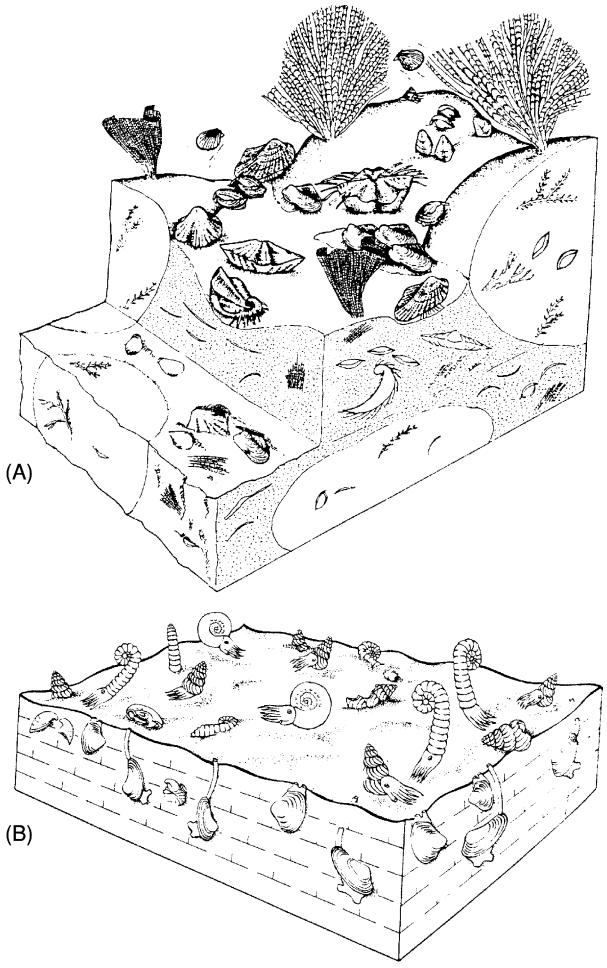


Figure 8 Reconstruction of (A) a characteristic Permian seafloor dominated by brachiopods, bryozoans, and crinoids and (B) a Triassic seafloor dominated by molluscs. After Lucas (2000) The epicontinental Triassic, an overview. *Zentralblatt für Geologie und Paläontologie Teil I* 7-8: 475–496.

dominated the Triassic seas (Figure 8). Most prominent of the Triassic molluscs were ammonoid cephalopods their rapid diversification during the Triassic provides a fossil record by which Triassic time has long been measured. Most Triassic ammonoids were ceratitidans with relatively simple suture lines. These were descended from only two ammonoid stocks that survived the PTB crisis: the otoceratids and the xenodiscids. Triassic ammonoid genera define three broad marine palaeobiogeographic provinces – Tethyan, Boreal, and Notal, the last of which is not well differentiated. The ammonoid palaeobiogeography of Triassic Panthalassa was complex and remains little understood. Triassic nautiloid cephalopods appear to have undergone relatively little change at the PTB, but reached great diversity in the Triassic, only to suffer an extensive (but not complete) extinction near the end of the period.

Bivalves were common Triassic molluscs and underwent a substantial diversification. Earliest Triassic assemblages are dominated by epifaunal pteriomorphs and detritus-feeding nuculoids, and they are very abundant as fossils. The Middle–Late Triassic saw a diversification of arcoid, mytiloid, trigonioid, and veneroid genera. The thin-shelled bivalves (so-called flat clams) *Claraia, Daonella, Halobia,* and *Monotis* are characteristic Triassic forms widely used in biostratigraphy. In contrast to ammonoids and bivalves, Triassic gastropods are relatively uncommon and not particularly diverse. A well-described Early Triassic (Smithian) gastropod assemblage from the western United States contains many genera that are also known from the Permian. Younger Triassic gastropod faunas are more diverse, but still contain numerous Permian holdover genera. The major Mesozoic change in gastropods took place after the end of the Triassic.

Brachiopods, bryozoans, and crinoids did not suffer total extinction at the PTB, although their numbers were greatly reduced (**Figure 6**). They were relatively minor, but persistent, components of Triassic marine faunas. More interesting is the distribution of corals and other reef-building organisms, which are virtually unknown in the Early Triassic (an exception are basal Triassic *Renalcis* biostromes in south China). In Middle Triassic time, Permian-type reef communities were re-established by *Tubiphytes*, bryozoans, calcisponges, and calcareous algae. The Carnian–Norian marine extinction was followed by a rapid turnover of the reef-building organisms, so that Norian reefs were characterized by abundant scleractinian corals, probably a result of the evolution of coral–zooxanthellae symbiosis. This presaged the extensive radiation of, and reef building by, corals that typified the Jurassic.

Insects

There was a major turnover in insect orders during the PTB biotic crisis, followed by a Triassic adaptive radiation, especially of beetles and cockroaches. A review of the Gondwanan Triassic record of plants and insects supports the concept of a co-evolution that led to the establishment of most modern insect orders by the end of the period.

Fishes

Fishes underwent a significant diversification during the Triassic. This is particularly evident in the appearance of new kinds of primitive actinopterygians, lungfishes, hybodontid sharks, and coelacanths. The extent of the extinction of fishes at the PTB biotic crisis is uncertain, but it appears to have been more significant within the marine realm. Conodonts, regarded by many now as a primitive fish group, were relatively unaffected by the end-Permian extinction. For them, the major biotic turnover occurred at the end of the Griesbachian, and this was followed by an explosive radiation at the start of the Olenekian. By the end of the Early Triassic, stocks had dwindled and there was a paucity of genera for the remainder of the period. The conodonts, nevertheless, continued to evolve rapidly, and their tooth-like elements now provide very useful biostratigraphic markers. After a long record, stretching throughout the Palaeozoic, conodonts became extinct at the end of the Triassic.

Tetrapods

Tetrapod vertebrates dominated Triassic landscapes and underwent at least two successive evolutionary radiations and extinctions during the period. Early Triassic tetrapod faunas were very similar to those of the Late Permian in being dominated by a relatively low diversity of dicynodont therapsids and capitosauroid/trematosaurid temnospondyls. Most notable is the dicynodont *Lystrosaurus,* the broad geographic distribution of which has provided classic evidence of the integrity of Triassic Pangaea (*Lystrosaurus* fossils have been found in Antarctica, South Africa, India, China, and Russia).

Middle Triassic tetrapod faunas remained dicynodont and temnospondyl dominated. However, by this time, the shift toward archosaur domination of the terrestrial tetrapod fauna had begun. By Late Triassic time, dicynodonts were rare, and the temnospondyls were greatly reduced in diversity. Instead, archosaurs were the most abundant terrestrial tetrapods. It was at this time that new groups of tetrapods appeared – turtles, dinosaurs (**Figure 9**), crocodiles, pterosaurs, and mammals – making the Late Triassic one of the most significant junctures in the history of vertebrate life. Indeed, it is fair to say that during the Late Triassic, the Mesozoic tetrapod fauna was born.

Marine reptiles also had their highest diversity during the Triassic. These reptiles were huhpehsuchians, nothosaurs, thallatosaurs, and placodonts, groups for which the entire diversification period was confined to the Triassic, and ichthyosaurs and plesiosaurs, groups that became prominent marine predators throughout much of the Jurassic and Cretaceous. There appears to have been a substantial extinction of marine reptiles (loss of 64% of families) at about the Middle–Late Triassic boundary. Apparently, the Reptilia successfully and explosively invaded the marine realm after the PTB crisis, but their diversity diminished rapidly, possibly due to

Figure 9 Dinosaurs appeared during the Triassic. These 3-m-long skeletons of the Late Triassic theropod *Coelophysis* are from Ghost Ranch in northern New Mexico.

the overall Late Triassic lowering of the sea, which reduced the epicontinental seaways in which most of the marine reptiles lived.

Glossary

compressional tectonics Deformation of the crust through compression (pushing together).
cycadophytes A group of gymnosperm plants with compound leaves, including the cycads.
diapsids The lizard-like and ruling reptiles, including lizards, snakes, crocodiles, and dinosaurs.
extensional tectonism Deformation of the crust through extension (pulling apart).
fore-arc The region between an island arc and an oceanic trench.
foreland The region in front of a deformed area of the crust.
freeboard The difference between mean sea-level and mean continental altitude.
graben A block of crust dropped down along faults relative to blocks on either side.
gymnosperm A vascular plant with seeds that are not covered by an ovary, such as conifers.
mesophytic The time of intermediate land plants, approximately equivalent to the Middle Triassic, Jurassic, and Cretaceous.
palaeophytic The time of ancient land plants, approximately equivalent to the Palaeozoic and Early Triassic.
pteridophyte A fern-like division of vascular plants that reproduce by spores.
retroarc foreland basin A zone of thickened sediment (basin) and extensional tectonism behind an island arc, floored by continental crust.
synapsids Mammal-like reptiles, including the ancestors of mammals.
transpressional tectonics Deformation of the crust by a combination of strike-slip (horizontal) motion and oblique compression.
trough An elongate depression in the crust with gently sloping borders.
xeromorphic scale-leaved conifers A group of conifers with thick, scaly leaves that retain water and thus allow the plants to live in relatively dry climates.

See Also

Atmosphere Evolution. **Fossil Invertebrates:** Molluscs Overview. **Mesozoic:** Jurassic. **Microfossils:** Conodonts. **Palaeoclimates**. **Palaeozoic:** Permian; End Permian Extinctions. **Pangaea**. **Time Scale**.

Further Reading

Alberti F von (1834) *Beitrag zu einer Monographie des Bunten Sandsteins, Muschelkalks und Keupers, und die Verbindung dieser Gebilde zu einer Formation*. Stuttgart: Cotta.
Benton MJ (ed.) (1993) *The Fossil Record 2*. London: Chapman & Hall.
Callaway JM and Nicholls EL (eds.) (1997) *Ancient Marine Reptiles*. San Diego: Academic Press.
Dobruskina IA (1994) Triassic floras of Eurasia. *Österreichische Akademie Wissenschaften Schriftenreihe Erdwissen Kommission* 10: 1–422.
Embry AF (1988) Triassic sea level changes: Evidence from the Canadian Arctic archipelago. *SEPM Special Publication* 42: 249–259.
Embry AF (1997) Global sequence boundaries of the Triassic and their identification in the western Canada sedimentary basin. *Bulletin Canadian Petroleum Geology* 45: 415–433.
Erwin DH (1993) *The Great Paleozoic Crisis: Life and Death in the Permian*. New York: Columbia University Press.
Kummel B (1979) Triassic. In: *Treatise on Invertebrate Paleontology, Part A, Introduction. Fossilization (Taphonomy) Biogeography and Biostratigraphy*, pp. 351–389. Lawrence, KS: Geological Society of America and University of Kansas Press.
Lucas SG (1998) Global Triassic tetrapod biostratigraphy and biochronology. *Palaeogeography, Palaeoclimatology, Palaeoecology* 143: 347–384.
Lucas SG (2000) The epicontinental Triassic, an overview. *Zentralblatt für Geologie und Paläontologie Teil I* 7-8: 475–496.
Sherlock RL (1948) *The Permo-Triassic Formations*. London: Hutchinson's Scientific and Technical Publ.
Tozer ET (1984) The Trias and its ammonoids: the evolution of a time scale. *Geological Survey Canada, Miscellaneous Report 35*. Ottawa: Geological Survey Canada.
Yin H (ed.) (1996) *The Palaeozoic-Mesozoic Boundary. Candidates of the Global Stratotype Section and Point of the Permian-Triassic Boundary*. Wuhan, China: University of Geosciences Press.
Ziegler PA (1989) *Evolution of Laurussia*. Dordrecht: Kluwer Academic Publ.

Jurassic

K N Page, University of Plymouth, Plymouth, UK

Introduction

The Jurassic System is the second of the three systems comprising the Mesozoic Era. It takes its name from the Jura Mountains of eastern France and Switzerland, where, as long ago as 1795, its distinctive character was recognized, although it was not until 1829 that Alexander Brongniart first used the term 'Jurassique'. The stratigraphical and palaeontological meaning of the system was first clearly defined, however, by Alcide d'Orbigny's publication of his classic *Palaeontologie Française, terrains Jurassique* from 1842 to 1849.This work established a system of remarkably modern-looking stages, most of which are still in use today.

D'Orbigny's stages or '*étages*' were designed to be of worldwide use and were based on the assumption that periodic mass extinctions followed by the rapid re-establishment of new faunas characterized stage boundaries. The existence of a particular 'fauna' therefore correlated rocks belonging to a specific stage. Later research showed that these patterns are little more than an artefact of environmental changes in north-western Europe at the levels taken to represent stage boundaries by d'Orbigny, with faunal changes being more ecological than catastrophic.

Each of d'Orbigny's stages included a sequence of fossil 'zones', representing the general stratigraphic ranges of specific taxa. This term was further refined by Albert Oppel who, from 1856 to 1858 in his *Die Juraformation Englands, Frankreichs und des Südwestlichen Deutschlands*, developed a sequence of such divisions for the entire Jurassic System, crucially using the units in the sense of time divisions. Oppel used a very similar sequence of '*grupen*' or '*etagen*' to d'Orbigny, and these form the basis of today's Jurassic System.

Jurassic Stratigraphy and Chronology

Chronostratigraphy: Stages, Standard Zones, Subzones, and Horizons

Chronostratigraphy aims to produce rigorous definitions for named subdivisions of geological time, using actual rock units as standards for reference. This approach includes the designation of a Global Stratotype Sections and Points (GSSP), through international agreement, for the base of each geological time period (or 'System') and each of their component stages. The sequence of stages recognized in the Jurassic and their actual or candidate GSSPs are shown in Table 1. Below the level of the stage, subdivisions at the level of chronozones and ultimately zonules can be used. In the Jurassic, the frequent abundance of ammonoid cephalopod molluscs in marine sequences and their wide geographical distribution have led to their use for correlating sequences of 'standard zones'. Such units have, from the days of Oppel, been used in the same sense as the modern concept of chronozones, although they are still frequently confused with biozones, where the use of fossils in correlation has no

Table 1 Jurassic stages and their definitions

Stage	*Historical type locality*	*GSSP or candidate GSSPs*
Tithonian (152–145 Ma)	South-Eastern France, after the Goddess 'Tithon'	Mont Crussol, Ardèche/Canjuers, Provence, France
Kimmeridgian (155–152 Ma)	Kimmeridge Bay, Dorset, England	Staffin Bay, Isle of Skye, Scotland; Mont Crussol, Ardèche, France
Oxfordian (157–155 Ma)	Oxford, Southern Central England	Savournon/Thoux, Provence, France; Weymouth, Dorset, England
Callovian (160–157 Ma)	Kellaways, Wiltshire, England	Albstadt-Pfeffingen, Württemberg, southern Germany
Bathonian (166–160 Ma)	Bath, South-West England	Cabo Mondego, Portugual; Bas Auran, Provence, France
Bajocian (174–166 Ma)	Bayeux, Normandy, France	Cabo Mondego, Portugual (GSSP)
Aalenian (178–174 Ma)	Aalen, Württemberg, Germany	Fuentalsaz, Iberian Cordillera, Spain (GSSP)
Toarcian (183–178 Ma)	Thouars, Deux-Sèvres, France	Peniche, Portugual
Pliensbachian (192–183 Ma)	Pliensbach, Württemberg, Germany	Robin Hood's Bay, North Yorkshire, England
Sinemurian (197–192 Ma)	Semur-En-Auxois, Burgundy, France	East Quantoxhead, Somerset, England (GSSP)
Hettangian (200–197 Ma)	Hettange, Lorraine, France	St Audries Bay, Somerset, England; Muller Canyon, Nevada, USA; Chilingote, northern Peru; Kunga Island, BC, Canada

explicit meaning for geological time. As such, ammonite chronozones are always treated as simple subdivisions of chronostratigraphical stages, a usage that would be impossible if they were biozones.

Most Jurassic ammonoid chronozones are divided into subchronozones, largely for historical reasons and to achieve a degree of nomenclatural stability at the chronozone level. Classically, ammonoid zones have offered a very precise and detailed relative time-scale with which to date Jurassic geological sequences and events. Over the last 30 years or so, however, even smaller subdivisions or units have been established, in particular in Europe, which are known collectively as 'horizons'. There are conceptually two types of sub-subzonal 'horizon' in use in Jurassic ammonoid stratigraphy – first, zonules, which are the smallest component of a chronostratigraphical hierarchy, and, second, biohorizons, which can be used in the sense of faunal 'events'.

Both classes of intra-subzonal unit provide a high-resolution time-scale for the Jurassic, which is probably unique for pre-Quaternary sequences, with subdivisions averaging less than 200 000 years, and even potentially as little as 120 000 years in the Lower Jurassic. Their practical value is potentially immense and only just beginning to be realized: one of the most dramatic demonstrations of this is the tracing of the basal Callovian *Kepplerites keppleri* Biohorizon circumglobally in the northern hemisphere from Europe to Russia to Japan to Alaska to Arctic Canada to Greenland and back to Europe again, thereby precisely correlating widely spaced regions with a single faunal 'event' that potentially lasted less than 200 000 years.

Other fossil groups, especially microfossils, have been used to construct true biozonal schemes for the marine Jurassic in particular, but the resolution of these schemes is usually inferior to that of the ammonite scale, with a single microfossil biozone spanning several ammonite chronozones. Indeed, the latter scale is typically used as a 'standard' against which biozonal schemes are correlated. Nevertheless, where ammonites are very rare or absent (e.g. in non-marine or restricted-marine facies or in boreholes), they can be very useful. Continued research can only improve these correlations, and detailed sampling of sequences of dinoflagellate floras, for instance, has begun to demonstrate that they may even ultimately provide a correlative scale at least as detailed as that provided by ammonite chronozones.

Other Stratigraphical Methods in the Jurassic

Considerable advances have been made in recent years in establishing chemostratigraphy as a potential tool for correlating Jurassic sequences, in particular for sequences where high-resolution biostratigraphical tools are rare or lacking (e.g. continental sequences), for linking terrestrial with marine systems, and for linking different faunal provinces where detailed biochronostratigraphical correlations have yet to be established. A variety of isotope systems have been assessed, in particular carbon, oxygen, and strontium, but also osmium, sulphur, and nitrogen. Trace elements, such as magnesium and manganese, have also been shown to have some potential.

Initial results are promising but still need to be adequately integrated with established high-resolution biostratigraphical schemes. Some attempts to infer correlations solely on the basis of geochemical data are less persuasive when independently assessed using, for instance, ammonite zonal and biohorizonal schemes.

Similarly, magnetostratigraphy has the potential to provide a globally useful framework for correlating existing regional biostratigraphical and chronostratigraphical schemes, but it is essential that an accurate calibration is achieved between the two different 'time-scales' if the results are to be meaningful and applicable.

The economic significance of Jurassic sequences, in particular as sources of oil in Europe, has led to a massive development of sequence stratigraphical analysis, and, for specific depositional basins, the tool has a high value for the correlation of geophysical logs in particular.

Jurassic Geochronology

Absolute time-scales for the Jurassic have been somewhat elusive, owing to the rarity of reliably dated and correlated calibration points, with as few as five such being used until relatively recently. As a result various methods have been used to estimate, for instance, stage duration, in particular the persistent fantasy that ammonite zones average 1 Ma in duration. That such zones can contain as few as one or more than 19 correlatable subunits (as biohorizons or zonules) emphasises that different zones are likely to be of widely differing durations. Such assumptions are no longer tenable.

A contemporary and more reliable time-scale has been developed in Canada, however, which uses at least 50 calibration points, mainly in biostratigraphically well-constrained volcanic and volcaniclastic rocks. Nevertheless, there remain problems with the dating of the later Jurassic, from the Callovian to the Tithonian, as reliable data remain sparse. In particular, analysis of the number of biostratigraphical and chronostratigraphical units recognizable for each of the stages in this interval in Europe suggests that the Callovian and Oxfordian stages are too short and

the Tithonian is too long. This dating problem may be compounded by problems of interprovincial correlation in the Late Jurassic, when ammonoid provincialism was strong (Figure 1).

The Jurassic World

Continents

The Early Jurassic inherited the pole-to-pole massive supercontinent of Pangaea (*see* **Pangaea**). The opening of the North Atlantic began the process of dismantling this huge landmass, initially with South America moving away from North America to form the marine 'Hispanic Corridor'. The latter linked the previously separated East Pacific regions and the west Tethyan regions of Europe. There is circumstantial evidence from faunal migrations that this passage may have been intermittently open during the Lower Jurassic, providing a migration passage for shallow-marine faunas, but it was not until the later Middle Jurassic that it became a more open passage.

During the Middle Jurassic, oceanic crust began to form in the Atlantic, as rifting began to separate Europe from Greenland and North America. Despite this gradual opening up, however, the general layout of the continents maintained strong east–west and north–south physical barriers to marine faunal migrations throughout the period, leading to well developed bioprovincialism (Figure 1) that did not finally break down until the mid-Cretaceous.

Climate

Greenhouse effects dominate Jurassic climates worldwide. Lithological indicators, such as coals, evaporites, and aeolian sands, show a general symmetry of climate zones about the palaeoequator. Floral data

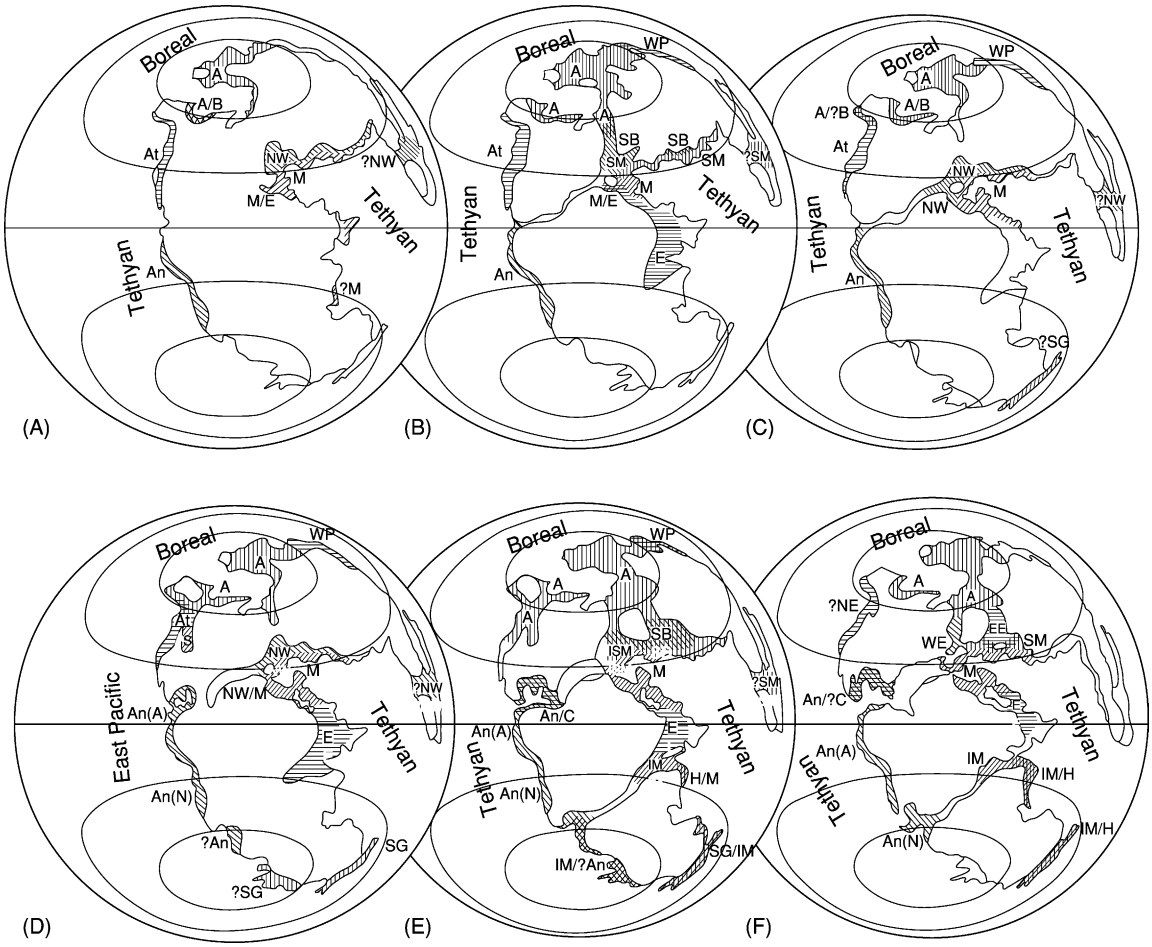


Figure 1 The evolution of the Jurassic world. (A) Hettangian–Lower Pliensbachian, (B) Upper Pliensbachian–Toarcian, (C) Aalenian–Lower Bajocian, (D) Upper Bajocian–early Lower Callovian, (E) Late Lower Callovian–Kimmeridgian, and (F) Tithonian–Valanginian. Ammonoid provinces as follows: Tethyan Realm – M, Mediterranean; NW, North-western European; SM, Sub-Mediterranean; H, Himalayan; E, Ethiopian; IM, Indo-Malgach; SG, Sula-New Guinean; C, Cuban; WP, West Pacific. East Pacific (Late Bajocian–Bathonian) or Tethyan Realm – At, Athabascan; S, Shoshonean; An, Andean. Boreal Realm – A, Arctic; SB, Sub-Boreal; WE, West European; EE, East European; B, Bering. Cross-hatching indicates where provincial barriers fluctuated.

suggest that the Jurassic world was one in which low latitudes were seasonally dry and were succeeded polewards in both hemispheres by desert, seasonally dry, warm temperate, and cool temperate biomes. Modelling studies have suggested that the tropical regions reached temperatures as high as 40°C during parts of the summer months, whilst winters were around 0–20°C. High-latitude oceans during the Jurassic were certainly significantly warmer than they are today and were generally ice-free.

Ocean palaeotemperature estimates for the European area range from 12°C to 30°C, again considerably warmer than today, with the Lower and Upper Jurassic appearing to be the warmest. The possible existence of Boreal and Austral realms, delineated by temperature, during parts of the Mesozoic, however, certainly lends support to the existence of cold conditions at the poles at certain times. General Circulation Model simulations of the Jurassic climate also reveal that high-latitude regions may have developed cold temperatures, providing adequate leeway to argue for either the presence or the absence of high-latitude ice. Only equivocal Jurassic tillites in Siberia have, however, been described. The Late Jurassic–early Cretaceous was characterized in many areas of the world by a sea-level lowstand, resulting in large semi-restricted epicontinental seas and widespread deposition of non-marine sediments. The latest Jurassic Purbeck group of the UK, which contains evidence of evaporite deposition, may be the local expression of the spread of an equatorial arid zone across Europe at this time.

A characteristic of Jurassic shelf sequences is a strong cyclicity, observed as sedimentary microrhythms, typically with an alternation of limestone and marl, although in clastic-dominated sequences this may manifest itself as an alternation between carbonate-rich and carbonate-poor mudrocks or even as coarser silty or sandy bands alternating with finer-grained claystones. The periodicity of these microrhythms is consistent with orbital forcing due to Milankovitch cyclicity (*see* **Earth:** Orbital Variation (Including Milankovitch Cycles)), with, for instance, the more carbonate-rich levels representing improved climates and consequently higher production by calcareous plankton. There has been much discussion about the potential of using microrhythms assignable to specific Milankovitch frequencies to improve or even calibrate Jurassic chronostratigraphy, although few attempts have been made to correlate results with high-resolution biostratigraphical or chronostratigraphical schemes, to ensure that local effects such as gaps in the record due to local processes (e.g. tectonic) do not compromise conclusions.

Key Events

The Early Jurassic saw a recovery from the devastating Late Triassic mass extinction, which was linked to the destruction of shelf seas and to climactic changes associated with the final assembly of the supercontinent of Pangaea. Radiations of many taxa followed, and extinction events that have been reported during the later Jurassic, for instance across the Pliensbachian–Toarcian and Callovian–Oxfordian boundaries, are probably consequences of no more than local or regional facies changes.

The Lower Toarcian in north-western Europe is characterized by a well-known anoxic event, whose causes and extent have been much discussed. Much of the latter depends on the dating of the event, which has been claimed to be global in extent. Nevertheless, precise correlations have not always been established and, even in Europe, there are possibly either two 'events' or a single diachronous anoxic event. For example, the ammonite chronology suggests that levels with the highest organic carbon contents are slightly older in Mediterranean/reas than in north-western Europe. It has also been suggested that the $\delta^{13}C$ minimum at this general level is related to a massive release of methane from gas-hydrate deposits, although a lack of synchroneity across Europe would make this scenario less likely. Perhaps more compelling, however, is the relatively simple explanation that anoxia developed in the stratified waters of the relatively enclosed and circulation-restricted marine basins of north-western Europe, perhaps as a result of a regional rise in sea-level, as suggested by facies changes and faunal migrations. The $\delta^{13}C$ minimum can be explained by the recycling of ^{12}C-enriched carbon derived from the remineralization of organic matter. Hence the 'global' significance of this event beyond Europe remains to be proved conclusively.

Jurassic Faunas and Floras, Including Key Groups with Stratigraphical Value

Benthic Algae (Including Chlorophyta (Dasycladales), Charophyta, and Rhodophyta)

Dasycladean algae that produce calcareous structures are locally sufficiently common in peri-Tethyan shallow-water carbonate-platform deposits to be useful locally as biostratigraphical markers, especially in the virtual absence of other stratigraphically useful groups. Coralline rhodophytes are sometimes important in Jurassic reefs and bioherms, contributing to their bound structure and to associated bioclastic sediments.

In freshwater carbonate sequences, the calcified stems and reproductive bodies ('oogonia') of charophytes lend themselves to preservation as fossils, and hence they may be locally abundant and even stratigraphically useful from the mid- to the Late Jurassic.

Planktonic Algae (Including Prynesisophyta, Coccosphaeales (Coccoliths) and Pyrrhophyta (Dinoflagellates))

Jurassic calcareous nannofloras, including coccoliths, are still relatively poorly known, especially in the Early Jurassic, owing to their great sensitivity to diagenesis. Nevertheless, they can be locally so abundant as to be rock-forming, especially in some fine-grained mud-rich carbonate sequences, such as pelagic deposits, where diagenesis has not destroyed traces of their structure – indeed they may well be the main contributor of calcareous material to many such fine-grained Jurassic marine carbonates. As knowledge improves, their stratigraphical value will improve, especially following the resolution of problems related to provincialism, including an apparently marked diachroneity of first occurrence across different regions.

The cysts of dinoflagellates, representing the dormant phase of the life cycle of a marine planktonic alga, are frequently abundant in Jurassic marine and quasi-marine sequences, especially those where the sediments are relatively organic rich. As such they can be excellent stratigraphical guide fossils in borehole studies, in particular in petroleum exploration, and, although correlative units are currently stratigraphically relatively long, the cysts' great diversity from the Middle Jurassic onwards hints at the potential, at least locally, to achieve a resolution virtually equivalent to that of ammonoid chronozones.

Protists

The calcareous or agglutinated tests of benthic foraminifera are frequently common in Jurassic marine sediments. In basins and platforms dominated by argillaceous sediments these associations are often dominated by Nodosariidae, although on some carbonate platforms Textularina may be more significant. Lower Jurassic assemblages are typically of relatively low diversity, following the end-Triassic mass extinctions, but diversity increases markedly through the Middle Jurassic. Planktonic foraminifera (Globigerinina) first appear in the Early–Middle Jurassic and by the early Upper Jurassic may be abundant in some sequences – although understanding of their distribution and biostratigraphy is often reduced by lumping them together under the term 'protoglobigerinids'. The stratigraphical use of Jurassic foraminifera is relatively limited, however, owing to the relatively long ranges of many taxa, although the group can be locally important for recognizing divisions on the scale of stages.

Jurassic radiolaria are most characteristic of relatively deep-water siliceous rocks, such as cherts, formed below the carbonate compensation depth. In such areas they can be valuable stratigraphical tools.

Porifera (Sponges)

Sponges (*see* **Fossil Invertebrates:** Porifera) can be locally common in Jurassic carbonate rocks, especially from the Middle Jurassic. In the Upper Jurassic, in particular, Hyalospongea (siliceous sponges) can locally be very important as rock formers in some fine-grained carbonate-platform sequences bordering Tethys, including building spectacular biohermal structures.

Cnidaria, Scleractinia (Corals)

The origin of hermatypic (reef-building) corals with symbiotic algae, in the Triassic, paved the way for the development of true coral reefs in the Jurassic, with a bound framework formed by the corals themselves – although typically in association with calcareous algae (including rhodophytes) and sometimes bryozoans and sponges. Coral patch-reefs are most typical where suitable warm shallow-water environments are developed, mainly in the Middle and Upper Jurassic.

Brachiopoda and Bryozoa

Although the last spiriferid brachiopods persist into the Lower Jurassic, the articulate orders Terebratulida and Rhynconellida dominate normal-marine Jurassic brachiopod faunas. Locally, in shallow-marine carbonate deposits these groups can be a major component of shelly faunas, even outnumbering bivalves. Despite strong provincialism, this local abundance gives the group biostratigraphical use, and resolutions more or less equivalent to that of ammonoid zones can be achieved across some basins or sub-basins. The remarkable inarticulate survivor *Lingula* persists in more restricted environments, especially in shallow quasi-marine environments but also occasionally in deeper-water situations where organic-rich mudrocks exclude other brachiopod groups.

Encrusting and coralline bryozoans with calcified skeletons are locally common in shallow-marine Jurassic carbonate sediments; the order Cyclostomata is particularly important and tends to dominate.

Mollusca

Jurassic bivalves are often extremely abundant in a great range of water-lain sediments, from fully

marine, through brackish, to freshwater – each environment has a characteristic range of taxa. As their morphology is tightly constrained by their ecology, they have been extensively used for palaeoecological studies – a process aided by the fact that virtually all Jurassic bivalve orders survive today. Some groups, in particular Trigonoida and some Pteroida (including some pectenids and the well-known oyster *Gryphaea*), can also have limited, albeit somewhat crude, biostratigraphical value. More useful, however, especially in some extreme Boreal areas in the Late Jurassic, are thin-shelled pseudo-planktonic buchids, and these form the basis of a biozonation, for instance in Arctic Canada and Russia. Certain Jurassic bioclastic limestones may be composed largely of bivalves in varying degrees of fragmentation, and occasionally epifaunal groups such as oysters and lithiotids may form small bioherm-like structures.

Jurassic gastropods are dominated by epifaunal archaeogastropods and mesogastropods and are considerably less diverse than the neogastropod assemblages of the later Cretaceous and Tertiary. Very occasionally, however, groups such as nereneiids may be very abundant in certain restricted-marine environments and may consequently even have local biostratigraphical value.

At the end of the Triassic the Ammonoidea faced a massive crisis, and perhaps only two genera survived into the Jurassic, the early phylloceratid *Rhacophyllites* and its direct descendant, the first ammonitine ammonite, *Psiloceras*. A massive evolutionary explosion followed, and by the Middle Jurassic five ammonoid suborders can be recognized, with a sixth appearing right at the end of the period (*see* **Fossil Invertebrates:** Ammonites). Phylloceratine and lytoceratine ammonites remain morphologically very conservative throughout the Jurassic and are most abundant in areas influenced by open ocean – studies of shell strength indicate that they could live at greater water depths than the other Jurassic suborders.

Other groups, such as the Ammonitina, Haploceratina, and Perisphinctina, are much more morphologically varied and thrived in epicontinental seas. They frequently show distinctive geographical distribution patterns, reflecting ecological and physical controls on individuals and populations (**Figure 1**). Such patterns are characterized as biogeographical provinces, and the inevitable consequence of using ammonites for correlation purposes is that every province, almost by definition, will have a different scheme of standard zones. These differences inevitably make interprovincial correlations at zonal level, and especially at subzonal and horizon levels, difficult. Nevertheless, faunal links often exist, frequently facilitating remarkably good interprovincial correlations, and ammonites remain the preeminantly high-resolution stratigraphical tools in the Jurassic.

Early Jurassic distributions continue the basic pattern of the Late Triassic, with a clearly distinguishable Arctic Boreal Realm and a much broader Tethyan Realm to the south. By the Middle Jurassic, however, communication with the western margin of the Americas was so restricted that a separate East Pacific Realm developed. The establishment of direct Boreal–Tethys connections in Europe in the late Middle and Upper Jurassic provided more possibilities to exchange faunas between the northern Arctic Sea and the trans-equatorial Tethys Ocean. Latitudinal controls on faunal distribution are nevertheless still evident, and a complex series of faunal belts developed, ranging from high-latitude restricted Boreal faunas, through sub-Boreal to sub-Mediterranean and eventually low-latitude diverse Mediterranean Province faunas. Analogous transitions from Boreal to more Tethyan faunas are also present in East Pacific regions at this time, although latitudinal faunal belts are only occasionally discernible southwards from the Jurassic equator towards the Antarctic.

Belemnoidea, in contrast, are never as diverse as Ammonoidea, although they can still show analogous bioprovincial distributions, with distinct Boreal and Tethyan assemblages. Their robustness in a sedimentological context gives them some value as stratigraphical indicators, occasionally at a resolution equivalent to around two ammonite zones. Much more important, however, is their use in geochemical studies, especially for assessing changes in oceanic chemistry (particularly related to temperature and climatic change), using oxygen, carbon, and strontium isotopes.

Crustacea, Including Ostracoda

Jurassic decapod crustaceans, although generally seen as body fossils only under exceptional conditions (such as rapid burial in a phosphate-rich environment), were apparently major contributors to the bioturbation of shallow-marine aerated sediments. Generally only burrow traces are seen, however, including the branching form *Thalassinoides*, which was made by a callianassid shrimp.

In contrast, the calcareous carapaces of microscopic Ostracoda can be very important environmental indicators, as different genera and species can be characteristic of marine or brackish-water deposits. They can also have value as stratigraphical indicator fossils. Although resolution tends to be at the level of substage and interbasin provincialism is potentially

well developed, when combined with results from the Foraminifera obtained in the same sieved residues, the group can be very useful for correlation purposes.

Uniramia, Insecta

Beetles and dragonflies are typical components of Jurassic insect faunas although their fossil record is very patchy in the Jurassic, prior to the more abundant record from Cretaceous amber. Although obviously representing terrestrial faunas, Jurassic insects not infrequently turn up in fine-grained organic-rich marine sediments where land was not too distant.

Echinodermata

By the Jurassic, Crinoidea had become much less important as producers of bioclastic sediment in shallow-marine environments than earlier, for instance in the Late Palaeozoic. They remain locally common, however, and planktonic and pseudo-planktonic forms can be locally significant, the former including the Upper Jurassic *Saccocoma* and the latter attached to drift wood.

Mobile benthic echinoderms, such as ophiuroids and holothurians, may have been more abundant in some marine Jurassic environments than their macrofossil remains may suggest, as their spicules or fragmented remains can be common in microfossil residues. Echinoids, including both burrowing 'irregular' and epifaunal 'regular' forms, are, however, more conspicuous in some marine deposits where seabed oxygenation was relatively good, and the former, in particular, can even have some local stratigraphical value.

Chordata (Vertebrates)

Jurassic fish faunas are characterized by a dominance of bony fishes (Osteicthyes), in particular advanced actinopterygians ('ray-finned fishes'). Teleosts, in particular, were on the rise. Chondrosteans (which include today's sturgeons) were in decline, as were coelocanths and lungfishes, which are both generally rare in the Jurassic. Whilst showing a reduced diversity compared with the Late Palaeozoic, chondrichthyans (cartilaginous fishes, including sharks and rays) remain relatively common, with neoselachian sharks becoming dominant (*see* **Fossil Vertebrates:** Fish). Otoliths (fish ear-bones) are well known from some Jurassic sediments.

Marine large-reptile faunas are very characteristic of the Jurassic, with dolphin-like ichthyosaurs, long-necked plesiosaurs, and shorter-necked pliosaurs being particularly significant – the latter including the world's largest known carnivore, *Liopleurodon*,

Figure 2 A typical marine scene from the Lower Jurassic, showing nektonic plesiosaurs, actinopterygian fishes, ammonites, and belemnites, with a benthos dominated by bivalves with gastropods and crinoids. Reconstruction based on a locality in southern England. (Reproduced with permission from © University of Bristol, UK.)

at over 15 m in length (Figure 2). Marine crocodiles include the highly adapted geosaurs, with limbs modified as paddles, but their terrestrial cousins also radiated extensively. Other terrestrial reptiles included true lizards (squamates). Pterosaurs become increasingly important through the Jurassic, although they do not achieve the great diversity of form typical of their Cretaceous descendants.

Of all terrestrial animals in the Jurassic, the dinosaurs will always remain the most emotive (*see* **Fossil Vertebrates:** Dinosaurs). From humble beginnings in the Triassic, the group rapidly diversified to dominate all terrestrial vertebrate faunas worldwide. Although Early Jurassic assemblages are still dominated by Triassic groups, during the Middle Jurassic, many well-known groups appeared, including the giant herbivorous sauropods, large carnivorous theropods, and armoured stegosaurs and ankylosaurs (Figure 3).

By the Upper Jurassic, some small dinosaurs had developed feathers, and by the Kimmeridgian the transition to birds was well underway with the famous *Archaeopteryx*. Although mammals had arisen in the latest Triassic, they remained small and

Figure 3 A typical Middle Jurassic terrestrial scene, showing dinosaurs (including a ceratosaur theropod, a stegosaur, a fabrosaurid ornithopod, and the sauropod *Cetiosaurus*), rhamphorynchid pterosaurs, a tritylodont mammal-like reptile (*Stereognathus*), a docodont mammal, and, associated with the pond, a goniopholid crocodile, a salamander, a disoglossid frog, and an actinopterygian fish. The vegetation includes horsetails, ferns, cycads, seed ferns, and conifers. Reconstruction based on faunal remains from a single locality in southern England. (Reproduced with permission from © University of Bristol, UK.)

relatively insignificant members of terrestrial faunas throughout the Jurassic.

Jurassic Terrestrial Floras

Jurassic floras are sporadically recorded where terrestrial sediments, especially those formed by rivers and in lakes, are well developed. Assemblages are typically dominated by ferns and various gymnosperm groups such as cycads, bennettites, caytonias, ginkgos, and conifers. Where conditions allowed the development of marsh vegetation, horsetails can also be abundant. Macrofloras are also well known, however, in certain shallow-marine sequences, although they are often somewhat different in composition from those of terrestrial (fluviolacustrine) facies, including more conifers and *Ptilophyllum* and *Sphenozamites* bennettites, with fewer ferns, ginkgos, cycads, and caytonias. Crucially, many families of extant ferns and conifers first appear in the fossil record in the Jurassic.

A characteristic feature of Jurassic floras is the relative lack of latitudinal variation, with northern 'temperate' floras in north-western Canada and Siberia being broadly similar to 'tropical' floras in the Early and Middle Jurassic of Europe, North America, Central Asia, and China. The main difference between these floras is a slight reduction in diversity in northern areas, with a dominance of ginkgos and lepidostrobaleans and with very few cheirolepidiacean conifers. Southern 'temperate' latitude floras, in contrast, were poor in ginkgos and dominated by the cheirolepidiacean conifers, with lepidostrobaleans being absent. These southern 'temperate' floras extended into very high near-polar latitudes in Antarctica, indicating the absence of polar ice.

The situation changed in the Late Jurassic owing to the development of arid conditions over much of the palaeotropics and parts of the southern mid-latitudes. The floras of these more arid regions became very low in diversity and were dominated by cheirolepidiaceans and a few ferns that were adapted to drier conditions (e.g., *Weichselia*). The floras of the northern high- and mid-latitudes and that part of the southern mid- and high-latitudes that escaped the aridification became much more diverse and more comparable with the floras of the palaeotropics in the Early and Middle Jurassic.

See Also

Earth: Orbital Variation (Including Milankovitch Cycles). **Fossil Invertebrates:** Ammonites; Porifera. **Fossil Vertebrates:** Fish; Dinosaurs. **Mesozoic:** Triassic; Cretaceous. **Pangaea**.

Further Reading

Arkell WJ (1956) *Jurassic Geology of the World*. London: Oliver & Boyd.

Behrensmeyer AK, Damuth JD, DiMichele WA, *et al.* (eds.) (1992) *Terrestrial Ecosystems Through Time*. Chicago: Chicago University Press.

Benton MJ (ed.) (1993) *The Fossil Record 2*. London: Chapman & Hall.

Cariou E and Hantzpergue P (eds.) (1994) *3ème Symposium International de Stratigraphie du Jurassique (Poitiers – France, 22–29 Septembre 1991*. Geobios Mémoire Spécial 17. Lyon: Université Claude Bernard.

Cariou E and Hantzpergue P (eds.) (1997) Biostratigraphie du Jurassique Ouest-Européen et Méditerranéen. *Bulletin du Centre Recherche Elf Exploration et Production, Mémoire* 17: 15–24.

Hall RL and Smith PL (eds.) (2000) *Advances in Jurassic Research 2000: Proceedings of the Fifth International Symposium on the Jurassic System (Vancouver, Canada, August 12–25, 1998)*. GeoResearch Forum 6. Zurich: Trans Tech Publications.

Ineson JR and Surlyk F (eds.) (2003) *The Jurassic of Denmark and Greenland*. Geological Survey of Denmark and Greenland, Bulletin 1. Copenhagen: Danmark og Grønlands Geologiske undersøgelse (GEUS).

Jenkyns HC, Jones CE, Gröcke DR, Hesselbo S, and Parkinson DN (2002) Chemostratigraphy of the Jurassic System: applications, limitations and implications for palaeoceanography. *Journal of the Geological Society* 159: 351–378.

Page KN (1995) Biohorizons and zonules: intra-subzonal units in Jurassic ammonite stratigraphy. *Palaeontology* 38: 801–811.

Page KN (1996) Mesozoic ammonoids in space and time. In: Landman NH, Tanabe K, and Davis RA (eds.) *Ammonoid Paleobiology,* pp. 755–794. Topics in Geobiology 13. London: Plenum Press.

Parisi G (guest editor) (2004) The 6th International Symposium on the Jurassic System, 16–19 September 2002. *Rivista Italiana di Paleontolgia e stratigrafa* 110(1): 428. Milano.

Riccardi A (ed.) (1996) *Advances in Jurassic Research, Proceedings of the Fourth International Symposium on the Jurassic System (Mendoza, Argentina, 1995)*. GeoResearch Forum 1, 2. Zurich: Trans Tech Publications.

Westermann GEG (ed.) (1992) *The Jurassic of the Circum-Pacific*. World and regional geology 3. Cambridge: Cambridge University Press.

Cretaceous

N MacLeod, The Natural History Museum, London, UK

Introduction

The Cretaceous System/Period is the last major subdivision of Mesozoic time. It was established by JJ d'Omalius d'Halloy in 1822 and was divided into Upper and Lower series/epochs by WD Conybeare and William Phillips that same year. Cretaceous rocks are currently assigned to 12 stage/age-level subdivisions, the combination of which represent an interval estimated to lie between 145.5 and 65.5 Ma. This is the longest single system/period of the Phanerozoic, and Cretaceous rocks are found on all continents. The Cretaceous is also the oldest system/period to be entirely represented in the ocean basins. Palaeogeographically, the Cretaceous represents the time during which Pangaea continued to fragment and continental plates began to move into their current positions. Prominent tectonic features of Cretaceous time include the rifting of Laurasia to form the northern and southern embayments of the proto-Atlantic Ocean, the joining of these embayments into a single north–south-trending Atlantic Ocean Basin, and the rifting of the southern supercontinent of Gondwana into South America, Antarctica, India, and Australia. During the Cretaceous, the sea-level stood high, flooding the interiors of most continental platforms and resulting in the establishment of marine chemical conditions that favoured the deposition of calcite. This, in turn, led to the widespread deposition, and subsequent preservation, of the rock type uniquely associated with the Cretaceous: chalk. As a result of this high sea-level, the Cretaceous was characterized by warm, equable, greenhouse-type climates over most areas, with temperatures that both exceeded and were more stable than those at present are. There was very little or no ice at the Cretaceous poles, and reefs, swamps, crocodiles, and even dinosaurs reached latitudes in the vicinity of – and, at times even greater than – 60°. Cretaceous atmospheric composition also differed from that of today, with higher levels of oxygen and carbon dioxide. The Cretaceous was a quiet time for magnetic reversals, but also a time of widespread volcanism. Biotically, the Cretaceous represents the culmination of many evolutionary–ecological trends begun in the Jurassic, including the diversification of many plant and animal lineages (e.g., diatoms, coccoliths, gymnosperms, angiosperms, foraminifera, ammonoids, molluscs, insects, dinosaurs, ichthyosaurs, plesiosaurs, mosasaurs, pterosaurs, and mammals). Finally, the Cretaceous includes two extinction events: an 'Aptian' event, which may be an analytic artefact, and the Cenomanian–Turonian event, in

which 27% of all marine genera disappeared as a result of the coincidental juxtaposition of sea-level rise, the (possibly volcanically accentuated) upwelling of oxygen-poor waters into the shallow chalk seas, and the effects of global cooling induced by more efficient marine circulation patterns. This period was also ended by a very large extinction event. (*see* **Mesozoic:** End Cretaceous Extinctions.)

Stratigraphy

Terrain Cretacé was the name originally given, in 1822, by the French geologist JJ d'Omalius d'Halloy, to a sequence of chalk beds, underlain by tufas, sands, and marls, that crop out in the structural basis of southern England, northern France, and Belgium. d'Omalius d'Halloy's usage followed that of William Smith's (*see* **Famous Geologists:** Smith) map, which had identified four sequences of strata between the Tertiary 'lower clay' and the (Jurassic) Portland stone. These were (from oldest to youngest) micaceous clay, also known as 'Brickearth', Greensand, brown or grey chalk, and white chalk. Later that same year, the English geologists WD Conybeare and William Phillips gathered these four units into two groups, the upper chalk facies and the lower, predominately clastic facies, and first used the term 'Cretaceous' to describe the entire stratigraphic package. Conybeare and Phillips' subdivision is reflected today in the fact that most time-scales recognize only two Cretaceous epochs (Upper and Lower), instead of the more typical three-fold subdivision of Upper, Middle, and Lower.

From 1840 to 1871, each Cretaceous epoch was further subdivided into six stages (Figure 1), based on rocks cropping out in France, Switzerland, and southern Holland. Cretaceous chronostratigraphy was originally based on molluscan biostratigraphy, especially the biostratigraphy of ammonites. These fossils were used to subdivide each stage into either threefold or twofold substage intervals. Beginning in the 1950s, however, Cretaceous biostratigraphic zonations were further refined through the use of microfossils, chiefly planktonic foraminifera (28 biozones) and calcareous nannoplankton (26 biozones). At least seven key radioisotopic tie points have been identified in Cretaceous sediments, providing good geochronometric control, especially for Upper Cretaceous stage boundaries.

The current base of the Cretaceous is undefined by a boundary stratotype, but is taken as being near the first occurrence of the ammonite *Berriasella jacobi*. The base of the Cenozoic Danian stage defines the top of the Cretaceous and was established in 1991 at the base of the 'boundary clay layer' associated

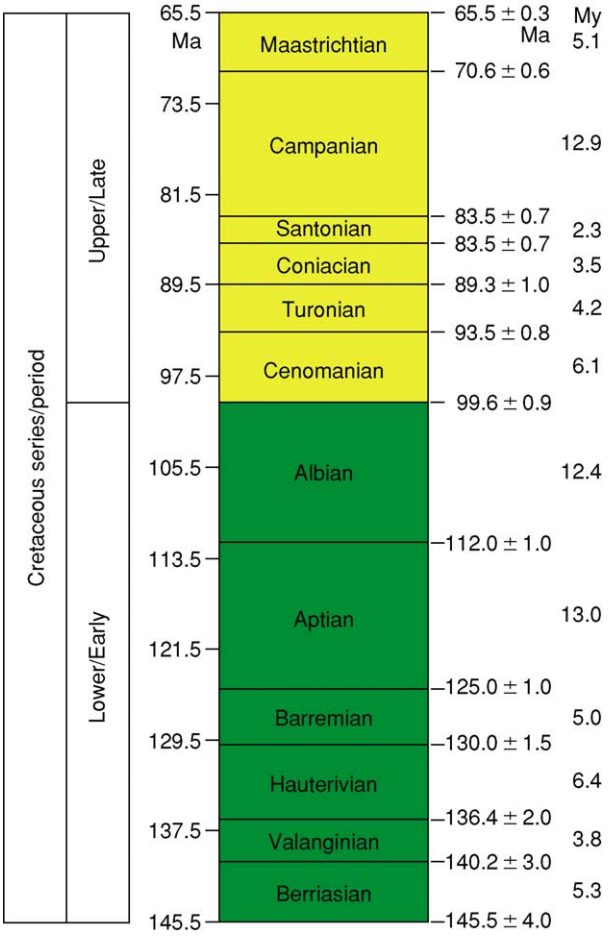


Figure 1 Chronostratigraphy and geochronometry of Cretaceous system/stages. Time-scale based on that of Gradstein and Ogg (*see* **Time Scale**). Stage thicknesses scaled to reflect relative durations. Interpolated duration estimates given in millions of years (far right column).

with a geochemical iridium anomaly and a major extinction of planktonic foraminifera and calcareous nannoplankton in the stratotype section outside the town of El Kef, Tunisia. This stratotype definition is somewhat unusual in that it is not taken at a biostratigraphic datum and coincides with a local extinction horizon rather with than a first occurrence horizon. It is also unfortunate that the El Kef stratotype was destroyed after 1991 due to local farming practices. Studies are currently under way to re-establish the stratotype in the Tunisian type area at an outcrop less susceptible to such damage.

The Cretaceous system has the greatest duration of any Phanerozoic stratigraphic system (76.5 My). Its distinctive sediments, especially the characteristic Upper Cretaceous chalk facies, are present on all large continental platforms. The Cretaceous is also the first stratigraphic system to be well represented in the deep sea, owing to the fact that most pre-Cretaceous

sediments deposited in the deep-ocean basins have been subducted.

Palaeogeography and Tectonics

At the beginning of the Cretaceous (Figure 2A), the post-Permian breakup of Pangaea had progressed to the point at which the northern (Laurasian) continents of North America, North China, Siberia, and Eurasia had rifted away from Gondwana (Africa, South America, India, Antarctica, Australia), though the latter remained coherent. Gondwana was, in turn, separated from North America by a narrow Atlantic Seaway, and from North China by a broad Tethys Ocean. Surrounding this region of intense tectonic activity, the Pacific Ocean occupied fully half of the surface of the planet.

In terms of continental landmasses, the primary Cretaceous event was the rifting of Gondwana. Near the Jurassic–Cretaceous boundary (~140 Ma), a large rift between Africa and India formed and propagated south until, by 150 Ma, Gondwana has

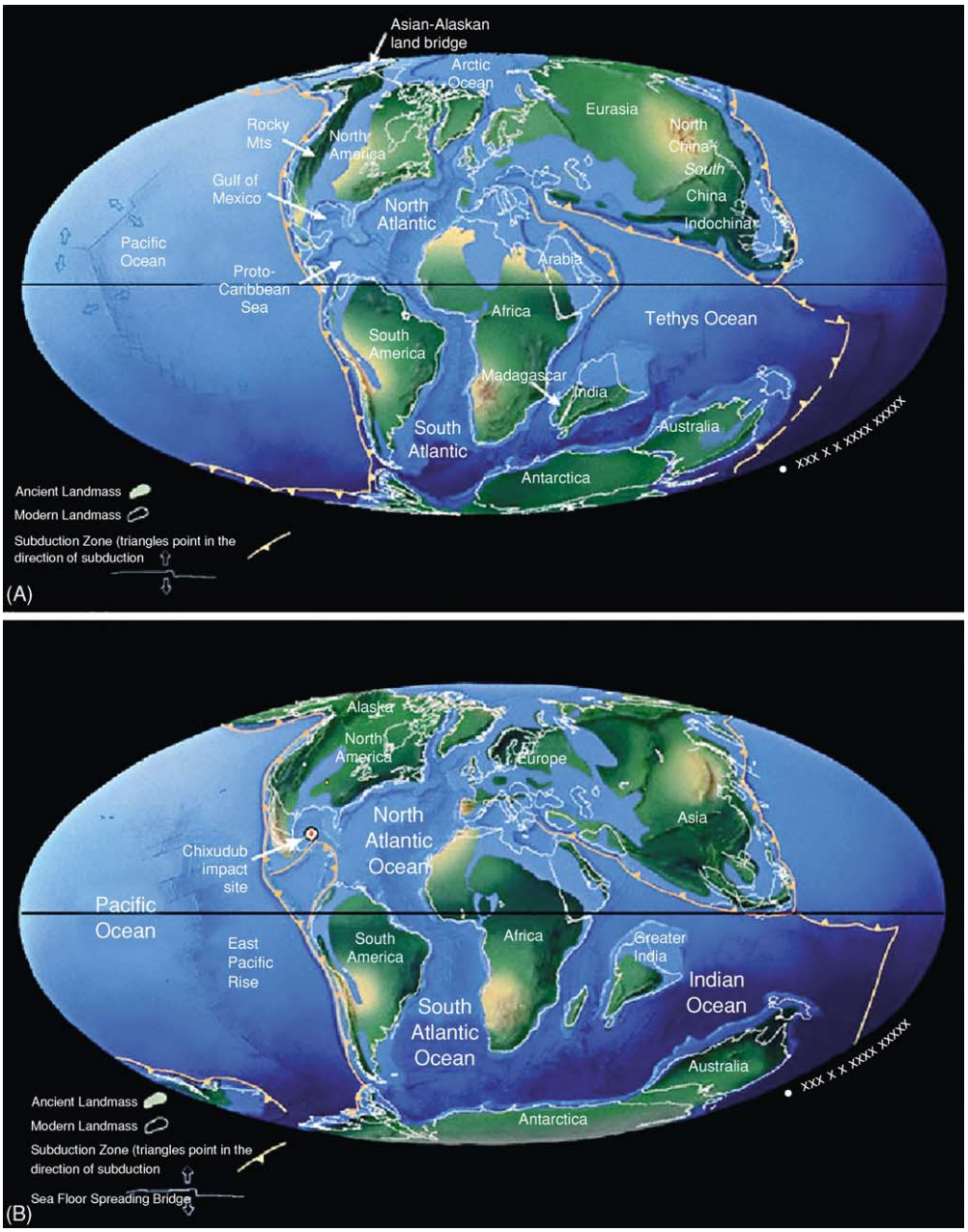


Figure 2 Reconstructions of Cretaceous palaeogeography at (A) 94 Ma (Cenomanian) and (B) 66 Ma (Maastrichtian). See text for discussion. Palaeogeographical maps generated as part of the GeoMap Project. Reprinted with permission.

Table 1 Cretaceous large igneous provice eruptions (from Courtillot *et al.*, 1996; Courtillot, 1999)

Name	*Location*	*Assoc. mantle plume*	*Age (Ma)*	*Stage*
Deccan Traps	India	Reunion	65	Maastrichtian–Danian
Madagascar Traps	Madagascar	Crozet(?)	88	Turonian
Ontong Java Plateau	SE Asia	Ontong-Java	122	Aptian
Rajmahal–Tasman Traps	India	Kerguelen	115	Aptian
Etendeka–Parana Traps	Brazil–Namibia	Tristan da Cunha	133	Valanginian

been effectively split into two parts: South America–Africa and India–Antarctica–Australia. Late in the Early Cretaceous (~120 Ma), a rift opened up along the southern coast of South America–Africa and propagated north to form a proto-South Atlantic Seaway. Early in the Late Cretaceous (~110 Ma), southward-propagating rifts had opened up between India and Antarctica as well as along the north-eastern coast of Australia. Seafloor spreading took place along all of these rifts throughout the Late Cretaceous (Figure 2B), causing (1) the South Atlantic Seaway to open northward where, by 100 Ma, it had joined with the Central Atlantic Seaway, (2) India to drift northward along the coast of Africa, and (3) Antarctica–Australia to drift southward. In order to compensate for the north-eastern drift of the Tethyan seafloor and the westward drift of the Atlantic seafloor, oceanic subduction systems developed along the north-eastern margin of the Tethys and along virtually the entire circumference of the Pacific Plate.

Patterns of marine circulation were also effected substantially by these tectonic reconfigurations. During the Early Cretaceous, tradewinds blowing westward across the Tethys Ocean would have set up a strong westward equatorial current that circled the globe owing to its passage between North America and the northern part of Gondwana down the Central Atlantic Seaway. Direct evidence for the existence of this current comes from the similarity of shallow marine faunas from submerged parts of the Europe and southern Asia through to (now) submerged seamounts (island systems in the Cretaceous) over 1000 miles west of Hawaii. Aside from this circumequatorial current, paired cyclonic gyres would have been present in the Pacific Ocean, leading to the evolution of distinctive northern and southern hemisphere marine invertebrate faunas. With continued continental fragmentation and northern drift through the Cretaceous, this very simple Early Cretaceous marine circulation pattern would have grown more complex, especially once the South Atlantic Seaway had opened up between South America and Africa.

In terms of physiochemical characteristics, the Cretaceous is noteworthy for the number of large igneous province eruptions (Table 1) and the number of large bolide impact craters (Table 2) that occurred during its span, as well as the large number of major marine-anoxic events (Aptian, Albian, Cenomanian, Turonian, Coniacian, Santonian, and Campanian) and very small number of magnetic polarity reversal events (the Long Cretaceous Normal interval stretches from the Aptian to the Santonian, some 40 million years). Cretaceous rates of weathering are inferred to have been relatively low, partly as the result of relatively high sea-levels and partly because of a relative lack of mountain building, especially during the Early Cretaceous. These, along with the proliferation of phytoplankton (see later), are thought to have been responsible for elevated levels of both carbon dioxide and free oxygen in the Cretaceous atmosphere, relative to modern concentrations.

Table 2 Large Cretaceous bolide impact crater

Name	*Location*	*Diameter (km)*	*Age (Ma)*	*Stage*
Gosses Bluff	Australia	22	122.5	Beriasian
Tookoonooka	Australia	55	128.0	Hauterevian
Carswell	Canada	39	115.0	Aptian
Steen River	Canada	25	91.0	Cenomanian
Kara	Russia	60	70.3	Campanian
Mansan	United States	35	73.8	Campanian
Lappajärvi	Finland	23	73.3	Campanian
Chicxulub	Mexico	170	65.5	Maastrichtian

Sea-Level and Sedimentation Patterns

As with other intervals of Phanerozoic history, the determination of Cretaceous sea-level history is necessarily tied to analyses of sedimentation patterns and is complicated by the fact that regional tectonic factors can modify, or in some cases even obscure, the global or eustatic signal. The first-order trend towards rising sea-level in the Jurassic culminated in the Tithonian and was followed by a eustatic sea-level fall of ~50 m into the Berriasian (Figure 3) The nadir of Late Hauterivian sea-level represents the lowest sea-level documented for the entire system. Major sea-level rises took place in the Early Barremian, Albian, and Early

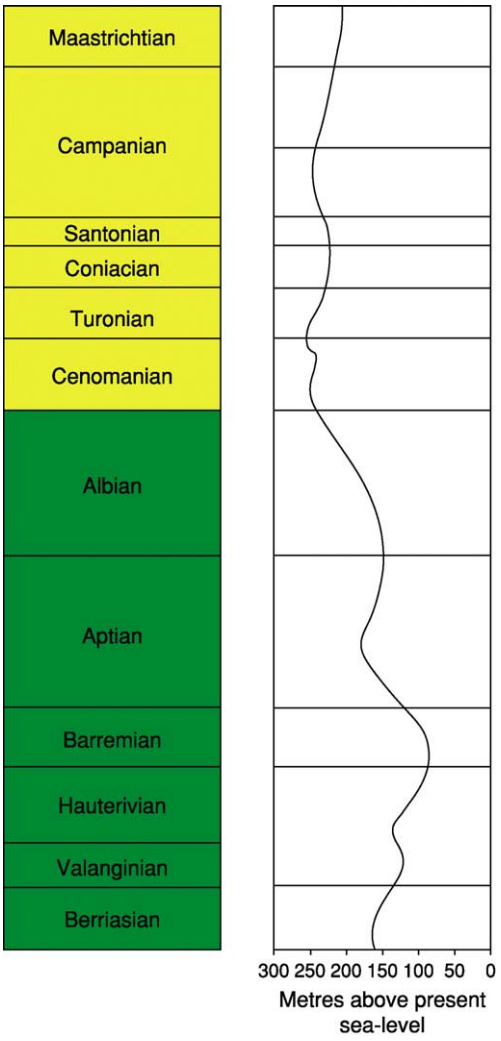


Figure 3 Global variation in sea-level throughout the Cretaceous. Sea-level curve partly based on data from Haq.

Campanian, with the Turonian transgression achieving the highest global sea-level stand, some 250 m above present sea-level, not only for the Cretaceous System, but also for the entire Mesozoic–Cenozoic interval. Global sea-level remained high throughout the Upper Cretaceous, but suffered quite a sudden (tens of thousands of years) and deep (50–100 m) sea-level fall more or less coincident with the Cretaceous–Tertiary (K–T) boundary. Due to the lack of evidence for widespread glaciation in the Cretaceous rocks of Gondwana, it is presumed that these sea-level fluctuations were largely driven by changes in the heat flow surrounding mid-ocean ridge systems, causing the ridge systems to swell and contract, with consequent effects on the volume of the deep-ocean basins.

The very high sea-levels achieved throughout the Late Cretaceous meant that large portions of the continental platforms were flooded to form very broad, but shallow, epicontinental seas (Figure 2). The substantial sediments that accumulated in these seas are largely responsible for the excellent Cretaceous geological and biological record. In North America and South America, collisions between the western margins of those plates and the eastern Pacific subduction centres resulted in the fold–thrust uplift of the Cordilleran mountain ranges (e.g., Sevier Orogeny) (*see* **North America:** Northern Cordillera; Southern Cordillera), along with associated volcanic and plutonic activity. Sediments from these mountains were shed to the east and west. In North America, this erosion led to deposition of the predominately clastic Great Valley sequences in California, which were subsequently deformed and uplifted during the Cretaceous accretion of microcontinental fragments (e.g., Wrangalia).

To the east of the Sevier–Laramide mountains, a large epicontinental sea (the Mowry Sea) encroached from the north and south as a result of the late Early Cretaceous sea-level rise. During its initial transgression (Albian), this sea was characterized by dysaerobic to anoxic conditions as evidenced by the abundant oil shales and black shales of this age. Clastic deposition characterized the northern part of the Mowry Sea during the Early Cretaceous, whereas carbonate-evaporite deposition characterized its southern arm. These two arms coalesced in the early Late Cretaceous (during the sea-level maximum), and a single interior seaway occupied the central portion of North America through to the Maastrichtian, during which time a more typical basinal carbonate-clastic depositional pattern become dominant. This same pattern of Early Cretaceous drowning of continental platforms also took place in South America, Europe, southern Asia, and Australia.

During the Upper Cretaceous, the characteristic chalk lithofacies developed in most large, epicontinental, marine ocean basins. These enormous chalk seas represented a singular environment that had no equivalent prior to the Late Cretaceous nor in all but the earliest part of the subsequent Cenozoic. Chalk is predominately an epipelagic sedimentary deposit composed of astronomical numbers of calcareous microfossil skeletons, chiefly nannoplankton and planktonic foraminifera. These organisms are present in the world's oceans today, but large areas of modern chalk deposits are not being created because the steady rain of calcareous from the water column is diluted by clastic sediments and by the dissolution of calcareous materials in deeper water. The shallow Late Cretaceous epicontinental seas, however, combined shallow depths with high productivity (because of their chemistry; see later) and low clastic input (because of their size) to produce near-ideal conditions for the development and preservation of plankton tests.

Moreover, these planktonic groups served as the basis for a highly productive and stable Late Cretaceous marine food chain, fostering the diversification of both marine invertebrate and vertebrate groups (see later). With the Late Maastrichtian sea-level regression, though, these chalk seas retreated from the continental interiors, although a few regional centres of chalk deposition continued into the Paleocene.

Climate

Continuing on from the Jurassic greenhouse, world climates in the Cretaceous were, if anything, even more stable, uniform, and equable, even at very high latitudes (Figure 4A). During the Early Cretaceous, climate bands with essentially modern latitude limits were found in both the northern and the southern hemispheres. Exceptions include a tropical zone largely confined to the western Tethys (as evidenced by widespread reef facies), a paratropical embayment reaching across southern Europe and into Eurasia (as evidenced by high-latitude bauxite deposits), and a very large arid region (as evidenced by widespread evaporite and calcrete deposits) in southern North America and northern South America. The overall equability of the Early Cretaceous climate is supported by the observation of coal deposits throughout Pangaea and Laurasia – even in areas reconstructed as being near the Early Cretaceous poles – along with crocodile fossils above 30° N latitude and, amazingly, above 60° S latitude!

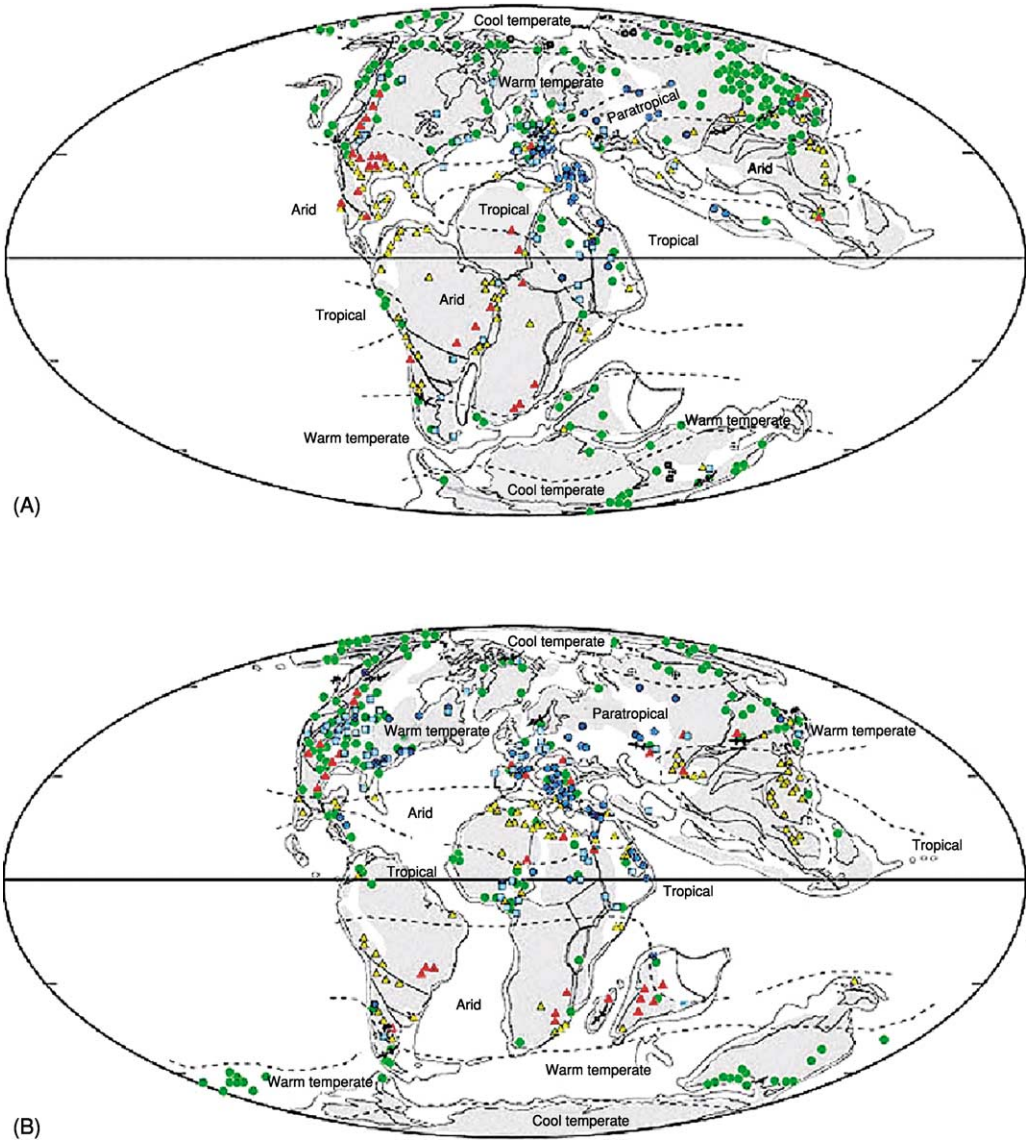


Figure 4 Environmental zones and climate variation (as inferred from sediments and biota) at (A) 120 Ma (Aptian) and (B) 80 Ma (Campanian). Palaeogeographical maps and environmental data generated as part of the GeoMap Project. Reprinted with permission.

The Late Cretaceous sea-level rise further intensified these climatic gradients (Figure 4B) by creating arid belts both north and south of the tropical equatorial zone and pushing the paratropical embayment well into Asia. However, an asymmetry developed between the north and south polar areas in the Late Cretaceous: the northern polar region contains coal deposits, crocodile fossils, and even dinosaur fossils well above the Arctic Circle, whereas these deposits stop at or near 60° S latitude. No doubt the fact that Antarctica had drifted to occupy a position at the Late Cretaceous south pole was a significant factor in the development of this climatic contrast between the hemispheres.

In the sea, $\delta^{18}O$ analyses of planktonic foraminifera indicate that both surface- and bottom-water temperatures rose steadily through the Lower Cretaceous, with a single, strong, high-temperature anomaly in the Upper Berriasian–Lower Aptian. Cretaceous marine temperatures peaked in the Turonian (average surface temperatures of ~18°C) and then went into a rapid decline to the Campanian. The Campanian–Maastrichtian interval was characterized by strong marine surface-water temperature instabilities, with several strong reversals between warm and cool phases that varied by as much as 4°C continuing to the K–T boundary. In contrast, marine bottom-water temperatures do not appear to have been subject to these strong variations.

Biota

Fossil Protists

The Cretaceous is, in many ways, the acme of the microfossil record. Two of the three modern phytoplankton groups, calcareous nannoplankton (coccoliths) and diatoms (Figure 5), underwent major family-level diversifications during this period, with diatoms being a particularly rapid family-level diversification in the

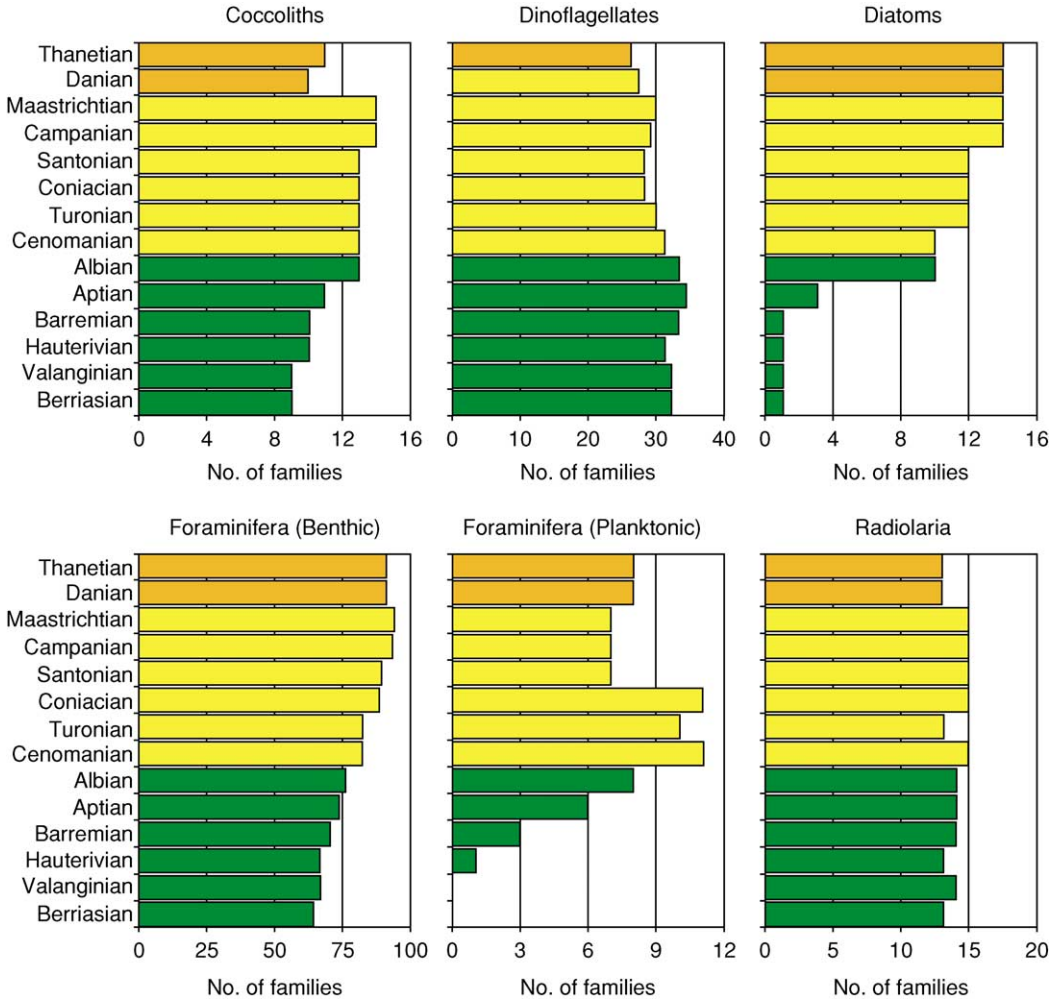


Figure 5 Fossil protist family-richness patterns through the Cretaceous and the first two stages of the Paleocene. Note different scales on each graph. Based on data from Benton MJ (1993) *The Fossil Record 2*. London: Chapman & Hall.

late Lower Cretaceous. Dinoflagellates maintained their very high diversities throughout the interval, albeit with a long-term drift to slightly lower family-richness values through the Late Cretaceous, and this continued into the Paleocene. Foraminifera responded to this change in the marine environment by maintaining a steady diversification of benthic forms through the late Cretaceous (probably a response to enhanced carbon deposition to the seafloor by planktivorous zooplankton and nekton). Planktonic foraminifera first appeared in the Hauterivian and, after an initial radiation that extended into the early part of the Upper Cretaceous (e.g., appearance of the first hedbergellids, heterohelicids, and guembelitrids), appear to have settled back to a steady-state family-level diversity in the upper part of the Late Cretaceous. Radiolaria appear to have maintained a more or less steady-state family-level diversity throughout the interval.

As a result of these radiations, the Cretaceous seabeds were, for the first time, blanketed with calcareous oozes. Of course, the most dramatic example of this was the extensive and economically important chalk deposits of North America and Europe. Massive Cretaceous chalk production appears to have occurred because of the very low Mg/Ca ratio of Cretaceous seawater. This chemical environment favours the production of calcite, which is then preserved because the small size of the nannoplankton-produced grains makes the (later uplifted) chalk deposit nearly impenetrable to groundwater. Widespread chalk deposition ended in the Middle Paleocene when the seawater Mg/Ca ratio began to rise from its all-Phanerozoic Cretaceous low.

Marine Invertebrates

Like most marine protist groups, corals, marine molluscs (chiefly gastropods, cephalopods, and bivalves), and marine arthropods (chiefly crustaceans and ostracods) underwent long-term family-level radiations through the Cretaceous and extending into the overlying Paleocene (Figure 6). Corals diversified strongly throughout the interval, but declined in abundance in the Late Cretaceous, presumably because the low Mg/Ca ratio made it more difficult to secrete their aragonite skeletons. The total mollusc patterns shown in Figure 6 mask strongly differing patterns among the

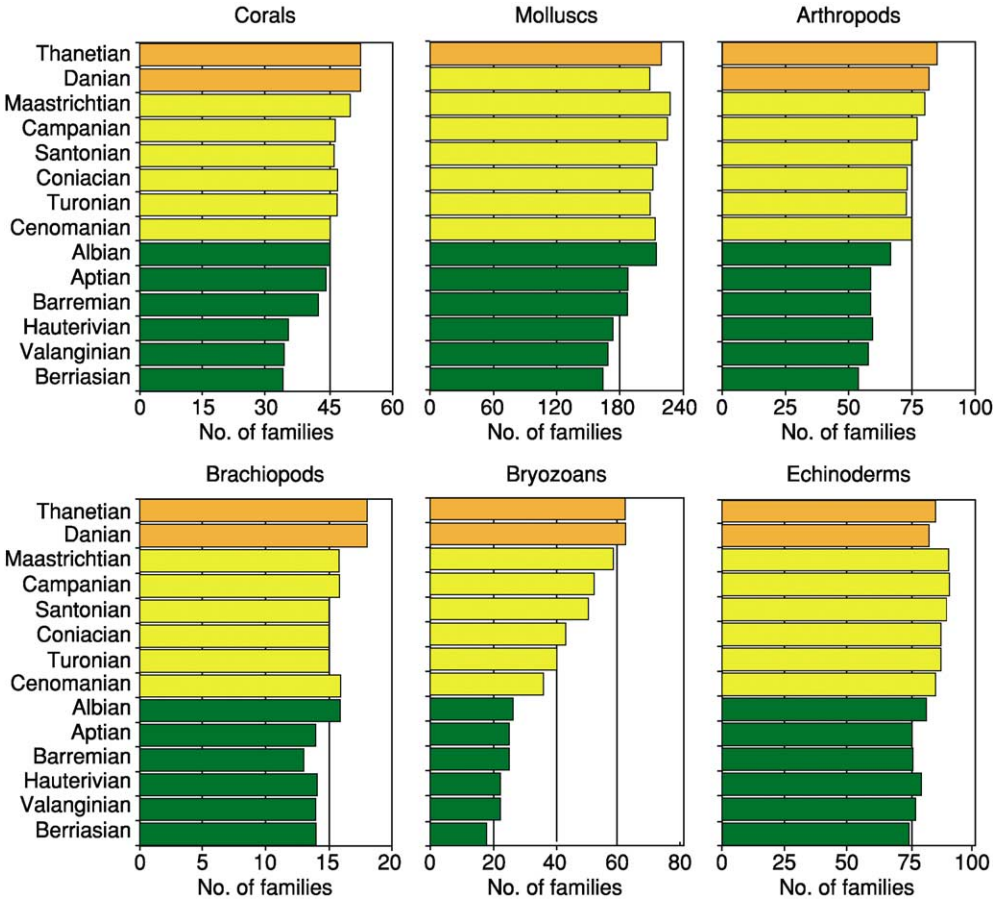


Figure 6 Fossil marine (invertebrate) family-richness patterns through the Cretaceous and the first two stages of the Paleocene. Note different scales on each graph. Based on data from Benton MJ (1993) *The Fossil Record 2*. London: Chapman & Hall.

constituent groups. For example, cephalopods exhibited a particularly impressive diversification in the latter half of the Early Cretaceous and across into the earlier part of the Late Cretaceous, whereas gastropods and bivalves had the greater part of their family-level diversification histories in the Late Cretaceous. Burrowing bivalves and gastropods of modern aspect (neogastropods) diversified strongly throughout the period. The Cretaceous was also a time of gigantism among several surface-dwelling bivalve lineages; the genus *Inoceramus*, for example, could be as much as a metre across (natural inoceramid casts are often mistaken by amateur geologists for 'dinosaur footprints'). One oddly shaped bivalve group, the rudists, even managed to replace corals as the principal reef builders of the shallow Late Cretaceous seas, owing to their superior chemical control of skeleton-secretion processes. The Cretaceous arthropod radiation was driven largely by crustaceans, presumably in response to the diversification of neogastropods, their principal prey item. During this time, gastropods and crabs continued their 'arms race', with gastropods developing ever more elaborate predator deterrent mechanisms (e.g., reinforced apertures, shell ribbing, and spines) and crabs responding through the development of improved claw designs (e.g., strength and shape). This crustacean diversification was particularly pronounced in the Albian and Campanian. Brachiopod family-richness values fell slightly in the Barremian, but recovered quickly and were maintained throughout the Late Cretaceous. Bryozoans exhibited perhaps the most striking richness-change pattern, with a strong and sustained diversification throughout the entire Late Cretaceous. This trend was driven by cheilostome bryozoans, which almost trebled their number of families through the course of the Late Cretaceous, though ctenostomes and cyclostomes also underwent modest diversifications through this interval as well. Echinoderms appear to have had the most conservative diversification history of any major marine invertebrate clade, with the slight drift to higher Late Cretaceous values being driven primarily by an asteroid (starfish) radiation.

Marine Vertebrates

The Cretaceous marine vertebrate faunas exhibit surprising similarities (and differences) between the richness histories of 'fish' (including sharks, rays, and bony fish) and marine 'reptiles' (including marine turtles, ichthyosaurs, plesiosaurs, and mosasaurs). Both groups (Figure 7) exhibit low, steady-state Early Cretaceous values and both exhibit much higher Late Cretaceous values. Moreover, both groups underwent the transition from low Early Cretaceous values to higher Late Cretaceous values, within a stage of one another. This similarity is all the more remarkable when the difference between the sizes of these faunas is taken into consideration. The Late Cretaceous richness increase was primarily driven by diversification within fish clades, whereas the driver of the marine 'reptile' pattern was chiefly a turtle–mosasaur diversification.

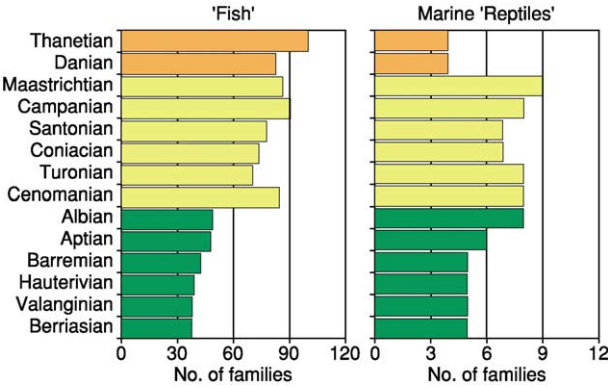


Figure 7 Fossil marine (vertebrate) family-richness patterns through the Cretaceous and the first two stages of the Paleocene. Note different scales on each graph. Based on data from Benton MJ (1993) *The Fossil Record 2*. London: Chapman & Hall.

Terrestrial Invertebrates

There were several groups of terrestrial and freshwater molluscs (e.g., pulmonate gastropods and unionid bivalves) as well as freshwater arthropods (e.g., crustacea and ostracodes), in addition to the ubiquitous Cretaceous insects. Among these, only the molluscs exhibit a sustained diversification trend (Figure 8). This pattern was driven primarily by the Late Cretaceous proliferation of both terrestrial and freshwater gastropods. Both non-insect arthropods and insects exhibited more or less steady-state diversification histories throughout this interval, with perhaps some suggestion of a Late Cretaceous increase in the former, driven primarily by a proliferation of Late Cretaceous chelicerate families (e.g., scorpions). It is interesting to note the very high richness values for Cretaceous insects, a group many would assume to have a 'poor' fossil record.

Terrestrial Vertebrates

Though dinosaurs understandably get most of the attention when it comes to Cretaceous terrestrial vertebrate faunas, it is important to note that essentially all modern terrestrial vertebrate groups had representatives in the Cretaceous. Although not as well known to the popular audience, salamanders, frogs, lizards, snakes, turtles, crocodiles, birds,

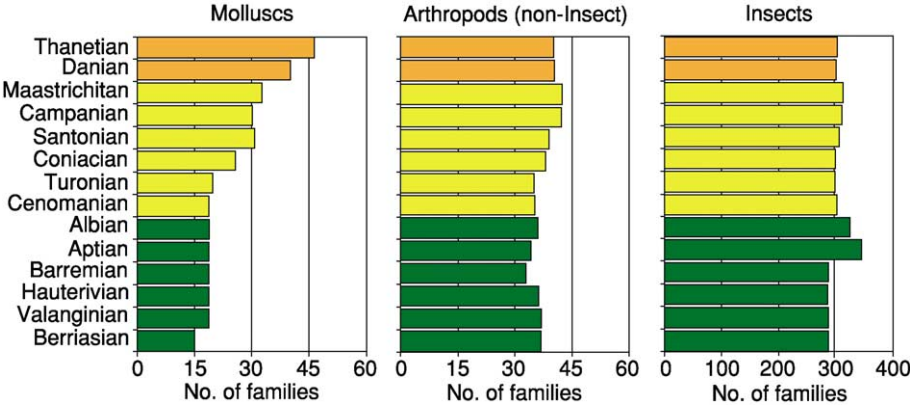


Figure 8 Fossil terrestrial (invertebrate) family-richness patterns through the Cretaceous and the first two stages of the Paleocene. Note different scales on each graph. Based on data from Benton MJ (1993) *The Fossil Record 2*. London: Chapman & Hall.

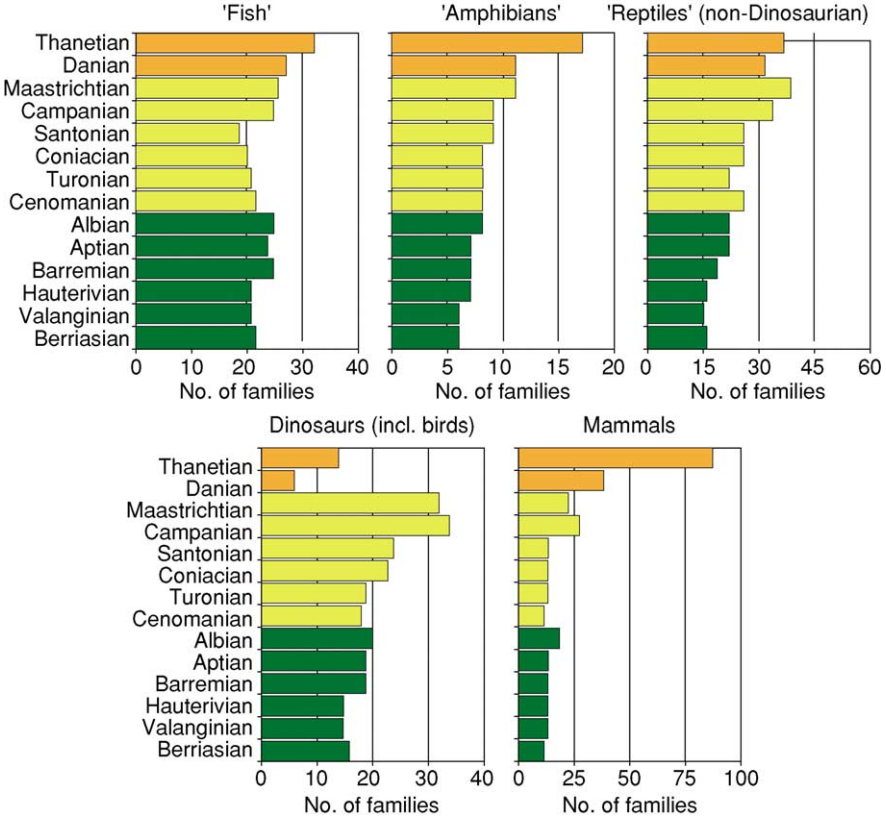


Figure 9 Fossil terrestrial (vertebrate) family-richness patterns through the Cretaceous and the first two stages of the Paleocene. Note different scales on each graph. Based on data from Benton MJ (1993) *The Fossil Record 2*. London: Chapman & Hall.

and mammals all shared the Cretaceous landscapes along with the non-avian dinosaurs, non-dinosaur archosaurs (e.g., crocodiles and champosaurs), and pterosaurs (Figure 9). Perhaps the most unusual Cretaceous diversification history is that of freshwater 'fish' (including sharks and rays as well as bony fish). This group began the Cretaceous with a healthy 20 or so families and underwent a modest diversification event in the latter part of the Early Cretaceous. The subsequent early part of the Late Cretaceous, when many other groups were diversifying, was a time of progressive diversity reduction in the freshwater fish fauna. This pattern was driven largely by reductions in the number of teleost and sarcopterygian clades. However, this declining pattern reversed dramatically in the Campanian (driven by a Late Cretaceous

radiation in teleost clades) and continued unbroken across the K–T boundary and into the Paleocene.

Both amphibians and non-dinosaurian reptiles (e.g., turtles, snakes, and lizards) exhibited similar progressive patterns throughout the Cretaceous that also extended into the Paleocene. Dinosaurs had strong diversification throughout the Late Cretaceous up to their Campanian peak, from which there was a marked retreat (involving both theropod and sauropodomorph forms) in the Maastrichtian. All non-avian dinosaurs became extinct by the end of the Maastrichtian, and the Danian values in Figure 9 all represent bird families. This said, there are persistent reports of dinosaur remains and signs (e.g., nests and trackways) in Danian strata and it would not be surprising if these reports were confirmed and accepted by most vertebrate palaeontologists in the near future.

Last, but by no means least, mammals exhibit very low family-richness values throughout most of the Early Cretaceous and the lower part of the Late Cretaceous. There is a hint, however, that mammalian diversification started in the Campanian, at about the same time as the final dinosaur diversification. Mammals also underwent a richness reduction in the Maastrichtian coincident with a similar reduction in dinosaur family-richness values. Indeed, mammals lost more families in the Maastrichtian than did dinosaurs! The fates of these two clades were decoupled by other events in the Maastrichtian, but it is false to regard their diversity histories as being mirror images of one another, as has been the case in many popular accounts and more than a few scientific treatises. Rather than ships passing in the night, the Cretaceous history of mammals and dinosaurs suggests several striking similarities, the interpretation of which would benefit from further investigation.

Terrestrial Plants

The Cretaceous bore witness to a fundamental transformation of the terrestrial flora, with the initial diversification of seed-bearing (angiosperm) plants. Though angiosperms first appeared in the Jurassic, 'naked seed' gymnosperms such as conifers, cycads, and ginkgoes dominated the earliest Cretaceous landscapes. Nevertheless, the strong plant diversification event documented in Figure 10 is almost entirely the result of family production within angiosperms. During the Cretaceous, over 30 new angiosperm families appeared, including beech, birch, fig, holly, magnolia, oak, palm, sycamore, and walnut. Angiosperm seed development is considered superior to that of gymnosperms because the seed is protected from damage inside the plant's ovary. This means that the plant can produce a much larger number of smaller seeds from the same energy investment required to make the less efficient gymnosperm cone. Flowers were another Cretaceous angiosperm development, and these literally opened up a host of new possibilities for the attraction of pollinator organisms, effectively co-opting other creatures into the plants' reproductive strategy. Like the first plants, the first angiosperms appear in Early Cretaceous lake and stream habitats, where quick-developing, opportunistic, weedy forms have an advantage. However, there were no grasses, and so no grassy plain or savanna habitats, in Cretaceous times.

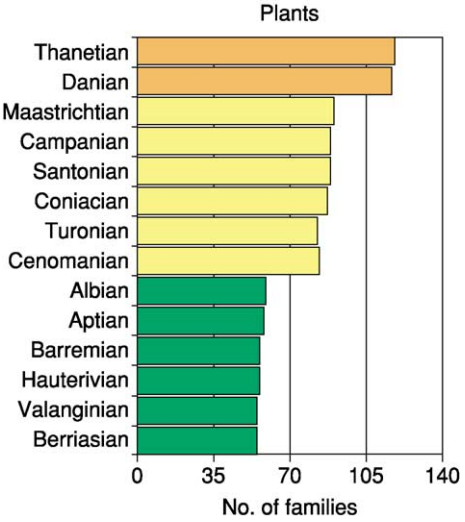


Figure 10 Fossil terrestrial (plant) family-richness patterns through the Cretaceous and the first two stages of the Paleocene. Based on data from Benton MJ (1993) *The Fossil Record 2*. London: Chapman & Hall.

Extinctions

Two extinction events are widely held to have occurred during the Cretaceous, an Aptian event and a Cenomanian event. There is considerable debate on whether the Aptian event is important, or even recognizable in most palaeontological datasets. In the late JJ (Jack) Sepkoski's family-level and genus-level compendia, an Aptian extinction-intensity peak stands slightly above what he and David Raup regarded as background extinction levels. The extent of this peak's distinctness has changed over successive versions of the Sepkoski dataset and is, to some extent, dependent on how time is 'binned' across the relatively long Aptian and Albian stages. In addition, an extinction peak in the vicinity of 119 Ma was predicted on the basis of Raup and Sepkoski's extinction-periodicity hypothesis.

Irrespective of these analytic results, specialist biostratigraphers and systematists have been hard

pressed to identify which Aptian organismal groups bore the brunt of this putative extinction event. Modest losses among planktonic foraminifera, benthic foraminifera, calcareous algae, and rudistid bivalves have been noted, but none of these appears to have had a dominant effect on the diversity histories of any of these groups, not to mention the biosphere as a whole. Possible extinction mechanisms also abound in the Aptian, including a local sea-level lowstand, an ocean-anoxic event, and a large volcanic eruption that emplaced the submarine Ontong-Java Plateau. Nevertheless, the Aptian extinction record appears to have a relatively weak claim to be a major turning point in Earth's biodiversity history.

The case for a major extinction event at the Cenomanian–Turonian boundary is much stronger. Beginning in the Upper Cenomanian, and continuing into the Lower Turonian, marine taxa suffered the lost of about 7% of all families with fossilizable hard parts and 26% of all constituent genera. This event occurred over an interval of from 3 to 5 Ma and was complexly structured both temporally and geographically. Ecologically, the extinction appears to exhibit a distinctly 'bottom-up' character, with deep-dwelling taxa (e.g., gastropods, bivalves; and benthic foraminifera, especially agglutinated-walled forms; deep-dwelling planktonic foraminifera; and bottom-feeding ammonoids) being differentially affected. Over a few metres of section, species-level extinction rates ranging from 20% to 50% (and occasionally as high as 80%) are commonly reported. At least locally, however, taxa in shallower water (including coccoliths, pelagic ammonites, and rudistid bivalves, as well as sharks, ichthyosaurs, and plesiosaurs) fell victim. Geographically, the extinction appears to have been centred on tropical and temperate biota, with virtually no extinctions being recorded above 60° north or south of the Cretaceous palaeoequator.

Geochemically, the extinction event is correlated with the widespread interruption of chalk formation in epicontinental seaways in favour of clastic deposition, especially in the form of the black shales and limestones that signal low-oxygen or dysaerobic conditions. In addition, this event is associated with pronounced $\delta^{13}C$ and $\delta^{18}O$ anomalies, a sea-level highstand, and an interval of widespread submarine and subareal volcanism. No convincing or consistent evidence for bolide impact has been recovered from Cenomanian–Turonian boundary sediments to date. Two Ir anomalies are associated with the Cenomanian–Turonian interval, but, owing to similarities between the trace-element signature of these anomalies and the trace-element composition of mid-Atlantic Ridge basalts, these anomalies have been interpreted as resulting from volcanic, rather than extra-terrestrial, inputs.

Virtually all of the popular extinction mechanisms, including bolide impact, have been advanced to explain the Cenomanian–Turonian extinction. The current geological and palaeontological consensus suggests that this is a multicausal event that occurred because a unique juxtaposition of independent factors affecting marine habitats coincided during this interval of Earth history. The most important proximal cause was probably anoxic deep-ocean waters invading middle and, in some cases, shallow shelf habitats as a result of eustatic sea-level rise. The Cenomanian–Turonian interval represents the highest stand of sea-level for the entire Phanerozoic as well as the highest Cretaceous sea-level stand. This primary mechanism was likely intensified by submarine volcanism (which can alter the buoyancy of dysoxic and anoxic deep-marine waters, causing them to rise further up the continental shelves than would have otherwise been the case) and global cooling resulting from improved marine circulation patterns that were probably a by-product of the sea-level highstand. This improved circulation may have increased the efficiency of heat transfer from tropical to polar regions, thus cooling the topics and (perhaps) exceeding the tolerance of many warm-adapted tropical species, including reef-building rudists and corals. As is the case with most modern extinction events, once a sufficient number of key species had been eliminated, other species would become increasingly susceptible to extinction because of complex ecological dependencies rather than physiochemical tolerances per se.

It has been suggested that the magnitude of the Cenomanian–Turonian extinction event has been overestimated as a result of the geographic migration of certain types of marine environments in the stratigraphic record, and because any single section or core presents a picture of changing local conditions superimposed over changing global conditions. Though all fossil records must be examined by biases of this sort, the widespread and, on the whole, consistent nature of the biotic patterns described from the Cenomanian–Turonian interval argues that such biases – though undoubtedly present – can be invoked to discount the entire event. Additional research will be needed in order to test these hypotheses and further sharpen our understanding of this fascinating interval of Earth history.

See Also

Atmosphere Evolution. **Biozones**. **Fossil Invertebrates:** Arthropods; Brachiopods; Bryozoans; Corals and Other Cnidaria; Echinoderms (Other Than Echinoids); Molluscs Overview. **Fossil Plants:** Calcareous

Algae. **Fossil Vertebrates:** Fish; Palaeozoic Non-Amniote Tetrapods. **Mesozoic:** End Cretaceous Extinctions. **Microfossils:** Foraminifera. **Plate Tectonics**. **Sequence Stratigraphy**. **Stratigraphical Principles**. **Time Scale**.

Further Reading

Benton MJ (1993) *The Fossil Record 2*. London: Chapman & Hall.

Gould SJ (ed.) (1993) *The Book of Life*. New York: WW Norton.

Hallam A (1992) *Phanerozoic Sea-Level Changes*. New York: Columbia University Press.

Hallam A and Wignall PB (1997) *Mass Extinctions and Their Aftermath*. Oxford: Oxford Science Publications.

Haq B, Hardenbol J, and Vail PR (1987) Chronology and fluctuating sea levels since the Triassic. *Science* 235: 1156–1166.

Stanley S (1989) *Earth and Life through Time*, 2nd edn. San Francisco: WH Freeman.

Stanley SM (1999) *Earth System History*. New York: WH Freeman.

End Cretaceous Extinctions

N MacLeod, The Natural History Museum, London, UK

Introduction

Although the end-Cretaceous extinction event goes by a variety of names in both the technical and popular literature (e.g., Cretaceous–Tertiary (K–T) mass extinction, K–T boundary extinction, Cretaceous–Paleocene (K–P) mass extinction, Cretaceous–Palaeogene (K–P) extinction), it is most closely associated with the uppermost Cretaceous period/stage—the Maastrichtian—and, to a lesser extent, the lowermost Palaeogene epoch/age—the Danian (**Figure 1**). As such, any review of effects and causes that may (or may not) have occurred during this episode of Earth history must begin with a review of the stratigraphy of these two intervals, not only because they provide the temporal, environmental, geographic, and tectonic context within which such phenomena must be understood, but also because the study cause–effect associations in Earth history is, by definition, a largely stratigraphic exercise.

Dumont first defined the Maastrichtian in 1850 as a distinct stratigraphic subdivision of the Upper Cretaceous well expressed in the sediments around the town of Maastricht in southern Holland. The current boundary stratotype was established in 1911 by the Comité d'étude du Maastrichtian as the Tuffeau section, exposed in the ENCI quarry at St. Pietersburg on the Maastricht outskirts. Unfortunately, only the upper Maastrichtian is exposed in this quarry and its upper boundary has traditionally been regarded as incomplete (though a 1996 restudy of the quarry concluded that a complete K–T boundary succession was present in certain man-made cave sections within the quarry). In terms of boundary stratotypes the Campanian–Maastrichtian boundary was established by the International Commission on Stratigraphy (ICS) in 2001 at the 115.2-m level in Grande Carrière Quarry, Tercis-les-Bains, Landes Province, in south-western France. A total of 12 criteria (all of equal weight) were defined as useful in recognizing the Campanian–Maastrichtian boundary, which falls just above the first appearance of the ammonite *Pachydiscus neubergicus* in this section.

The overlying Danian was established by Desor in 1847 for the stratigraphic successions present at Stevens Klint and Faxse in Denmark. Desor originally regarded the Danian as a Cretaceous stage because it is characterized in these two localities by chalk lithofacies. Later, it was shown that these sections were equivalent temporally to the Montian stage of Belgium that had long been regarded by continental European stratigraphers as the lowermost stage of the Tertiary. After a short debate concerning whether to place the new Danian–Montian stage in the Cretaceous or the Tertiary, the latter was accepted, largely on the basis of similarities between this fauna and that of the Midway Formation of the US Gulf Coast, which had also long been regarded as being basal Tertiary. These correlations notwithstanding, controversy regarding the correct placement of the Danian continues, largely on the basis that a number of characteristically 'Cretaceous' taxa (including bryozoans, brachiopods, echinoids, gastropods, bivalves, and, perhaps, planktonic foraminifera) did not become extinct until the end of the Danian where chalk deposition also effectively ceased worldwide. Some have regarded the Danian as the time interval between the (Maastrichtian) 'white chalk' exposed at Stvens Klint and the basal Selandian conglomerate exposed at Hvalløse in Jutland.

The Danian boundary stratotype was established by the GSSP in 1996 at the base of the boundary clay in the El Haria section, near the town of El Kef,

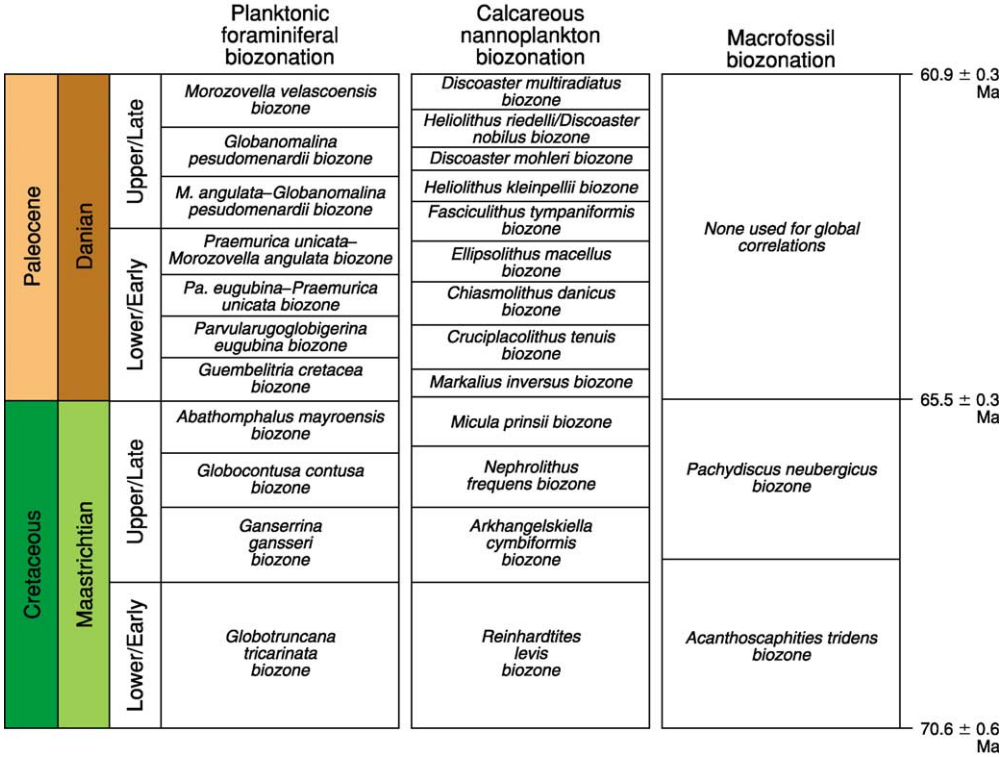


Figure 1 Chronostratigraphy, biostratigraphy, and geochronology of the (Cretaceous) Maastrichtian through (Palaeogene, Paleocene) Danian interval. Boundaries between Early/Lower and Late/Upper Maastrichtian and Danian represent undated, informal units based on biostratigraphic zones.

Tunisia. Like the Campanian–Maastrichtian boundary, multiple criteria are used to recognize the Maastrichtian–Danian boundary (which is also the Cretaceous–Tertiary, Cretaceous–Palaeogene, and Mesozoic–Cenozoic boundary), including an iridium (Ir) anomaly and a major extinction horizon for foraminifera and calcareous nanoplankton. The boundary stratotype for the base of the Selandian Epoch (= top of the Danian) has not been established to date.

Like almost all chronostratigraphic boundaries, the Maastrichtian–Danian interval is difficult to correlate because:

1. most so-called boundary successions are profoundly incomplete, often juxtaposing sediments of markedly different ages (e.g., Campanian beneath Eocene), and
2. biozones (see Figure 1) contained within the interval are demonstrably diachronous and often facies limited.

For example, the traditional uppermost Maastrichtian planktonic foraminiferal zone—the *Abathomphalus mayroensis* taxon range zone—is based on a open ocean, deep-dwelling species not present in shallower marine environments, including those represented by the El Kef boundary stratotype. The result has been confusion regarding what constitutes a complete record of the uppermost Maastrichtian–lowermost Danian sequences of events, especially between different ecological realms and locations remote from one another.

The Nature of the Maastrichtian–Danian Turnover Patterns

Historical Concept

Despite the problems associated with attempts to achieve a high-resolution, temporal correlation of end-Cretaceous and Early Danian sediments, it is abundantly clear that something extraordinary happened to the Earth's biosphere in the Late Cretaceous. In many lineages characteristic elements of the Late Cretaceous marine and terrestrial biota are simply not present in the Early Paleocene fossil record. The standard list of Late Cretaceous victims includes coccoliths and planktonic foraminifera (both of which survive into the Danian but with an almost complete species-level taxonomic turnover), rudistid and inoceramid bivalves (both of which were common constituents of the Late Cretaceous chalk seas and both of which became entirely extinct),

ammonite cephalopods (the traditional index fossil group of Cretaceous biostratigraphy, all of whose representatives became entirely extinct), marine reptiles (e.g., ichthyosaurs, plesiosaurs, mosasaurs; all entirely extinct), flying reptiles (=pterosaurs, extinct), and all nonavian dinosaurs (extinct). (Modern birds are nested phylogenetically within the dinosaur clade and so are considered dinosaurs in terms of their ancestry. Recent discoveries of feathered Cretaceous dinosaurs in China confirms this phylogenetic link and puts the 'obvious morphological distinctions' between extinct nonavian dinosaurs and modern birds—which are often cited by those who would try to preserve phylogenetically unwarranted taxonomic distinctions between these two groups—in a somewhat different light.) On one hand, this might not seem a very long list, especially given the fact that these groups comprise a relatively small collection of suborders, superfamilies, families, and genera in a much larger Late Cretaceous biota. Nevertheless, several victim groups were among the set of dominant marine and terrestrial players on the Late Cretaceous ecological stage whose demise affected many other groups in both direct and indirect ways. Also, since many evolutionary advantages flow from ecological incumbency, the removal of these groups from the ecological scene (by whatever means) opened up opportunities for evolutionary–ecological transformation and diversification among Late Cretaceous survivor lineages that simply would not have been possible otherwise. (This dependency on the unpredictable elimination of competition between lineages is referred to as contingency in the evolutionary literature, and it is widely reported that, were it not for the contingent elimination of nonavian dinosaurs in the Late Cretaceous, the explosive radiation of mammals—in which our own lineage, primates, took part—would not have taken place. Although this is true in a general sense, it is not the case that we owe our existence solely to the end-Cretaceous extinction since many other antecedent as well as subsequent contingencies also played important roles in mammal-primate evolution.)

In order to understand the character of this extinction event, the obvious questions to ask are as follows:

1. When did these extinctions occur?
2. Did they affect only these traditional victim groups?
3. Did any groups diversify or prove unusually extinction-resistant over this interval?
4. Were these Late Cretaceous extinctions associated with any physical event(s) that could have been responsible for the extinctions?

Before answers to these questions can be attempted, however, some further aspects of the Late Cretaceous–Early Tertiary fossil record must be understood.

Sources of Bias in the Fossil Record

Although all information we possess about fossils is traceable ultimately to observations made on actual specimens, many factors conspire to constrain the scope and depth of this knowledge for different fossil groups. For example, the contrast between the Late Cretaceous planktonic foraminiferal and dinosaur records could hardly be more striking even though both have played roughly equivalent roles in the understanding of end-Cretaceous extinction patterns and processes. Planktonic foraminifera are very common constituents of Maastrichtian and Danian marine sediments deposited over a broad range of marine depth habitats. Species can be recovered from these sediments worldwide, sampled in great temporal detail (owing, not least, to the small size of foraminiferal shells), and placed within a very highly resolved biostratigraphy–taxonomy, albeit the latter of whose phylogenetic dimensions are not especially well known. Dinosaurs, on the other hand, are comparatively rare constituents of Maastrichtian terrestrial sediments and occur in only a narrow range of terrestrial habitats. Uppermost Maastrichtian dinosaurs are known to occur in only one area of the world (western USA) and cannot be sampled in great temporal detail, because of the spotty occurrence of complete or near-complete skeletons. Dinosaur biostratigraphy and taxonomy are not as well known as planktonic foraminiferal taxonomy (e.g., it is difficult to assess the shifting patterns of erosion-prone terrestrial sections for chronostratigraphic relations to stratotype successions, since new dinosaur species turn up each year). Nevertheless, owing to the complex nature of vertebrate skeletons, more is known about dinosaur phylogeny than planktonic foraminiferal phylogeny. The Late Maastrichtian–Danian fossil records of other groups can be thought of as ranging through a spectrum of temporal, geographic, taxonomic, and phylogenetic resolutions, the boundaries of which are set by dinosaurs and planktonic foraminifera.

Not only are there intrinsic differences between the quality of the fossil records of different groups, there are also several fundamental sources of uncertainty about the details of all fossil records, which constrain interpretations based on palaeontological observations. First, lack of a detailed, global chronostratigraphy for Maastrichtian–Danian sediments means that the vast majority of fossils can only be located as occurring—or not occurring—within these relatively coarse temporal intervals. Thus, although it is

possible to say that the extinction of plagioptychid bivalves and azhdarchid pterosaurs both occurred in the Maastrichtian, it is not possible to place these extinctions accurately within the span of the Maastrichtian, or to say whether they occurred simultaneously with respect to one another, without going back and restudying the original material. In particular, it is not appropriate to assume that simply because an extinction event is listed as Maastrichtian, all species composing the group in question ranged through the entire Maastrichtian and then simultaneously became extinct at a horizon coincident with the Maastrichtian–Danian boundary.

It is standard practice for biostratigraphic range charts to represent the chronostratigraphic ranges of fossil groups as solid lines joining individual occurrence horizons that denote intervals of time along a temporal axis (Figure 2). Because we experience time as a continuum it is tempting to regard these axes as representing time as a continuous variable. In fact, these charts represent time as a discontinuous variable with the vertical range line always being drawn through the entire interval, irrespective of whether the actual time of extinction is known. This graphing convention often gives the (erroneous) impression that all extinction events occur at stage/age boundaries and that there is some pronounced tendency for extinctions to occur together in time.

Similarly, use of higher taxonomic categorizations (e.g., families, genera) as proxies for species in extinction studies often leads to unappreciated distortions of the fossil record by inexperienced interpreters. The taxonomic categories of family and genus are used most often in extinction studies because these are regarded as more stable and comparable than species-level data across the broad scope of life's diversity. Because these are composite categories, however, the presence of a family at one point in time may represent a rather large number of species, whereas at another point in time (especially if the latter is close to the group's extinction event) the actual number of species represented may be much smaller (Figure 2).

Pseudoextinction is another problem. In an evolving lineage two types of morphological transformations can occur. The first (anagenesis) results in the progressive transformation of the entire species from

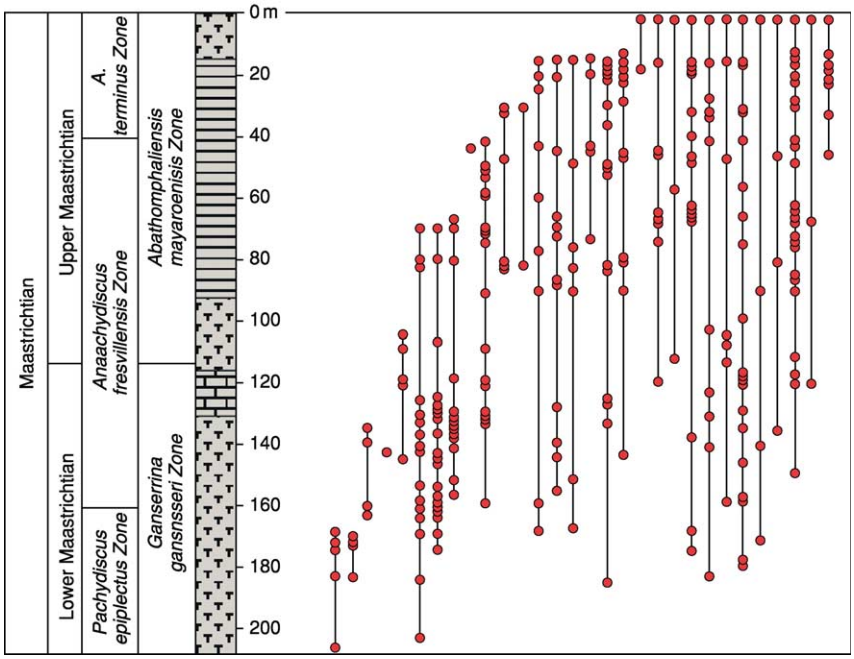


Figure 2 Maastrichtian ammonite biostratigraphy at Zumaya, Spain. Note that the majority of species (and, by implication, the genera and families they represent) disappear from this section during the course of the Maastrichtian, not at the Maastrichtian–Danian boundary. In fact, the last appearance datum of no ammonite species has been found to coincide with the Maastrichtian–Danian boundary in this section. Moreover, of the 12 species whose last appearance is within 1.5 m of the boundary, approx.half have exceedingly rare or patchy biostratigraphic distributions, which suggests patchy, uneven, extinction-susceptible population structures. In this section, the mass extinction of ammonites is equivalent to the near coincident disappearance of 6-7 previously abundant species, a phenomenon that would hardly be noticed in other parts of the Cretaceous ammonite record. Note also that, because of the patchy occurrence pattern characteristic of all ammonite species in this section, the idea that some ammonite species may have survived into the lowermost Danian interval cannot be rejected statistically. Redrawn from Marshall and Ward (1996) Sudden and gradual molluscan extinctions in the latest Cretaceous of western European Tethys. *Science* 274: 1360–1363.

one morphological condition or state into another. The other (cladogenesis) results when a single species is split into two or more daughter species, one of which may continue to exhibit the morphological condition of the ancestral population(s). Whereas the extinction of a lineage by the physical death of the last individual within the population must be regarded as a true extinction event, the physical death of the last individual within an extant population that simply happens to exhibit an ancient or atavistic morphological condition is not the same sort of event. This latter situation is termed pseudoextinction (Figure 3). Although the phenomenon of pseudoextinction need not be distinguished from true extinction in most routine biostratigraphic studies (after all, the pseudoextinction event does take place at a particular time), the differences between pseudoextinction and true extinction are profound in the context of extinction studies. Mammal systematist/palaeobiologist David Archibald has estimated that as many as 25% of the extinctions recorded in three different Early Puercan (=lowermost Danian) mammal lineages were pseudoextinctions.

Raw observations can also be deceiving in terms of the simple observation of fossil groups' last occurrence distributions. Since the last observed occurrence of each species is only an estimate of its true extinction coordinate in time and space (*see* **Biozones** for additional discussion), the vagaries of preservation and sampling will conspire to distort the observed extinction record for each species (and, by

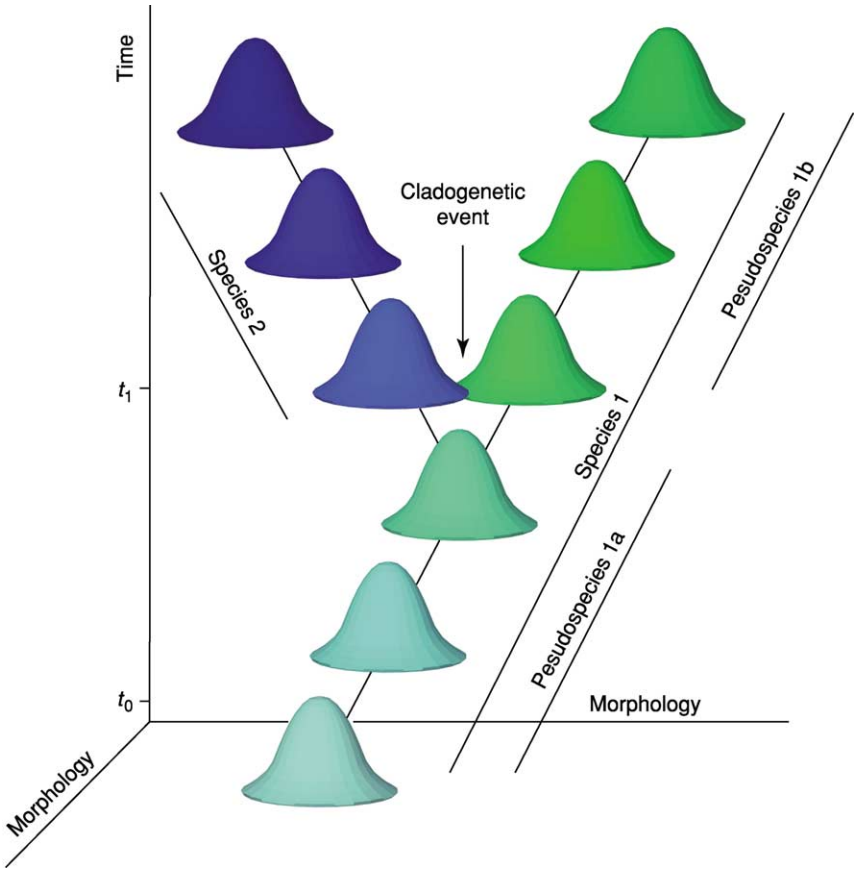


Figure 3 An example of the distinction between true extinction and pseudoextinction. Morphological change often occurs during the course of a lineage's evolutionary history. Morphological changes that result in a daughter species diverging from its ancestor are said to have arisen via cladogenesis (=lineage splitting). In the diagram cladogenesis has taken place approximately halfway up the time axis and resulted in the production of Species 2 from Species 1. Cladogenetically produced species are true species with a definite beginning (the cladogenetic event), history, and end (the extinction event). However, after a lineage has undergone so much within-lineage (=anagenetic) morphological change that no individuals at time t_1 exhibit the same morphological states as individuals in ancestral populations (time t_0), biostratigraphers often find it useful to differentiate between the t_0 and t_1 populations by giving a different species name to the latter. This convention, in effect, results in the 'extinction' of the t_0 species, not because of the death of the last individual belonging to the lineage, but merely as the result of a nomenclatural change. Such nomenclatural extinctions are termed pseudoextinctions. Without a detailed knowledge of the phylogenetic and nomenclatural history of a lineage, it is often impossible to distinguish true extinction from pseudoextinction.

extrapolation, for each genus, family, etc.) to a greater or lesser extent (Figure 4). This phenomenon was termed the Signor-Lipps Effect by the nominal authors in 1982, although Alan Shaw offered an essentially identical description of the same phenomenon as early as 1964. There is no way to correct for the Signor-Lipps Effect. Using the distribution of gaps between known occurrences within a species' or higher taxon's stratigraphic range, and making use of several simplifying assumptions, it is possible to estimate a confidence interval above which it is appropriate to regard the species or group as being truly extinct. Nevertheless, there is no method whereby an investigator may retrospectively pinpoint a taxon's 'true' extinction level. For groups of taxa various scenario-based extinction geometries can be evaluated using extensions of the confidence-interval method, but this approach will always identify a spectrum of geometries, ranging from randomized patterns to a strictly simultaneous geometry, as being equiprobable.

Finally, simple taxonomic uncertainty conspires to complicate data reported in the palaeontological literature. Certain species or higher groups that appear to become extinct at particular stratigraphic levels appear again, later in time at the same locality or in a different region altogether. Sometimes the morphological condition of the specimens at the second appearance is similar to that of predisappearance specimens and sometimes it is not. If the second appearance is regarded as a continuation of the predisappearance taxonomic entity it is referred to as a Lazarus taxon and should not be regarded as having undergone anything other than a local extinction at

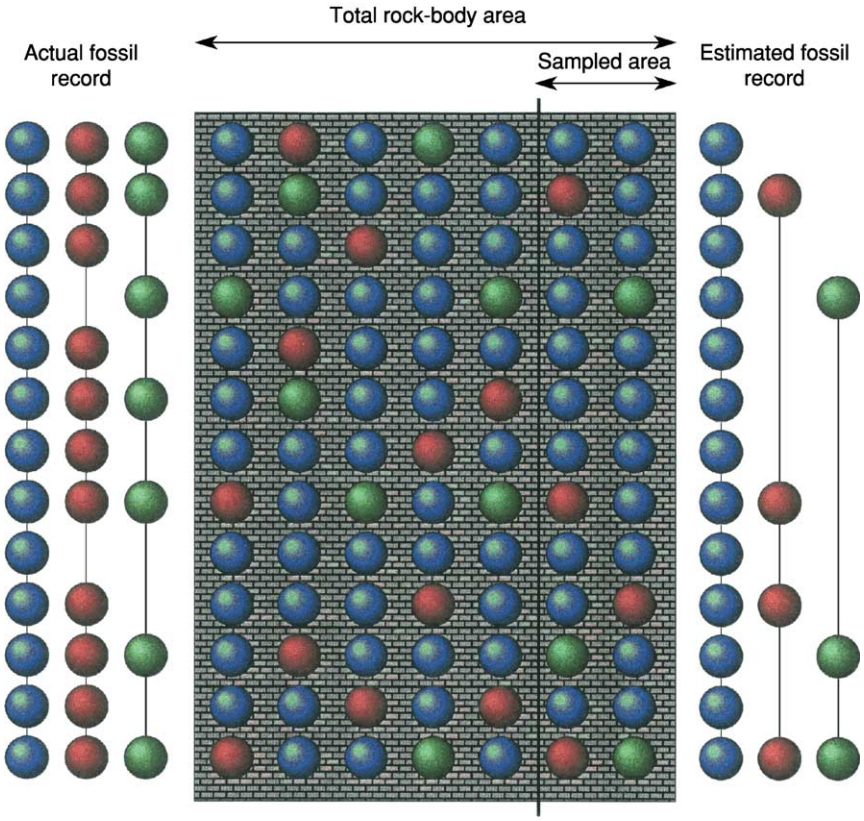


Figure 4 A hypothetical example of the Signor-Lipps Effect. The total fossil record of any species is provided by the sum total of all occurrences of the fossil within a body of rock. However, because a relatively small portion of that total rock body is ever available for sampling, the known fossil record represents a subset of the total fossil record. Because of this distinction between the total and known fossil records, and because of additional biases that might arise as a result of changes in the relative abundance of fossil species over time, changes in facies within single stratigraphic successions, and sampling patterns, estimates of the pattern of last-appearance data (e.g., coincident vs progressive) are usually biased towards the recovery of progressive patterns. In this example, note that despite the fact that all three species range through the entire section, ranges based on the subset of occurrences in the two right-hand columns yield a sequential extinction pattern. This phenomenon has been used to argue that progressive extinction patterns imply an abrupt extinction mechanism. Such arguments are invalid—as was pointed out by Signor and Lipps (1982)—because the progressive bias applies to genuinely progressive as well as genuinely abrupt true extinction patterns. The best one can do is not to reply on the apparent pattern of biostratigraphic last-appearance data as reliable estimates of true extinction rates.

the lower horizon. However, if the second appearance specimens differ in some way from their characteristic predisappearance forms such that their placement within the previously existing lineage is questionable, the second appearance may regarded as a suspect or 'Elvis' taxon, in which case the status of the group's extinction is uncertain.

Taking all of these potential biasing factors into consideration, the following sections outline our current understanding of Maastrichtian–Danian biotic turnover patterns for six different ecologically subdivided biotas. In all cases these data are based on the family-level compilations provided by *The Fossil Record 2* (Benton, 1993) as the most up-to-date single source of stratigraphic information across the broad spectrum of fossil groups. Although not as detailed taxonomically as JJ Sepkoski's *Compendium of Fossil Marine Animals* (2002), the former has the advantage of encompassing terrestrial as well as marine taxa and being compiled, reviewed, and adjudicated by specialists in each group.

Marine Microfossils

Marine microfossils, including protistan autotrophs and heterotrophs, have traditionally been thought to be one of the broad taxonomic groups most affected by the end-Cretaceous extinction. Review of the family-level fossil record largely bears this out. Among the major marine microfossil groups (Figure 5), only diatoms fail to exhibit a Maastrichtian extinction-intensity peak. For coccoliths, benthic foraminifera, and radiolaria this peak is more-or-less isolated, suggesting operation of a causal process or processes that were confined to the Maastrichtian. For dinoflagellates and planktonic foraminifera though, the Maastrichtian peak appears to be part

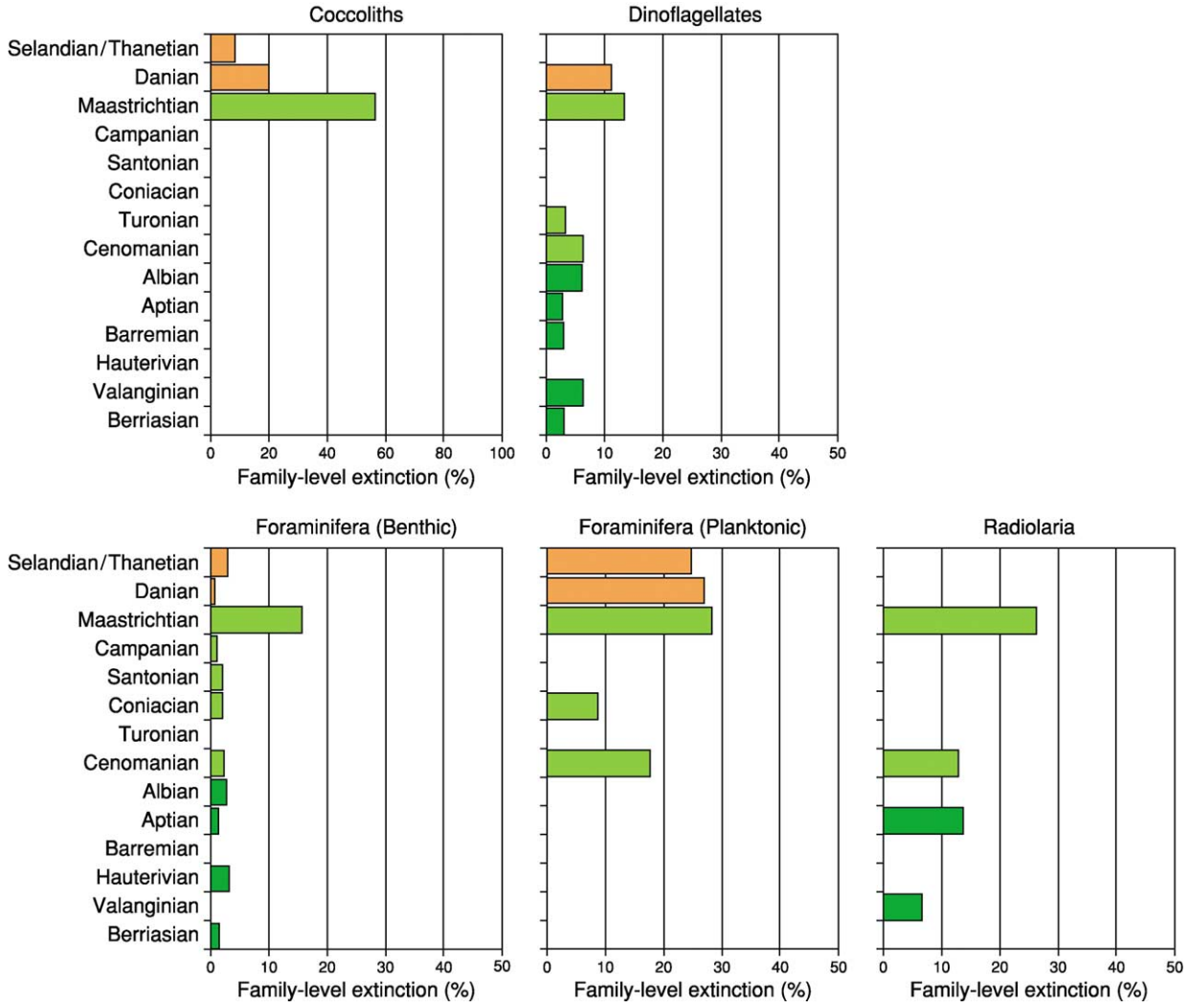


Figure 5 Cretaceous–Paleocene, stage-level extinction patterns for marine microfossil families. Note change in scale in the coccolith diagram. Data tabulated from Benton (1993) *The Fossil Record 2*. London: Chapman & Hall.

of a larger pattern that encompasses both the Maastrichtian and Paleocene intervals. Overall Maastrichtian extinction intensities across all six marine microfossil groups are less than 20%, suggesting an overall species loss similar to that of the total Maastrichtian estimate (see above).

Interpretation of this Maastrichtian–Danian microfossil record is complicated by several factors. A considerable controversy regarding the correct interpretation of Cretaceous planktonic foraminiferal species routinely found in lowermost Danian sediments continues. Some specialists regard Cretaceous isotopic values obtained from the analysis of particular species' skeletons, along with the widespread chaotic disruption of bedding patterns in lowermost Danian sediments, as indicative of widespread shelf failure at the Maastrichtian–Danian boundary with consequent reworking of Cretaceous species into Danian sediments. This shelf failure was presumably caused by the physical shock of bolide impact along with subsequent earthquakes. Others regard the recovery of Danian isotopic results from other Cretaceous species, the pristine preservation of many millions of Cretaceous microfossil skeletons in Danian sediments (fully comparable with those of undoubted Danian species and distinct from obviously reworked Cretaceous species), and the fact that Danian occurrences of Cretaceous species exhibit a clear biogeographic signal of greater penetration into the Danian in higher latitudes (where the effect of boundary disturbances is known to be reduced) as evidence for the survivorship of some species into Danian times. (There are also specialists who cite phylogenetic criteria as bearing on the survivorship question despite the fact that these arguments can be rejected on both logical and analytic grounds.) Two families are involved in this controversy, Globotruncanidae and Rugoglobigerinidae.

Relatively low Maastrichtian extinction intensities for dinoflagellates and diatoms have been accounted for by noting that the biology of these groups includes resting cyst stages and that these may have enhanced their overall survivorship potential. Species-level data for both groups from the K–T section on Seymour Island, Antarctica do not support this interpretation. Seymour Island cyst-forming dinoflagellates and diatoms exhibit progressive turnover patterns across the K–T boundary, suggesting that extinction-inducing environmental changes were not confined to any single horizon. There is an increase in diatom resting spores in the Upper Maastrichtian interval of this high-latitude section, but this occurs throughout the succession and is not confined to any single stratigraphic horizon. Moreover, since no modern diatom resting spore has been successfully revived after more than two years' dormancy, this sets an inferred maximum duration of environmental disruption that could be tolerated before wholesale extinction of the indigenous diatom flora—for which there is no evidence—would occur.

Marine Invertebrates

Extinctions among marine invertebrate groups exhibit a range of patterns (Figure 6) different from that of marine microfossils. Here, family-level data suggest that poriferans were among the most strongly affected groups, whereas corals exhibit scarcely any effect at all. Between these extremes molluscs, brachiopods, bryozoans, and echinoderms exhibit broadly progressive extinction patterns that rise throughout the Late Cretaceous to a Maastrichtian peak and then fall off into the Paleocene, whereas marine arthropods exhibit a coral-like indifference to whatever factors were driving these long-term changes within other groups.

Given its iconic status within the pantheon of end-Cretaceous victims, the mollusc record is particularly interesting. Throughout the Cretaceous mollusc extinction intensities are comparatively low. The traditional mollusc victim families of Inoceramidae, and the rudistid families Radiolitidae and Hipporitidae all record last appearances in the Maastrichtian, but these three losses alone represent over 20% of all marine bivalve family-level extinctions. Overall, Maastrichtian mollusc losses are subequally split between bivalves (14 families lost), gastropods (8 families lost), and cephalopods (12 families lost). Note that, in contrast to the popular perception of ammonites being iconic K–T victims, at the family level bivalves suffered more than ammonites during the Maastrichtian. It should also be noted that all of these groups suffered the bulk of their losses within the Maastrichtian and not at the K–T boundary itself.

Bryozoans are another group with an especially intriguing Cretaceous extinction record. Numbers of extinct bryozoan families appear to rise in a more-or-less uniform stage-level pattern to a Maastrichtian peak that represents just over 12% of the total Maastrichtian assemblage. Although all three bryozoan orders are affected by this protracted extinction, the ecological hammer appears to have fallen more forcefully on cheilostomes than on ctenostome or cyclostome forms. Interestingly, none of these extinctions are thought to have effected long-term patterns of bryozoan diversification substantially. The relatively comparable level of Maastrichtian extinction intensity exhibited by the bryozoans' sister group, the brachiopods, is a bit misleading in that the data presented in Figure 6 have been expressed as percentages.

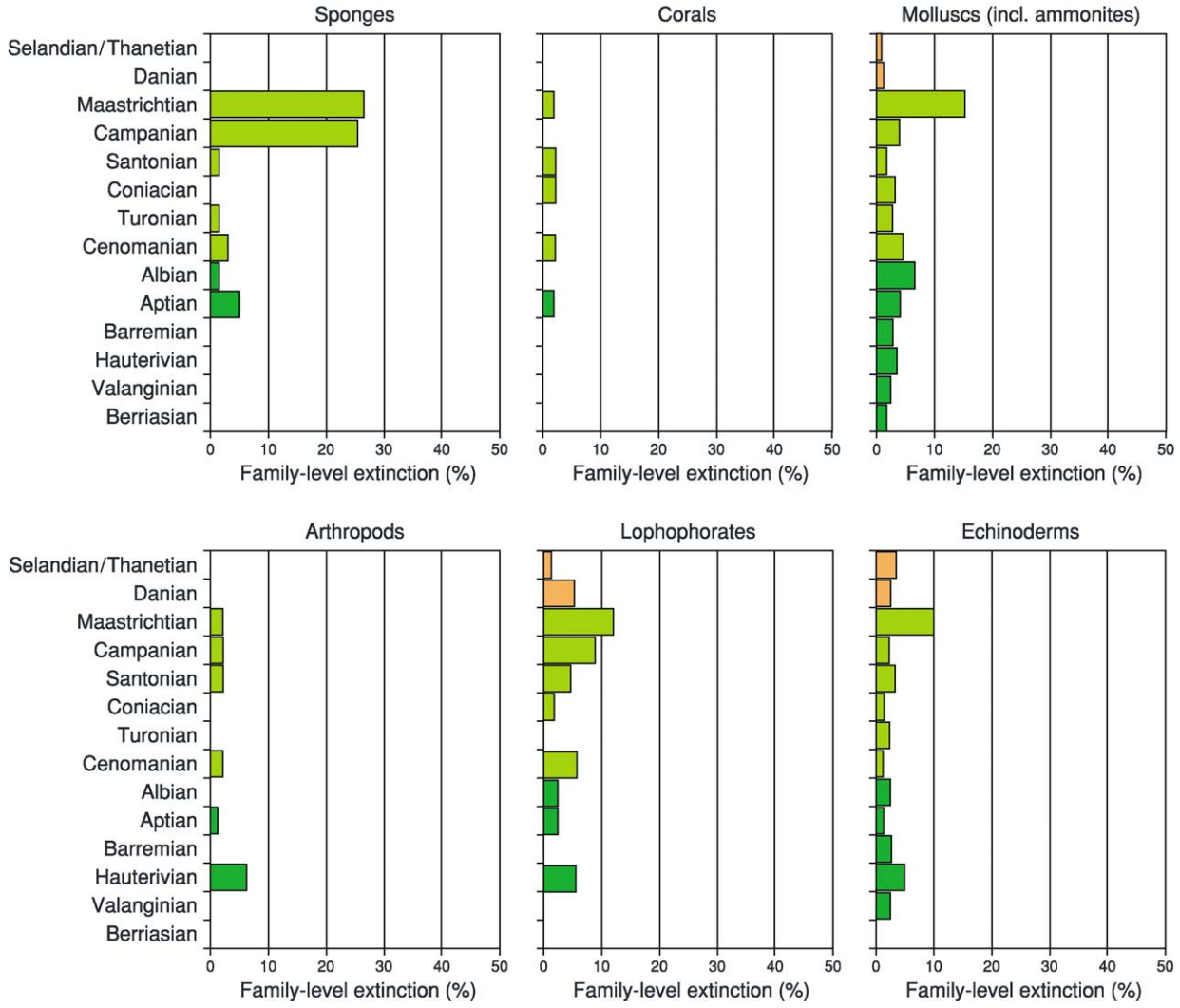


Figure 6 Cretaceous–Paleocene, stage-level extinction patterns for marine invertebrate families. Data tabulated from Benton (1993) *The Fossil Record 2*. London: Chapman & Hall.

In fact, overall Maastrichtian brachiopod family richness is less than a third that of bryozoans, and the actual number of brachiopod families that became extinct during this stage was comparably reduced. Among echinoderms, irregular echinoid families bore the brunt of the Maastrichtian extinction pulse.

Taken as a whole, it is difficult to avoid the impression that marine invertebrates were being subjected to much longer-term extinction pressures than their protistan counterparts with the Maastrichtian representing the apotheosis of trends that first become evident in the early part of the Late Cretaceous. Moreover, once the Maastrichtian catharsis was reached, it had relatively less overall effect on marine invertebrate family-level diversity than did the protistan reductions. As for the patterns of species-level extinction distributions, despite two decades of intensive study the fine-scale extinction pattern is known only for ammonites in two Late Maastrichtian sections (Seymour Island, Antarctica and Zumaya, Spain). In both cases extinctions observed extinction horizons are not concentrated at the K–T boundary, but rather are spread throughout the entire interval (see Figure 2).

Marine Vertebrates

This group, largely comprising fish, ichthyosaurs, mosasaurs, sauropterygians, and marine turtles, has rarely received much attention in Maastrichtian extinction studies despite the iconic status of mosasaurs and plesiosaurs as stereotypic K–T victims. One sometimes also sees ichthyosaurs mentioned as participants in the end-Cretaceous extinction event, but this is erroneous. The last ichthyosaur became extinct in the Albian. Overall (Figure 7), marine vertebrates

exhibit a distinct and rather abrupt Cenomanian peak, followed by a protracted Late Cretaceous buildup to a subordinate Maastrichtian peak, after which extinction intensities decline throughout the Paleocene. Although it should be noted that the pattern shown in Figure 7 is dominated by fish data, and that marine reptiles do show a much more distinct Maastrichtian extinction peak, it is also true that comparatively few Maastrichtian marine reptile families were present and that these were formed from relatively few genera. Moreover, no marine reptile extinction horizon is known to coincide with the K–T boundary.

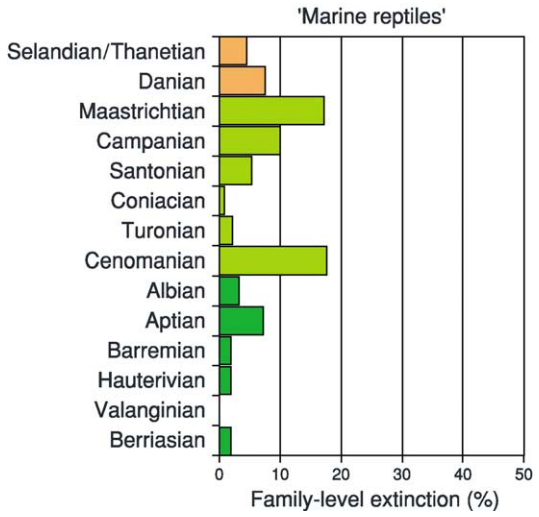


Figure 7 Cretaceous–Paleocene, stage-level extinction patterns for marine vertebrate families. Data tabulated from Benton (1993) *The Fossil Record 2*. London: Chapman & Hall.

Terrestrial Invertebrates

Relative to comparable marine taxa, freshwater and terrestrial invertebrates (gastropods, bivalves, chelicerates, crustaceans, ostracods, and insects) exhibit very modest levels and episodic patterns of extinction intensity throughout the Cretaceous with the Maastrichtian stage representing a decidedly subordinate intensity peak (Figure 8). Combined data from this ecological realm are inevitably biased by the very large number of insect families that, in all Cretaceous stages, are represented by values well over an order of magnitude larger than those for any other group. Despite this faunal size discrepancy though, Maastrichtian levels of percent extinction intensity for both freshwater and terrestrial mollusc and noninsect arthropod families are both less than those for insect families. The general character of Maastrichtian family-level extinctions for these former groups appears fully consistent with the qualitative notion of background, rather than mass, extinction. It should also be noted in passing that insects suffer their greatest Cretaceous extinction in the Albian rather than in the Maastrichtian.

Terrestrial Vertebrates

The Maastrichtian terrestrial vertebrate extinction record (freshwater fish, amphibians, reptiles, dinosaurs, and mammals) is by far the most diverse of any ecological realm (Figure 9). Even more than their marine counterparts, terrestrial fish faunas provide no evidence for heightened Maastrichtian extinction intensities. Indeed Maastrichtian fish families exhibit the lowest family-level extinction intensity of any Cretaceous stage. This remarkably quiescent

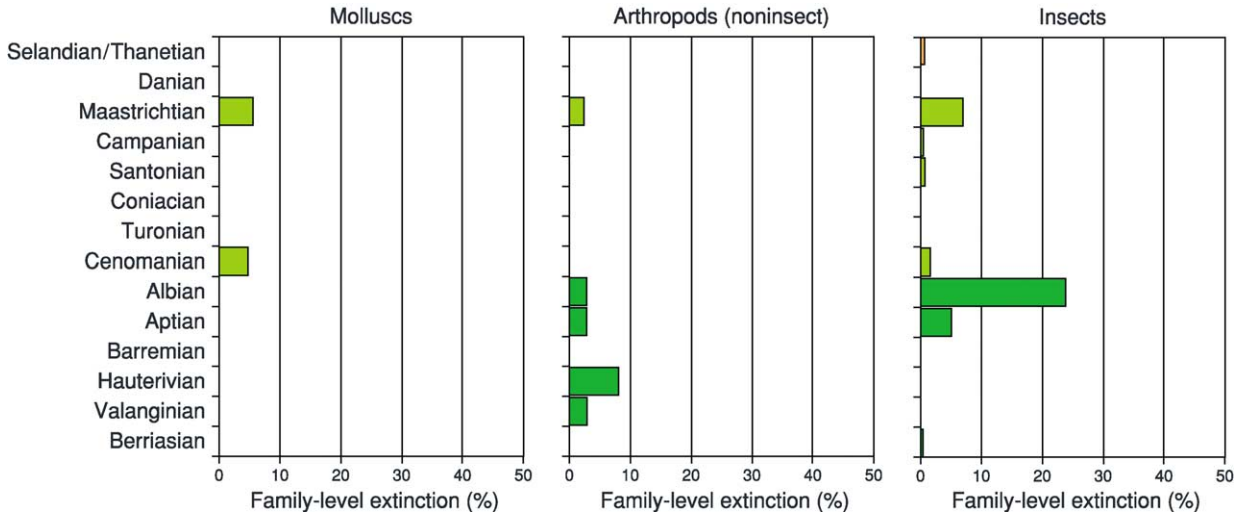


Figure 8 Cretaceous–Paleocene, stage-level extinction patterns for terrestrial–freshwater invertebrate families. Data tabulated from Benton (1993) *The Fossil Record 2*. London: Chapman & Hall.

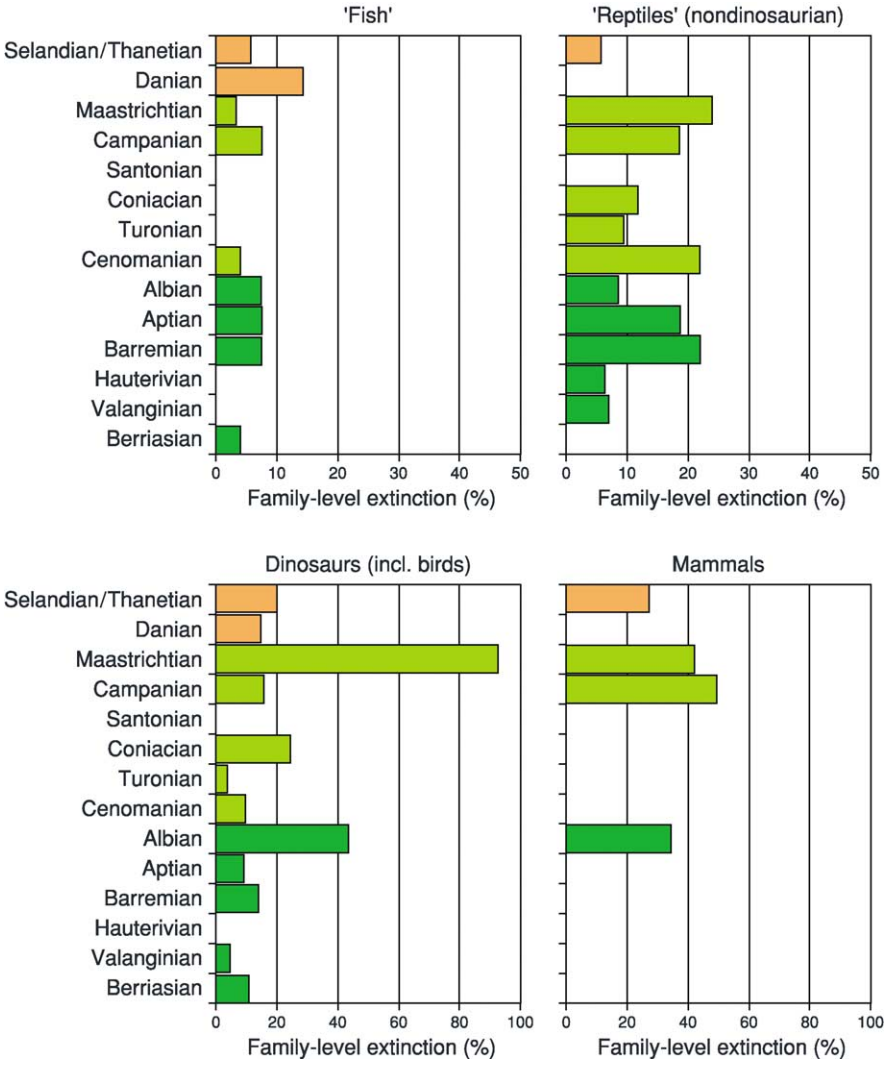


Figure 9 Cretaceous–Paleocene, stage-level extinction patterns for terrestrial–freshwater vertebrate families. Note change in scale in the dinosaur and mammal diagrams. Data tabulated from Benton (1993) *The Fossil Record 2*. London: Chapman & Hall.

record is exceeded only by that of amphibians, which fail to record any family-level extinctions in any Cretaceous stage. Cretaceous extinction levels for non-dinosaurian reptiles are much higher, but neglect to exhibit a pronounced Maastrichtian peak despite the loss of five crocodylomorph families. This pattern is reversed, however, in the Dinosauria with its impressive 21 dinosaur families lost over the course of the Maastrichtian. Supporters of catastrophic dinosaur extinction scenarios hold that most, if not all, of these extinction occurred at the K–T boundary despite the fact that no 'in-place' dinosaur bone has been found within more than a metre of that horizon to date. Finally, as noted for Cretaceous family richness patterns, the Cretaceous mammalian extinction record bears an intriguing similarity to that of the Dinosauria, with coincident Albian and Maastrichtian peaks along with a unique Campanian peak that exceeds that of the Maastrichtian in terms of overall extinction intensity. Although dinosaurs (including birds) and mammals represent the two most severely affected clades among all terrestrial groups, and two of the most severely affected across all ecological realms, it should nevertheless be kept in mind that, in terms of overall extinction intensity, dinosaurs and mammals represent only 10% of all Maastrichtian land-based families and less than 5% of the overall Maastrichtian biota.

Plants

As has been noted elsewhere in the extinction literature, Cretaceous plants appear to have been relatively extinction-resistant (Figure 10). There is a hint of a slight increase in Maastrichtian plant extinction intensities, but nothing that could be spoken of as being extraordinary relative to other Late Cretaceous

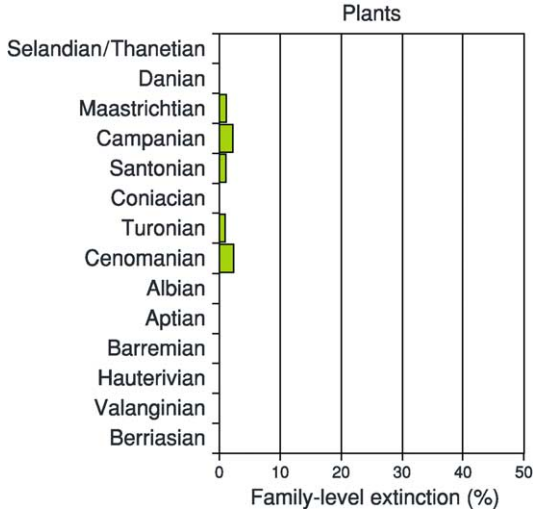


Figure 10 Cretaceous–Paleocene, stage-level extinction patterns for terrestrial plant families. Data tabulated from Benton (1993) *The Fossil Record 2*. London: Chapman & Hall.

stages. In addition, the pattern of extinctions suggested heightened extinction susceptibility beginning in the Campanian and continuing into the Danian. Over this time interval the Maastrichtian extinctions that do occur are focussed in gymnosperm lineages.

Extinction Causes

There are three single-cause models and one composite cause model currently used to account for the end-Cretaceous extinctions. The former are based on changes in global sea-level, increases in large-province volcanism, and bolide impact.

Sea-Level

Three distinct, short-term, sea-level drops occurred during the Maastrichtian. These were superimposed on a long-term declining trend from the all-Cretaceous sea-level highstand at or near the Cenomanian–Turonian boundary. The oldest of these is an ~100-m drop across the Campanian–Maastrichtian boundary, followed by a slightly larger and more acute drop across the Early-Late Maastrichtian, followed by a significantly larger (~200-m drop) and still more abrupt drop that took place at or just before the end of the Maastrichtian. These sea-level drops are reflected in $\partial^{18}O$ isotopic values that suggest a long-term cooling of global climates throughout the Turonian–Maastrichtian interval (presumably due to decreased albedo and reduced atmospheric CO_2) coupled with extreme sea-surface temperature instabilities in the Maastrichtian. Aside from further reducing the already-reduced area of submerged continental shelf available for colonization and foraging by marine taxa, these sea-level drops would be expected to have resulted in intensified global climate gradients (due to reduced equator-to-pole heat transfer efficiency), increased seasonality, and the development of local centres of bottom water anoxia. Sea-level has long been a suspected cause of many 'mass extinction' events in the geological past. A total of seven major sea-level lowstands have occurred over the past 250 My, and of these three exhibit a stage-level association with peaks in the extinction–intensity curve.

Large Igneous Province (LIP) Volcanism

The Maastrichtian witnessed a particularly large flood-basalt volcanic event when the (then) island continent of India moved over the Reunion mantle plume or 'hotspot' currently located in the southern Indian Ocean. Radioisotopic dates indicate that these Deccan volcanics were erupted in a series of quasi discrete events that range in age from 65.7 ± 1.3 Ma to 65.4 ± 1.3 Ma. These dates suggest that over 1×10^6 km^3 of extrusive volcanics were emplaced over an ~300 000- to 500 000-year interval that straddled the Maastrichtian–Danian boundary (65.5 Ma). The primary climatic effect of these eruptions would have been a brief global cooling (due to increased albedo brought about by injection of SO_2 into the stratosphere), longer-term global warming (due to release of volcanogenic CO_2, possibly along with the melting of frozen methane hydrates on the continental shelves), and acid rain (due to the SO_2). Secondary effects would include enhanced marine anoxia (due to global warming and sluggish marine circulation), disruption of weather patterns (due to the thermal anomaly created by the cooling lava), enhanced local-regional earthquake activity, and increase rates of subareal weathering. The emplacement of large igneous-volcanic bodies such as the Deccan Traps has long been suspected as a cause of many mass extinction events in the geological past. There have been six associations between large LIP volcanic eruptions and extinction-intensity peaks during the past 250 My.

Bolide Impact

Few people interested in science can be unaware of the dramatic discovery that an ~10-km (diam.) comet or asteroid collided with the Earth at the precise time of the Maastrichtian–Danian boundary. Indeed, the iridium (Ir) anomaly in marine and terrestrial sediments that marks this event has been accepted as the primary defining feature of this boundary in temporally complete K–T successions worldwide. Subsequent subsurface geological

investigations, first in cores from boreholes drilled near the town of Chixculub on Mexico's Yucatan Peninsula, and later from regional Bouger gravity surveys, revealed the presence of a 250-km-wide, circular, multiringed basin that corresponds to predictions of the crater size such an impact would be expected to leave. Radioisotopic dates of melted basement rock from the original Chixculub cores confirmed the age of this crater as 65.5 ± 0.1 Ma. In 1996 this date was accepted as the official age of the Maastrichtian–Danian (=K–T) boundary.

The physical and climatic effects of a Chixculub-size impact would be expected to be similar to those of an LIP volcanic event. Unlike a volcanic event, a thermal blast would precede the arrival of a massive shock wave at localities thousands of miles away from the impact site itself. At the same time, a variety of materials would be injected into the stratosphere, resulting in a sharp-but-short global cooling phase, followed by a longer-term global warming. Swarms of major earthquakes would also have been produced locally as the Earth's crust adjusted to the deformation. Some have predicted global darkness brought about by this stratospheric dust injection, plunging the Earth into a months- to years-long night, but recent reviews of this proposed effect have reduced the duration of impact-related global darkening considerably. Acid rain would be produced as a result of both the bolide's passage through the atmosphere and its ejection of a large volume of vaporized evaporitic material from the impact site, though, once again, the severity of this acid-rain phase has been downgraded recently from initial estimates. The initiation of global wildfires as a result of melted ejecta is another part of the original bolide impact scenario that, once again, has been downgraded as a result of subsequent research findings. The remaining, accepted, primary and secondary phenomena (see above) are all remarkably similar to the predicted effects of an LIP volcanic event, with the bolide impact effects capable of being distinguished from the latter only in terms of:

1. different trace element signatures in impact vs volcanogenic glass spherules,
2. the production of different mineralogic artefacts (e.g., multiple sets of shock lamellae in the crystal lattices of mineral grains), and
3. the shorter time period over which the primary and secondary effects would have operated.

With respect to the latter, it should be noted that, although the intensity of the postimpact environmental effects would be greater than those generated as a result of LIP volcanic eruptions, bolide impact effects would also be expected have attenuated at a much more rapid rate over time since—unlike LIP eruption event scenarios—disruptive event renewal would not have been possible. In effect, the important ecological distinction that can be drawn between bolide impact and volcanic disruptions is that of a single, massive, disruptive shock to global ecosystems after which normal conditions are reestablished versus a continuous series of less intense shocks spread over a much longer time interval. Bolide impacts had long been discounted as a major extinction cause in the past but interest in this mechanism has undergone a dramatic turn-around since 1980. Still, of the six major bolide impact events (crater diam. $\geq$50 km) that are known to have occurred over the past 250 My, only three exhibit stage-level associations with peaks in the extinction-intensity curve.

Multiple Causes

Although it has not been defined precisely—especially in terms of positive or negative feedbacks between different disruption sources—a commonly stated body of opinion suggests that the most reasonable stance is to admit that all of the previous mechanisms, as well as others, took place in the Maastrichtian and may have contributed to precipitating the end-Cretaceous extinctions, each in their own way. The interesting feature of this model is not the existence of causal mechanisms but their apparently coincident timing. This model gains credibility not from the detail with which its adherents can currently construct scenarios of predicted environmental effects, but from:

1. the variety of extinction patterns present among the various groups affected by the overall extinction event (see above),
2. the lack of a consistent and/or simple ecological signal among organisms that inhabit different ecological realms,
3. the fact that all local peaks in extinction—intensity curves, regardless of size, are associated with the time series of at least one of these component mechanisms, and
4. the fact that, over the past 250 My, all three large extinction events occur during a time of confluence between two or more of these component mechanisms.

Opinion surveys among research palaeontologists with regard to the most likely cause of the end-Cretaceous extinctions were conducted in 1984 and 1996. On both occasions 'geologists' (=researchers with expertise outside palaeontology) were overwhelmingly in favour of the single-cause, bolide impact explanation for end-Cretaceous extinctions—especially the dinosaur extinctions—whereas palaeontologists (including experts in dinosaur and nondinosaur groups) were overwhelmingly in favour

of the multiple-cause model. If this survey were to be repeated this author is quite sure the results would be similar.

See Also

Atmosphere Evolution. **Biozones**. **Fossil Invertebrates:** Arthropods; Brachiopods; Bryozoans; Corals and Other Cnidaria; Echinoderms (Other Than Echinoids); Molluscs Overview. **Fossil Vertebrates:** Fish; Palaeozoic Non-Amniote Tetrapods. **Microfossils:** Foraminifera. **Plate Tectonics**. **Sequence Stratigraphy**. **Stratigraphical Principles**. **Tertiary To Present:** Paleocene. **Time Scale**. **Volcanoes**.

Further Reading

Archibald JD (1996) *Dinosaur Extinction and the End of an Era: What the Fossils Say. Critical Moments in Paleobiology and Earth History.* New York: Columbia University Press.

Benton MJ (1993) *The Fossil Record 2*. London: Chapman & Hall.

Galvin C (1998) The great dinosaur extinction controversy and the K-T research program in the late 20th Century. *Earth Sciences History* 17: 41–55.

Glasby GP and Kunzendorf H (1996) Multiple factors in the origin of the Cretaceous-Tertiary boundary: The role of environmental stress and Deccan Trap volcanism. *Geologische Rundschau* 85: 191–210.

Hallam A and Wignall PB (1997) *Mass Extinctions and Their Aftermath*. Oxford: Oxford Science.

Hoffman A and Nitecki MA (1985) Reception of the asteroid hypothesis of terminal Cretaceous extinctions. *Geology* 13: 884–887.

MacLeod N, *et al.* (1997) The Cretaceous-Tertiary biotic transition. *Journal of the Geological Society of London* 154: 265–292.

Marshall CR and Ward PD (1996) Sudden and gradual molluscan extinctions in the latest Cretaceous of western European Tethys. *Science* 274: 1360–1363.

Ryder G, Fastovsky D, and Gartner S (1996) *The Cretaceous-Tertiary Event and Other Catastrophes in Earth History*, Special Paper 307. Boulder, CO: Geological Society of America.

Sarjeant WAS (1999) Dinosaur extinction: sudden or slow, cataclysmic or climatic? *Geoscience Canada* 23(3): 161–164.

Sepkoski JJ Jr (2002) *A Compendium of Fossil Marine Animals*. Bulletins of American Paleontology. 363. Ithaca, NY: Paleontological Research Institute.

Sharpton VL and Ward PD (1990) *Global Catastrophes in Earth History: An Interdisciplinary Conference on Impacts, Volcanism, and Mass Mortality*, Special Paper 247. Boulder, CO: Geological Society of America.

Shaw A (1964) *Time in Stratigraphy*. New York: McGraw-Hill.

Signor PW III and Lipps JH (1982) Sampling bias, gradual extinction patterns and catastrophes in the fossil record. In: Silver LT and Schultz PH (eds.) *Geological Implications of Impacts of Large Asteroids and Comets on the Earth*, Special Paper 190, pp. 291–296. Boulder, CO: Geological Society of America.

Silver LT and Schultz PH (1982) *Geological Implications of Impacts of Large Asteroids and Comets on the Earth*, Special Paper 190. Boulder, CO: Geological Society of America.

Williams ME (1994) Catastrophic versus noncatastrophic extinction of the dinosaurs: Testing, falsifiability, and the burden of proof. *Journal of Paleontology* 68: 183–190.

METAMORPHIC ROCKS

Contents

Classification, Nomenclature and Formation

G Hoinkes and C A Hauzenberger, University of Graz, Graz, Austria
R Schmid, ETH-centre, Zurich, Switzerland

Introduction

Many names for metamorphic rocks are inconsistently used. A recommended classification of these rocks, which is acceptable worldwide, does not yet exist but will, however, be presented in near future by the IUGS Subcommission on the Systematics of Metamorphic Rocks (SCMR). In this chapter, a classification principle and definitions for metamorphic rocks are presented which have been extracted from provisional recommendations of SCMR, presented on their homepage at 'http://bgs.ac.uk/SCMR/' (see at end of this article **Further Reading** list) and from our own considerations.

How to Classify and Name a Metamorphic Rock

Metamorphic rocks originate, if not already of metamorphic origin, from sedimentary and igneous rocks. Therefore, their bulk chemical composition is extremely variable, and because they can have formed under pressures (*see* **Ultra High Pressure Metamorphism**) and temperatures (*see* **Thermal Metamorphism**) ranging between those existing some kilometres beneath the Earth's surface and those existing when rocks start to melt, under the influence of varying fluid compositions or metasomatic processes, the mineralogical composition (*see* **Minerals:** Definition and Classification) is far more variable than that of igneous or sedimentary rocks. This contrasts with the small number of well-known and most frequently used names for metamorphic rocks.

The definitions of these names unfortunately are not based on the same features or properties. Some definitions use purely mineralogical, others structural or a combination of structural and mineralogical properties, others chemical parameters or genetic considerations, or otherwise (Tables 1a, b, and 2). This means that only some of the terms can be used to derive a coherent classification system encompassing all the possible varieties of metamorphic rock types, because such a system requires that each subdivision in that system use only one criterion for each subdivision. If it is not built up in this manner, it will contain many gaps, and will not be able to offer names for rocks falling within such gaps.

The Principal Solution of This Problem

The available rock names are split into three groups, and four major rules are set up on how to use the terms in these groups

Main Specific Rock Names

Eight terms (Table 1a) are applied to frequently occurring metamorphic rock types and in their definition describe the essential mineral mode of these rock types. They are valuable because the nearest macroscopic characterization of a metamorphic rock is achieved by indicating its mineral mode. Therefore, the first rule of the classification system proposed here is that when a rock sample fulfils the requirements of the definition of a main specific rock term listed in Table 1a the pertinent rock name is the correct name for the sample in question. The rule excludes in this case one of the structural root names (see below) being applied. It would give rise to ambiguities, if two different names for one and the same rock type were in use. The rule should, therefore, be strictly followed.

Minor Specific Rock Names

Terms define the mineral mode less well or define other properties of the rock (Table 1b). Here, a second rule similar to the first units is applied, but less strictly.

Table 1a Definitions of the main specific rock names which have to be given preference over the equivalent structural root names schist, gneiss, or granofels

Amphibolite: *Metamorphic rock mainly (to more than 50% vol.) consisting of green, brown, or black amphibole and plagioclase. The modal content of pyroxene is larger than 30% vol. and larger than the percentage of any one of the other mafic minerals. Plagioclase is a major mineral (>5% vol.).*

Other common minerals in amphibolite are quartz, chlorite, epidote, zoisite, biotite, garnet, titanite, scapolite, and calcite. Their presence should be indicated by prefixing them, e.g., *garnet bearing clinopyroxene amphibolite*, where garnet is a minor and clinopyroxene a major constituent. Only in the case of the special type of amphibole or plagioclase being present, these constituents should be prefixed as well (e.g., *bytownite amphibolite*).

Calc-silicate rock: *Metamorphic rock mainly composed of Ca-rich silicate minerals such as wollastonite, vesuvianite, diopside-hedenbergite, titanite, grossular- and andradite-rich garnet, prehnite, meionite-rich scapolite, zoisite, clinozoisite, epidote, and pumpellyite.*

Less Ca-rich silicates such as plagioclase, tremolite, etc., are common additional minerals as well as opaque minerals. Also carbonate (calcite or aragonite ± dolomite) may be present in an amount of up to 50% by vol.

Rocks of similar mineral composition, formed by metasomatism or contact metamorphism, should be classified as skarns or calc-silicate hornfelses, respectively.

Eclogite: *Plagioclase-free metamorphic rock composed of more than 75% omphacite and garnet, both of which are present as major minerals, the amount of neither of them being higher than 75% vol.*

Granulite: *High-grade metamorphic rock, in which anhydrous Fe-Mg-silicates are dominantly anhydrous. Cordierite may be present and is not counted as either a hydrous or anhydrous mineral. The term should not be applied to ultramafic rocks, calc-silicate rocks, marbles, and ironstones. This rule implies that granulites contain more than 10% vol. of feldspar and/or quartz.*

The mineral composition of the rock is to be indicated by prefixing all of the major constituents present, e.g.:

- Biotite-sillimanite-garnet-plagioclase granulite: rock sample composed of the major minerals plagioclase, garnet, sillimanite, and biotite, where plagioclase?garnet?sillimanite?biotite.
- Pyroxene-plagioclase granulites ± garnet, amphibole: rock series in which each individual sample contains as major minerals pyroxene (cpx and/or opx) and plagioclase, and sometimes also garnet and/or amphibole.

The following collective terms for certain groups of granulites are proposed:

- Granulites containing the phase assemblage orthopyroxene, quartz, K-feldspar or mesoperthite and ±plagioclase may be called *charnockitic granulites*.
- The term *mafic granulites* embraces all granulites containing pyroxene±amphibole±garnet as the dominant ferromagnesian phase assemblage. They contain plagioclase, rarely K-feldspar and quartz, and very rarely aluminosilicates (only at high P).
- In contrast, those granulites containing garnet±biotite±cordierite as the dominant ferromagnesian phase assemblage may be called, due to the lack of a better term, *felsic granulites*, even if they are frequently restitic and may contain a large amount of garnet. They contain plagioclase and/or K-feldspar or mesoperthite, ± quartz and ±aluminosilicates.
- The collective term *quartzo-feldspathic granulites* or *quartz-feldspar granulites* should only be applied, if all members of the rock group contain both of the minerals quartz and feldspar.

Marble: *Metamorphic rock containing more than 50% vol. of calcite and/or dolomite and/or aragonite.*

Pure marble contains more than 95% vol. of these carbonate minerals, whereas the remainder are classified as *impure marble*.

Phyllite: *Metamorphic rock, in which the individual grains are large enough to be seen by the unaided eye (>0.1 mm) and which is characterized by a lustreous sheen and a well-developed schistosity resulting from the parallel arrangement of phyllosilicates.*

Phyllite is usually of low metamorphic grade.

Pyriclasite: *Metamorphic rock mainly (to more than 50% vol.) consisting of pyroxene (cpx and/or opx) and plagioclase. The modal content of pyroxene is larger than 30% vol. and larger than the percentage of any one of the other mafic minerals (>5% vol.).*

Other common minerals in pyriclasite are garnet and/or amphibole and/or biotite. The presence of them may be indicated by prefixing them, e.g., *biotite bearing garnet pyriclasite*. Only in the case of the special type of pyroxene or plagioclase being present, these constitutents should be prefixed as well, e.g., *cpx pyriclasite*, if only clinpyroxene is present, or *cpx-opx pyriclasite*, if both pyroxenes are present.

The term was created and first defined by Berthelsen in 1960.

Slate: *Metamorphic rock, in which the individual grains are too small to be seen by the unaided eye (<0.1 mm) and in which the schistosity is developed on the grain scale.*

Slate is usually of very low metamorphic grade and rich in phyllosilicates.

A minor specific rock name may be used in cases where it is appropriate and makes sense to do so.

Structural Root Names

The three terms 'schist', 'gneiss', and 'granofels' are defined in (Table 2). The third rule simulates that for a rock sample, for which no major or minor specific rock term can or should be applied, one of the three structural root names schist, gneiss, and granofels can to be applied. These names are delineated from each other by one single criterion, the quality of schistosity developed within them (Table 2). If the schistosity is well developed, it is a schist, if poorly developed, it is a gneiss, and if there is no schistosity present, it is a granofels. Thus, any metamorphic rock, for which no specific name is available, can be described by one of these three terms. In order to indicate the mineral mode of the sample, all minerals present in the rock in a volume of more than 5% (= major minerals) have to be arranged in front of the structural root name

Table 1b List of minor specific rock names names which need not be given preference over the equivalent structural root names schist, gneiss, or granofels, because they only poorly define the modal composition of a rock and should therefore only be applied when it is appropriate to do so. The list only presents some examples of minor specific rock names and is not complete

1.Rock names defined by mode and rock colour:
Blueschist, Greenschist, Greenstone, Whiteschist.
2.Rock name refering to the protolith and its alteration:
Spilite.
3.Genetic to semi-genetic rock names:
Restite, Rodingite, Skarn.
4a.Names for cohesive fault rocks:
Mylonite, Protomylonite, Mesomylonite, Ultramylonite, Augen mylonite, Blastomylonite, Phyllonite. Pseudotachylite.
4b.Names for cohesive or incohesive fault rocks:
Cataclasite, Protocataclasite, Mesocataclasite, Ultracataclasite, Fault breccia
4c.Names for incohesive fault rocks:
Fault gouge
5.Names for composite rocks:
Migmatite
6.Names for contact metamorphic rocks:
Hornfels
7.Names for impact rocks:
Impact breccia, Impact melt rock, Impactite

Table 2 Definitions of the three structural root names, and of the structural terms used in these definitions

Schist: *A metamorphic rock displaying a schistose structure.*
The term schist may also be applied to a rock displaying a linear rather than a planar fabric, in which case the expression 'lineated schist' is applied
For phyllosilicate-rich rocks, the term schist is reserved for medium- to coarse-grained varieties, wheras finer-grained rocks are termed phyllites or slates.
The mineral composition of the rock is to be indicated by prefixing all of the major constituents present.
Gneiss: *A metamorphic rock displaying a gneissose structure.*
The term gneiss may be also applied to a rock displaying a linear rather than a planar fabric, in which case the expression 'lineated gneiss' is applied.
The mineral composition of the rock is to be indicated by prefixing all of the major constituents present
Granofels: *A metamorphic rock displaying a granofelsic structure.*
The mineral composition of the rock is indicated by prefixing all of the major constituents present
For granofels containing layers of different composition the expression 'layered (or banded) granofels' may be used.
Note: In English there was no term for a massive rock devoid of a planar or linear fabric, until R. Goldsmith proposed, in 1959, the name granofels for such a rock type.
Schistosity: *A type of foliation produced by metamorphic processe, characterised by the preferred orientation of inequant mineral grains or grain aggregates.*
A schistosity is said to be well developed if inequant mineral grains or grain aggregates are present in a large amount and show a high degree of preferred orientation.
If the degree of preferred orientation is low or if the inequant grains or grain aggregates are only present in small amounts, the schistosity is said to be poorly developed.
Schistose structure: *A type of structure characterized by a schistosity which is well developed, either uniformly throughout the rock or in narrowly spaced repetitive zones, such that the rock will split on a scale of one cm or less.*
Gneissose structure: *A type of structure characterized by a schistosity which is either poorly developed throughout the rock or, if well developed, occurs in broadly spaced zones, such that the rock will split on a scale of more than one centimetre.*
Granofelsic structure: *A type of structure resulting from the absence of schistosity, such that the mineral grains and aggregates of mineral grains are equant, or if inequant, have a random orientation.*
Mineralogical or lithological layering may be present.

chosen, in order of increasing modal abundance, and hyphenated amongst themselves, but not with the rock name. Optionally, minor minerals, present in a volume of $\geq 5\%$, may be mentioned by prefixing them with '.... bearing' (see Flow Chart (Figure 1)). This comprises the fourth rule of the classification system.

A flow chart How to name a Metamorphic Rock is presented in Figure 1, which when followed, will enable each Earth scientist to find the correct name for each metamorphic rock. The flow chart consists of subsequent questions, and on each sides of the questions, comments which help in gaining the right

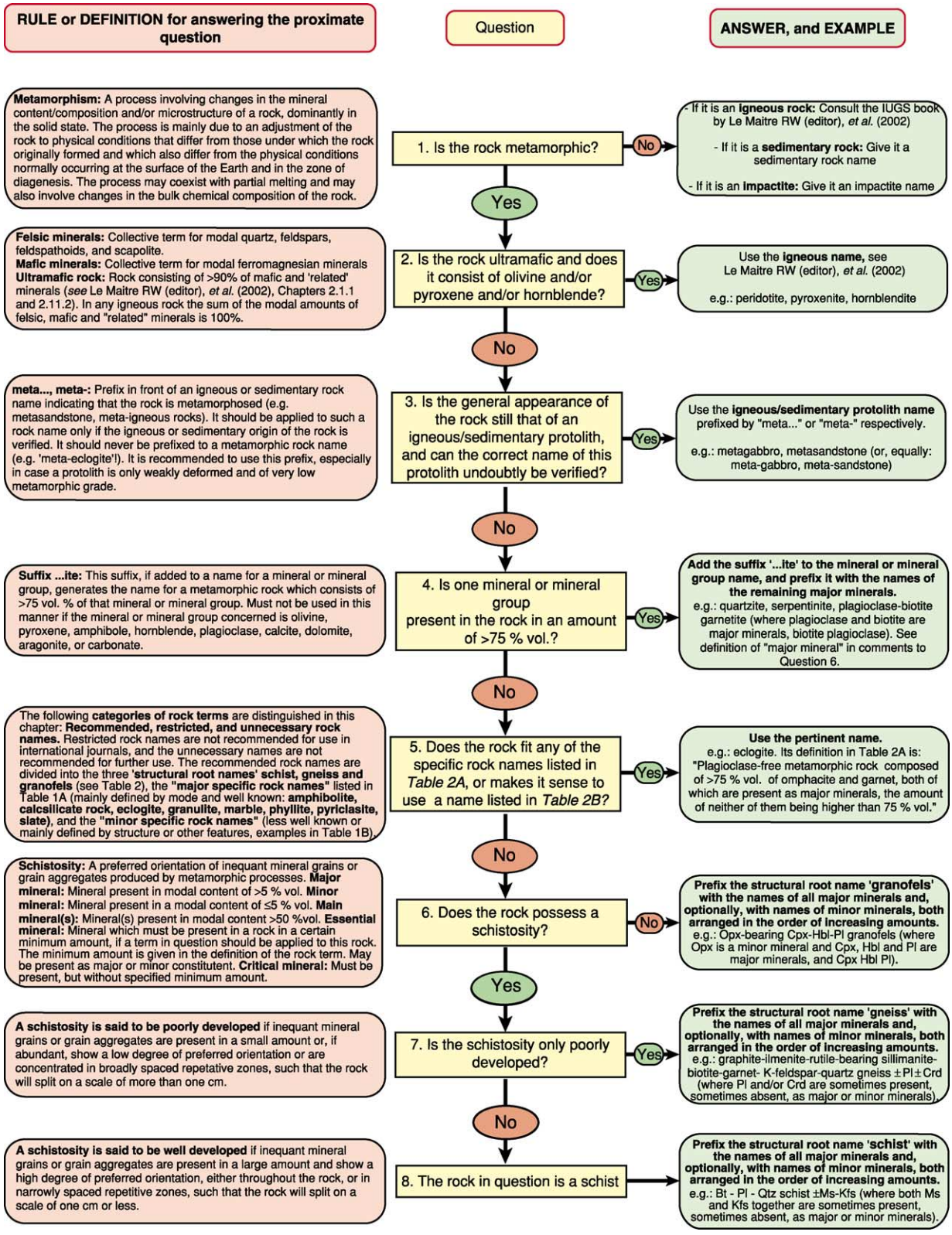


Figure 1 Flow Chart 'How to name a Metamorphic Rock'.

Table 3 Terms for structures partially used in the rock definitions of this chapter

Structure: *The arrangement of the parts of a rock mass irrespective of scale, including spatial relationships between the parts, their relative size and shape and the internal features of the parts.*

Dependent on the scale of the feature, the terms micro-(thin section or smaller scale), meso-(hand specimen scale), and mega-(outcrop or larger scale) can be used as a prefix.

Fabric: *The relative orientation of the parts of a rock mass.*

This is commonly used to refer to the crystallographic and/or shape orientation of mineral grains or groups of grains, but can also be used on a larger scale. Preferred linear orientation of the parts is termed linear fabric, preferred planar orientation planar fabric and the lack of a preferred orientation is referred to as random fabric.

Augen: *Distinctive large lenticular mineral grains or grain aggregates in a finer grained matrix of a rock, having the shape of an eye in cross-section, around which the foliation of the matrix is wrapped, commonly forming symmetric or asymmetric trails.*

Cleavage: *The property of a rock to split along a regular set of parallel or sub-parallel closely spaced surfaces.*

More than one cleavage may be present in a rock.

Fault: *A dislocation surface along which there has been displacement of the rock mass on one side relative to the one on the other side.*

Foliation: *Any repetitively occurring or penetrative planar feature in a rock body. Examples include:*

- *layering*
- *preferred planar orientation of inequant mineral grains (planar fabric)*
- *preferred planar orientation of lenticular or elongate grain aggregates.*

More than one kind of foliation with more than one orientation may be present in a rock. Foliations may become curved or distorted. The surfaces to which they are parallel are called s-surfaces. Where possible, the type of foliation should be indicated.

Fracture: *A general term for any surface of break in a rock mass, whether or not it causes displacement, due to mechanical failure.*

Fracture includes cracks, joints, and faults.

Joint: *A single fracture in a rock with or without a small amount (<1 cm) of either dilational or shear displacement.*

Joints may be infilled during or after their formation.

Layer: *One of a sequence of near parallel, tabular-shaped rock bodies.*

Layering: *The feature displayed by layered rocks.*

Band: *Used as a synonym for layer.*

Banding: *Used as a synonym for layering.*

Lineation: *Any repetitively occurring or penetrative visible linear feature in a rock body. This may be defined by:*

- *alignment of the long axes of elongate mineral grains (= mineral lineation)*
- *alignment of elongate mineral aggregates*
- *alignment of platy mineral grains parallel to and around an axis*
- *parallelism of hinge lines of small scale folds (= crenulation lineation)*
- *intersection of two foliations (= intersection lineation)*
- *slickenside striations or fibres.*

More than one kind of lineation, with more than one orientation, may be present in a rock. Lineations may become curved or distorted. The lines to which they are parallel are called l-lines. Where possible the type of lineation should be indicated.

answers. The comments comprise definitions or rules to be taken into consideration, and a practical example to each of the questions. Complementary information to the individual Questions 1 to 8 of the flow chart is given below. Terms for structures used in some of the rock definitions of this chapter are presented and defined in Table 3.

Additional Information to Questions 1 to 8 of the Flow Chart

Question 1: Is the Rock Metamorphic?

There are two possible cases in which it may be difficult to answer this question, especially when this has to be done after macroscopic inspection of a rock, and before thin section or analytical data are available:

i. The rock looks like an unmetamorphosed sedimentary or igneous rock but may, in fact, be weakly metamorphosed. When in doubt as to whether it is metamorphosed or not, choose the name of the protolith. When, during later laboratory work, it can be established that the rock is in fact metamorphosed, prefix the protolith name with 'meta…' or 'meta-', respectively (see comment in the flow chart to Question 3).

ii. The rock is mafic or ultramafic, the mafic minerals are olivine and/or pyroxene and/or amphibole, the microstructure is granoblastic, and a schistosity is not or only weakly developed (see definition of schistosity in the comment to Question 6 in the flow chart). It might be an igneous rock which was emplaced at deep crustal levels, where it crystallized and/or re-equilibrated at ambient pressures and temperatures. At these pressures and temperatures, igneous and metamorphic microstructures and phase assemblages converge, and thus the following is proposed:

- When the rock is ultramafic, use the pertinent ultramafic rock name in the IUGS-publication by Le Maitre (ed.) and others, published in 2002. For

the definition of 'ultramafic rock', see comment to Question 2 in the flow chart.
- When the rock is mafic; answer Question 1 with 'yes' and proceed down the flow chart. You will then have the choice to either use an igneous name prefixed by 'meta-' (Question 3) or call your rock a 'granofels' (Question 6).

iii. The rock is an impactite. Even if it is metamorphic, first have a look at the paper by Stoeffler and Grieve about impactites, on the homepage of SCMR at http://bgs.ac.uk/SCMR.

Question 5: Does the Rock Fit Any of the Specific Rock Names Listed in Table 1a, or Does it Make Sense to Use a Name Listed in Table 1b?

i. A name for a metamorphic rock should at best characterize its modal mineral contents as well as its structure. Whereas the structure can easily be described by prefixing specific rock names with schistose, gneissose, or granofelsic, indication of the minerals which are present in a metamorphic rock and of their relative amounts is less easy to perform. Therefore, the specific rock names listed in Table 1a, which are defined by mineral content and which are well known, are important and have, therefore, to be given preference over the the equivalent structural root names, schist, gneiss, or granofels. As an example: a rock containing the essential minerals plagioclase and amphibole in the amounts defined in the definition of amphibolite (Table 1a) should always be called amphibolite, and never an amphibole-plagioclase gneiss. The names in Table 1a should be prefixed by those major minerals, which are not essential minerals (see comment to Question 6 in the flow chart).

ii. The specific rock names in Table 1b are defined using structural or other features or properties. The list is by far from complete. If such special features or properties should be highlighted in the name rather than mineral content, such rock names may be used, giving them preference over the the equivalent structural root names, schist, gneiss or granofels. As an example: during field observations the term mylonite may be applied to a fault rock which is derived from a biotite-quartz-plagioclase gneiss. In a hand specimen collection of a museum, the same rock may be called a very fine grained biotite-quartz-plagioclase schist, with or without mentioning that it originates from a fault zone.

Questions 6–8:

i. *(i) Schistosity as a parameter for the subdivision of remaining rocks into schist, gneiss, and granofels:* It is obvious that the parameter 'perfection of schistosity' used for the subdivision into schist, gneiss, and granofels is only a qualitative one. Mainly rocks at the transition 'well developed' and 'poorly developed schistosity' will be qualified by some colleagues as schists, and by others as gneisses. Fortunately, no severe complications are expected to arise from this.

ii. ***Mentioning the mineral contents of metamorphic rocks when applying structural root names:*** The main information about a metamorphic rock comes from its mode which is semiquantitatively fixed. This is done by arranging the names of the major minerals in order of increasing modal abundance and hyphenated amongst themselves in front of the structural root name or, if not present in each individual sample of a group of rocks, behind the root name, preceeded by '±' and separated by a comma, or if not co-existing, by a hyphen, respectively. It is very important that all major minerals are prefixed with reference to a single rock sample or all those major minerals consistently occurring in each specimen of a group of rocks. Ambiguities would arise, if this rule was not obeyed.

The presence of minor minerals (present in a volume of $\leq 5\%$ by volume) may be accounted for by prefixing them, hyphenated with the suffix '-bearing', either in front of the major minerals or behind the structural root name. When not present in each sample of a set of rocks, a '±' is set in front of the respective mineral or the paragenesis of minerals.

SCMR is preparing a list of abbreviations of mineral names containing the abbreviations already published, as well as additional ones. These abbreviations may preferentially be used when prefixing minerals in front of rock names.

Formation of Metamorphic Rocks

Introduction

In petrology, the term metamorphism refers to a change in a rock's mineralogy, structure and/or composition during geological processes that occur predominantly in a solid state under conditions between diagenesis and large-scale melting. Most metamorphic rocks retain some of their parental heritage such as chemical composition and to some extent primary structural features such as bedding. The process of metamorphism is driven by changes in physical and/or chemical conditions, usually as a result of large-scale geological processes (plate tectonics).

Temperature Increasing temperature promotes recrystallization of fine-grained rocks as well as

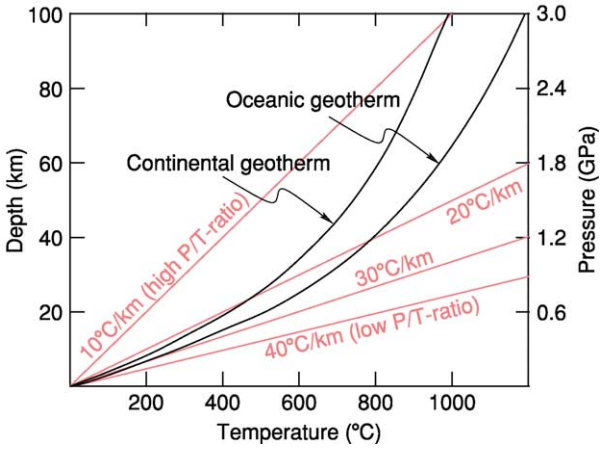


Figure 2 Typical geothermal gradients observed in the Earth's crust. Geotherms in oceanic crust are higher than in continental crust.

the crystallization of new minerals. The temperature experienced by a rock is controlled by the production, conduction, and advection of heat during geological processes (*see* **Thermal Metamorphism**). The main source of heat production in the lithosphere is radioactive decay, which is typically some microwatts per cubic metre. Minor contributions come from mechanical (e.g., friction) and chemical (latent heat of reaction) processes. The transport of heat within the lithosphere occurs by conduction and/or advection. The first describes the diffusive heat transport, the latter the active transport of heat, for example, by magmatic intrusions or fluids. Geothermal gradients describe the temperature increase with depth (Figure 2). Typical geothermal gradients are in the range of 15–30°C km.

Pressure Pressure increases with depth due to the weight of overlying rocks and is called lithostatic pressure (*see* **Ultra High Pressure Metamorphism**). The relationship between depth and pressure ($P = N/m^2$) is given by the density (ρ) of the rock ($\rho = m/V$). The density of crustal rocks is ~2500–2800 kg/m^3; that of mantle rocks ~3100–3300 kg/m^3. Thus, pressure increases by ~0.027 GPa (= ~270 bar) per kilometre in the crust and by ~0.032 GPa (= ~320 bar) per kilometre in the mantle.

Fluid During most metamorphic processes, an intergranular fluid of varying chemical composition is interacting with the solid phases of a metamorphic rock. The pressure of the fluid is generally considered to be close to the lithostatic pressure except for porous rocks in shallow crustal levels where the fluid forms a continuous network extending to the Earth's surface. In that case, the fluid pressure is equivalent to the overlying fluid weight, which is the hydrostatic pressure. Intergranular metamorphic fluids are usually dominated by H_2O. Additionally, CO_2 and minor amounts of CH_4 and N_2 as well as NaCl may be present in the fluid phase. Fluids can come from meteoric sources, juvenile magmatic sources, rocks dehydrating during prograde metamorphism, and trapped sedimentary brines. Fluids play an important role in enhancing the speed of metamorphic reactions, crystallization, and crystal growth, and in material transport within the crust.

Time The duration of metamorphism has an important influence on the formation of metamorphic rocks. The shortest period of time is realised by impact metamorphism (*see* **Impact Structures**), which is completed within seconds. Contact metamorphism can last from tens of thousands of years to almost a million years. Typical time spans for regional metamorphism are in millions to tens of millions of years. During this time, the rocks are subject to changes in temperature, pressure, and possibly fluid composition. Usually the structure and minerals formed at the highest temperature and pressure are preserved. However, during exhumation some late-stage mineral forming reactions may lead to a retrograde overprint of the peak assemblages.

Classification of Metamorphism

Regional metamorphism (*see* **Regional Metamorphism**) occurs over wide areas, affects large volumes of rocks, and is associated with tectonic processes such as plate collision and crustal thickening (orogenic metamorphism) and ocean-floor spreading (ocean-floor metamorphism). During the subsidence of large basins, sedimentary piles can be affected by PT conditions of 200–400°C and 0.05–0.2 GPa. These conditions mark the transition from diagenesis to metamorphism and are named burial metamorphism.

i. ***Orogenic Metamorphism*** is associated with various phases in the course of an orogenic cycle and involves compressional and extensional regimes. The pressure–temperature conditions cover a wide range (300–1000°C, 0.3–3 GPa), depending on the specific mountain building processes.
ii. ***Ocean-floor metamorphism*** is caused by circulating hot seawater, resulting in hydration of the formerly water-free magmatic mineral assemblages of the newly formed oceanic crust at mid-ocean ridges. The temperature range may vary considerably (200–700°C), depending on the proximity of the spreading centre. Pressure is typically low (<0.5 GPa).

Local metamorphism occurs in relatively small areas around magmatic intrusions (contact metamorphism), meteorite impacts (impact metamorphism), or certain fault zones (dislocation metamorphism).

i. ***Contact metamorphism*** is caused by igneous intrusions as a result of the thermal effects of hot magma on the surrounding cooler country rock. Temperatures may be up to 1000°C at relatively low pressures (0.05–0.5 GPa). Pyrometamorphism is a special form of contact metamorphism at very high temperatures (>1000°C) and very low pressures (<0.1 GPa), affecting small fragments of country rock in a volcanic or subvolcanic environment.
ii. ***Hydrothermal metamorphism*** is caused by hot H_2O-rich fluids. Metasomatism is commonly associated with this type of metamorphism.
iii. ***Impact metamorphism*** is due to an enormous amount of kinetic energy which is released to surface rocks by a large-scale meteorite (asteroid, or comet) impact. Extremely high temperatures partly vaporise and melt the rocks. The impact also produces some ultra-high pressure minerals (*see* **Ultra High Pressure Metamorphism**), such as stishovite, usually not found in the Earth's crust and also lead to amorphous (isotropic) and other anomalous behaviour of minerals.

Metamorphic Processes and Reactions

Mineral assemblages observed in metamorphic rocks are directly related to their chemical composition (X) and to a specific temperature (T) and pressure (P). Rocks can be considered as mechanical mixtures of phases (minerals, melts, fluids, gases), which can be mixtures of phase components. As an example, we will look at a rock which consists of the minerals garnet (grossular; $Ca_3Al_2Si_30_{12}$), kyanite (Al_2SiO_5), and quartz (SiO_2). At a given temperature and pressure, these three mineral phases have a unique molar Gibbs free energy (G), a quantity that describes the energy state of the rock. The Gibbs free energy of the rock is the sum of the Gibbs free energies of its phases, in this case garnet, kyanite, and quartz. However, the same chemical system can be described by a single mineral, the plagioclase end-member anorthite ($CaAl_2Si_2O_8$), if the following relationship is applicable:

$$1\,\text{Grossular} + 2\,\text{Kyanite} + 1\,\text{Quartz} = 3\,\text{Anorthite} \quad [1]$$

According to the thermodynamic laws, the mixture with the lowest (most negative) total Gibbs free energy is the stable mixture, while the other mixture is metastable. Within a P-T diagram (*see* **Metamorphic Rocks: PTt - Paths**) we find exactly one line (= univariant reaction curve), where the Gibbs free energies of the assemblage garnet, kyanite, and quartz, and of anorthite are equal (eqn [2]; Figure 3). Below this reaction curve anorthite has a lower Gibbs free energy than 1 garnet + 2 kyanite + 1 quartz and hence is stable (eqn [3]) above the reaction curve 1 garnet + 2 kyanite, + 1 quartz are stable.

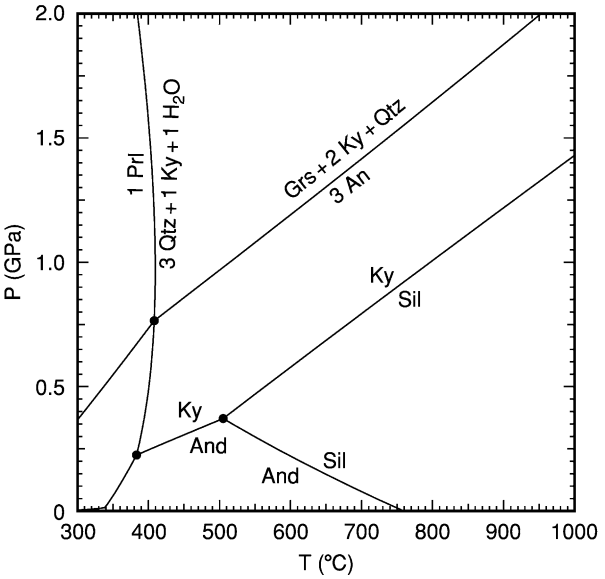


Figure 3 PT-diagram showing the reactions Grossular (Grs) + 2 Kyanite (Ky) + Quartz (Qtz) = 3 Anorthite (An) and Pyrophyllite (Prl) = 3 Qtz + Ky + H_2O. In addition, the stability fields of the Al_2SiO_5 phases Ky, Andalusite (And), and Sillimanite (Sil) are shown.

$$\Delta G_r = 3G_{An} - G_{Grs} - 2G_{Ky} - G_{Qtz} = 0 \quad [2]$$

$$\Delta G_r < 0 \quad [3]$$

The slope of a reaction in a P-T diagram can be approximated by a simple relationship of change of reaction volume (ΔV) and reaction entropy (ΔS):

$$dP/dT = \Delta S_r/\Delta V_r \text{ (Clausius} - \text{Clapeyron equation)} \quad [4]$$

Reactions involving only solid phases (minerals) phases tend to have a relatively large change in volume (ΔV_r) compared to their change in entropy (ΔS_r) and show relatively flat-lying reaction curves in a P-T diagram (eqn. [1], Figure 3). Reactions involving a fluid phase (e.g., H_2O) are different. As an example, take the pyrophyllite (Prl) dehydration reaction (Figure 3):

$$1\,\text{Prl} = 3\,\text{Qtz} + 1\,\text{Ky} + 1\,H_2O \quad [5]$$

The volume of H_2O decreases significantly with increasing pressure so the slope of the reaction curve can

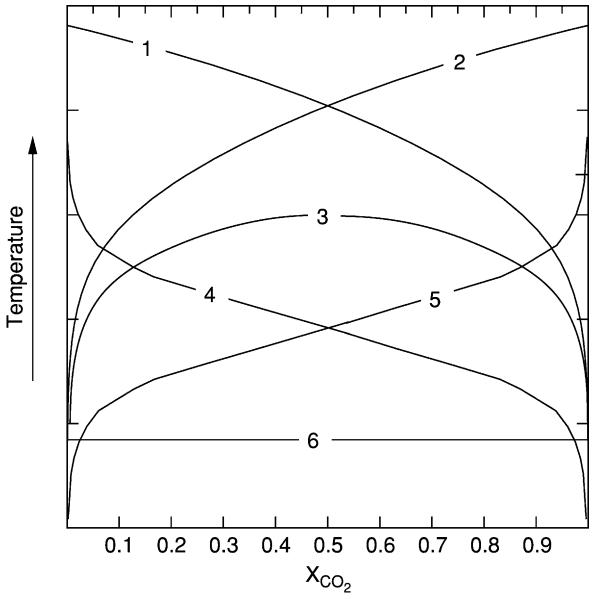


Figure 4 Different reaction types in T-X_{CO_2} space.

become negative. Reactions involving only an exchange of cations (e.g., the exchange of Fe and Mg between garnet and biotite) typically have a small ΔV_r and a large change in entropy (ΔS_r), and hence a steep slope in *P-T* space.

The stability of carbonates and hydrous silicates is not only influenced by P and T, but also by the fluid composition in terms of CO_2 and H_2O (X_{CO_2}) which can be shown in a *T*-X_{CO_2} diagram. The reactions that take place in marbles and carbonate-bearing rocks can be grouped into six types, each of which has a distinctive shape to the equilibrium curve on a *T-X* diagram (Figure 4): (i) dehydration reactions; (ii) decarbonation reactions; (iii) dehydration-decarbonation reactions; (iv) carbonation-dehydration reactions; (v) hydration-decarbonation reactions; and (vi) fluid-absent reactions).

Metamorphic Structures

Metamorphic rocks can form in many different stress environments and at different heating rates. They contain structures ranging from massive to highly foliated and show extremely fine to very coarse grain sizes. Recrystallization and/or mineral growth during metamorphism are the main mechanisms, producing structures that are diagnostically important in recognizing metamorphic rocks. Minerals formed by metamorphic processes are called blasts. Porphyroblasts occur as coarse grained crystals in a finer grained matrix (Figure 5A). Poikiloblasts contain numerous inclusions of other minerals (Figure 5B).

Structural criteria for determining the relative ages for deformation and metamorphic mineral growth are characterized by the relationship between the internal schistosity (S_i) within the mineral blasts and the external schistosity (S_e) of the dominant matrix of the rocks. Within polydeformed areas, it is necessary to explain the relation between the mineral growth and the respective deformational phase. Predeformational growth is defined, when S_i is discordant to S_e, or when a break between S_i and S_e develops. During the younger deformational overprint (S_2), the foliation in the matrix often wraps around the blasts and pressure shadows develop (Figure 5C). Syn-deformational growth is defined by a blast that consist of sigmoidal, spiral shaped inclusion patterns, that develope during rotational growth by shearing along the schistosity planes (Figure 5D). The shape of the inclusion patterns are, therefore, best indicators for establishing the shear direction during metamorphic mineral growth. Figure 5D shows no pressure shadows around the blasts and a continuously developed inclusion pattern from S_i (S_1) to S_e (S_2). Post-deformational growth shows blasts that contain inclusion trails that are parallel or continuous to the external schistosity ($S_i = S_e$) (Figure 5E). Corona structures are observed when a mineral breaks down to a new phase or phase assemblage which prevents the progress of reaction. This is commonly observed in granulite-facies rocks. Symplectites develop when a mineral becomes unstable over a very short period of time, which is the case in the exhumation of eclogites (Figure 5G).

Metamorphism of Different Protoliths

Protoliths (other than already metamorphosed rocks) may be of either magmatic or sedimentary origin. For simplicity, magmatic protoliths may be divided into three different groups: ultramafic, mafic, and quartzo-feldspathic rocks. The most important sedimentary protoliths are pelitic (clay-stones), quartzo-feldspathic, carbonate rocks and mixtures of the three groups, such as marls (carbonate-rich pelitic sediments to clay-rich carbonates). Figure 6 shows metamorphic rock names in relation to protolith and grade of metamorphism.

Ultramafic protoliths are mainly peridotites and related rocks from the Earth's mantle and reflect a rather simple MgO-rich silicate composition. The most abundant metamorphic products are serpentinites.

Mafic protoliths are mainly mafic igneous rocks (basic to intermediate intrusive and extrusive rocks) and to a much lesser extent marls of unusual composition similar to basalts. With increasing metamorphic grade they result in very significant metamorphic rock types, depending on the *P/T*-ratio (Figure 2). Low to

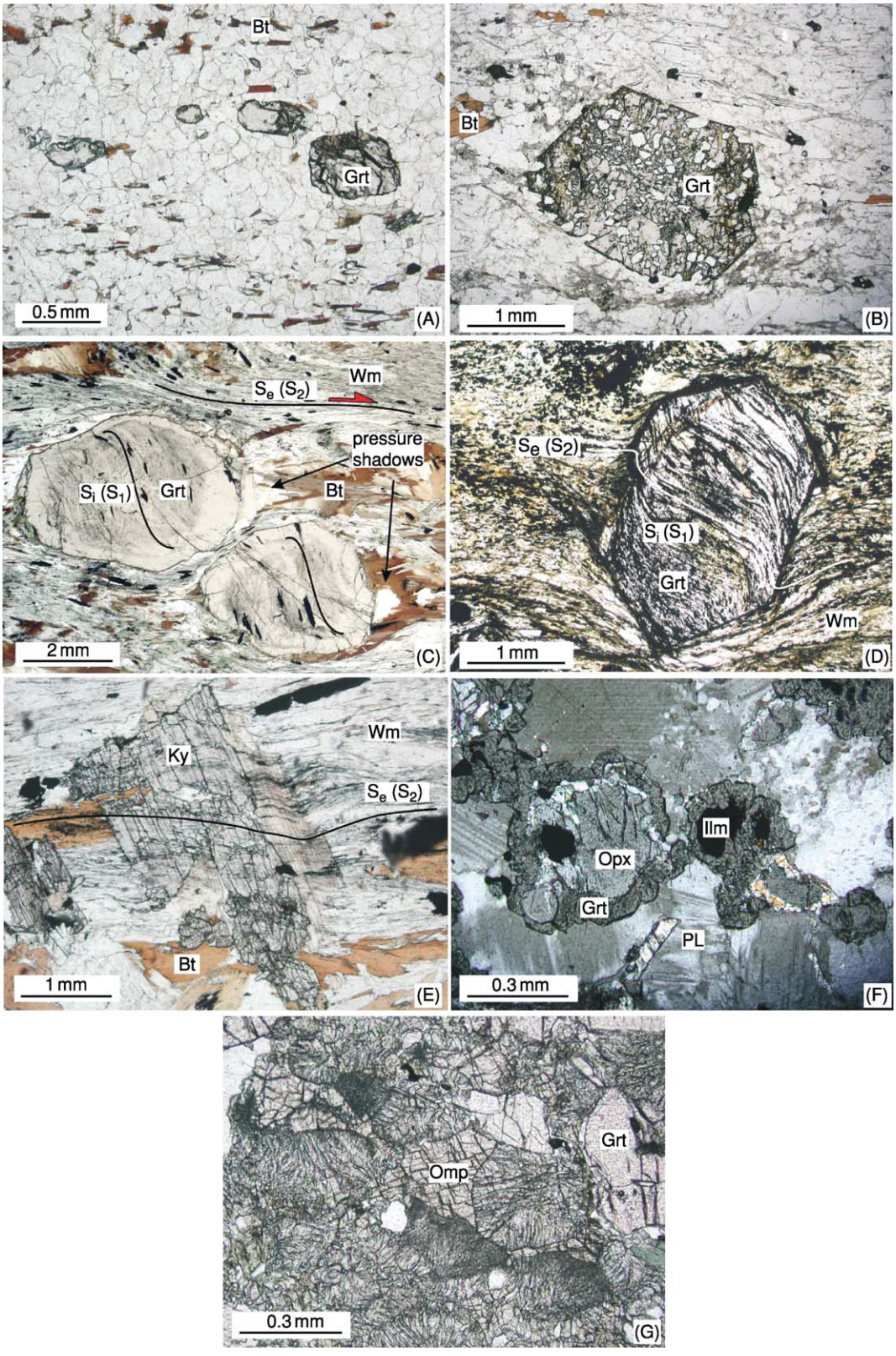


Figure 5 Typical mineral structures observed in metamorphic rocks; (A) A garnet porphyroblast occurs in a finer grained granoblastic matrix; (B) Poikiloblastic garnet containing numerous inclusions of other minerals; (C) Pre-deformational garnet porphyroblasts is indicated by a discordant internal foliation S_i (S_1) in garnet. During the younger deformational overprint (S_2), the foliation in the matrix S_e commonly wraps around the blasts and pressure shadows develop; (S) Syn-deformational growth is defined

Continued

medium P/T-gradients lead to greenschists, amphibolites, mafic granulites, and/or eclogites, whereas high P/T-gradients are responsible for blueschists and eclogites.

Quartzo-feldspathic protoliths are either represented by intermediate to acid magmatic rocks (intrusive and extrusive) or by clastic sediments derived from these magmatites. If derived from sediments, they form metamorphic schists, paragneisses, migmatites and felsic granulites with increasing grade. If derived from granitoids, the metamorphic products are meta-granitoids, orthogneisses, migmatites, and felsic granulites with increasing grade.

Pelitic protoliths are mainly composed of clay minerals derived from weathered and eroded continental crust. With increasing metamorphic grade, they are transformed to slates, phyllites, mica-schists, and granulites.

Carbonate rocks mainly originate from biogenic sediments. Pure carbonate sediments are SiO_2-free. The metamorphic products are marbles, composed of calcite and/or dolomite. Impurities by SiO_2-rich clastic sediments result in silicate marbles during metamorphism.

Marls may resemble mafic protoliths in terms of chemical composition but usually differ by significant amounts of $CaCO_3$. Low-grade metamorphic marls are calcareous micaschists, whereas the formation of Ca-rich silicate minerals during higher-grade metamorphism results in calcsilicates.

Diagnostic Mineral Assemblages with Prograde Metamorphism

Ultramafic rocks

The primary mineral assemblage of peridotite is represented by forsterite–enstatite±diopside. In order to enable prograde metamorphism, these anhydrous assemblages must first be hydrated to serpentinites in the course of ocean floor metamorphism. The starting assemblage for prograde metamorphism of diopside-free peridotite is the low temperature form of serpentine (chrysotile), brucite, and/or talc, depending on the amount of enstatite and forsterite in the protolith. At about 200°C, antigorite starts replacing chrysotile and at 300°C antigorite is the only serpentine phase stable up to more than 500°C, where it breaks down to forsterite and talc. This assemblage may be replaced by anthophyllite at slightly more than 600°C and low pressures. From 670°C, the peridotite assemblage forsterite and enstatite is also the stable metamorphic assemblage. In peridotites containing considerable amounts of CaO, metamorphic diopside and tremolite are present. Diopside is also a common stable phase in serpentinites at greenschist-facies conditions but is replaced by tremolite and forsterite at amphibolite-facies conditions (*see* **Metamorphic Rocks:** Facies and Zones). It comes in again with enstatite at the amphibolite–granulite-facies transition replacing tremolite and forsterite. Additional Al_2O_3 stabilizes

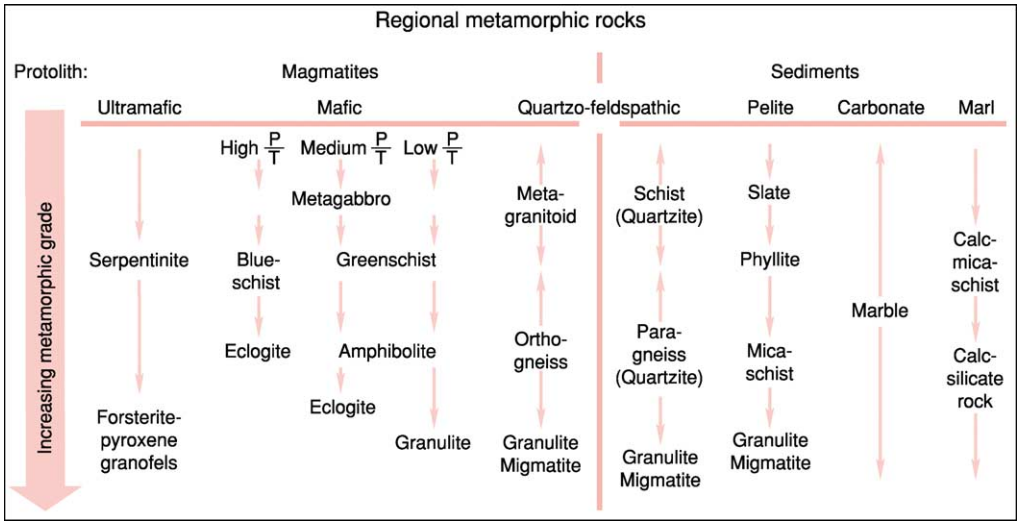


Figure 6 Chart showing the metamorphic rock name in relation with protolith and metamorphic grade.

by a blast that consists of sigmoidal, spiral shaped inclusion patterns, that developed during rotational growth by shearing along the schistosity planes; (E) Post-deformational growth is seen in discordant growth of kyanite and biotite which may contain inclusion trails parallel to the external schistosity ($S_i = S_e$); (F) Orthopyroxene forms a corona structure around garnet and ilmenite; and (G) Symplectite formation of diopside and plagioclase after omphacite.

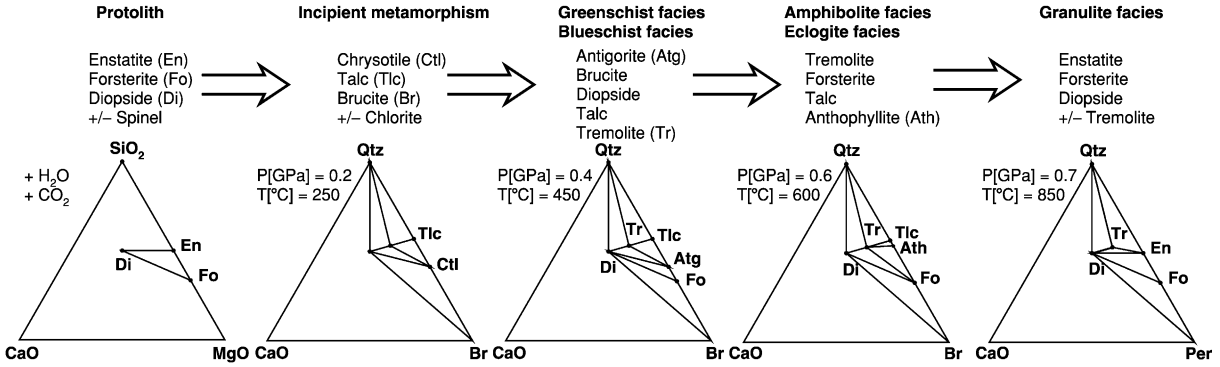


Figure 7 For graphical presentation of mineral assemblages and chemical compositions of ultramafic rocks, the ternary system SiO_2 – M_gO – CaO with H_2O and CO_2 in excess is used. The significant mineral assemblages for ultramafic rocks at different metamorphic grades are shown here.

chlorite in ultramafic rocks up to the amphibolite–granulite-facies transition where it is replaced by spinel (Figure 7).

Mafic Rocks

Basaltic or gabbroic protoliths are usually anhydrous and must first be hydrated and carbonated at low temperatures (by ocean floor metamorphism) to allow prograde metamorphic reactions to proceed. A typical low-T assemblage is albite + chlorite + carbonates. Compared to other protoliths, mafic lithologies are particularly sensitive to very low-grade metamorphic conditions. Incipient metamorphism is recognized by the occurrence of various zeolite minerals (i.e., laumontite). The subsequent prograde assemblages depend strongly on the geothermal gradient. Index minerals along a path with a high P/T-ratio (Figure 2) are pumpellyite and lawsonite (stable up to ca. 0.8 Gpa at 300°C), glaucophane (stable up to ca. 1.6 GPa at 500°C), and omphacite. Subgreenschist-facies metamorphism along a medium P/T-gradient is recognized by prehnite (at pressures below 0.3 GPa) or pumpellyite at higher pressures. The transition to greenschist-facies assemblages occurs at about 280°C, due to the breakdown of prehnite and pumpellyite to epidote and actinolite. Together with the matrix minerals albite+chlorite+carbonates, they form the characteristic assemblage of a greenschist. The amphibolite-facies assemblage plagioclase+hornblende grows at ~550°C from chlorite+epidote. In course of this reaction garnet may be formed. White micas, particularly paragonite, as well as kyanite coexisting with hornblende, are indicative of pressure-dominated amphibolite-facies conditions (epidote–amphibolite-facies). Additionally to the change in the ACF-topology, the transition from greenschist- to amphibolite-assemblages is also marked by a significant change in the chemical composition of feldspars from albite to oligoclase–andesine and amphiboles from actinolite to Tschermakitic hornblende. Eclogite-facies conditions are reached when plagioclase is replaced by omphacite and quartz. The lower pressure limit of the amphibolite/eclogite transition is *ca.* 1.2 GPa at 600°C. In metamafics metamorphosed along low P/T-gradients (Figure 2), the transition to granulite-facies conditions occurs between 700–800°C. Index mineral phases of mafic granulites are clinopyroxene and orthopyroxene, which form at the expense of hornblende (Figure 8).

Quartzo-Feldspathic Rocks

For the chemographic presentation of mineral assemblages, the AKF-system is applied (Figure 9). The chemical composition of quartzo-feldspathic rocks plot in the Al-poor part of the triangle and therefore these compositions contain few and not easily diagnosed assemblages. Contrary to clastic sediments, magmatites need to be hydrated to make metamorphic reactions possible. However, granitoids often remain in a dry state preventing the rock from completely metamorphosing. Otherwise, incipient metamorphism is demonstrated by phengitic muscovite, chlorite±stilpnomelane. Stilpnomelane is only stable in lower greenschist- and blueschist-facies conditions whereas muscovite remains stable up to about 650–700°C, then breaking down to K-feldspar, quartz, and an Al_2SiO_5-modification. However, the chemical composition of muscovite changes significantly with T and particularly with *P* and may be used as a geobarometer. In the assemblage K-feldspar–biotite–muscovite–quartz, the phengite content of muscovite expressed as Si per formula unit decreases with temperature but increases with pressure due to substitution of the Tschermak-molecule $Mg_{+1}Si_{+1}Al_{-1}Al_{-1}$. Metamorphism along a medium P/T-gradient leads to only a few metamorphic minerals, such as almandine-rich garnet at lower amphibolite facies

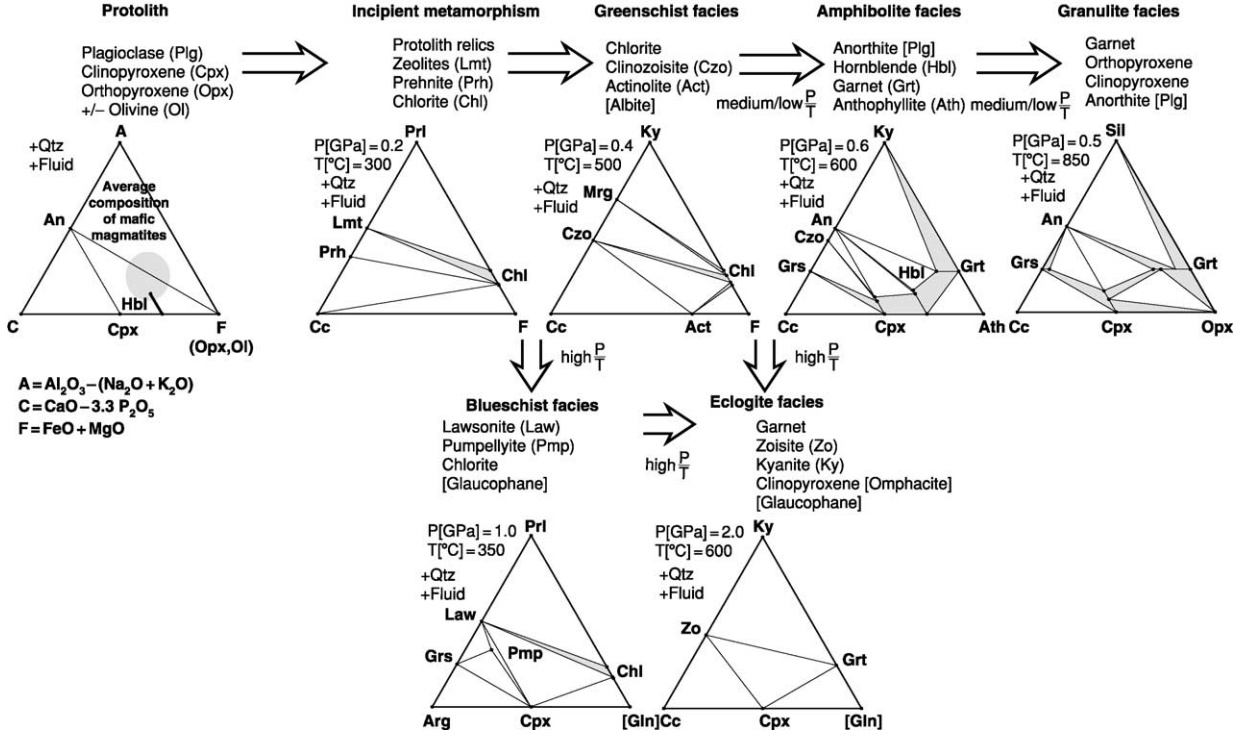


Figure 8 Al_2O_3 (A) is an important component in mafic rocks besides CaO (C) and FeO + MgO (F), hence the ACF-projection is commonly used for the graphical representation of mineral assemblages and chemical compositions of metamafics. In order to include silicates, hydrates, and carbonates in the ACF-projection SiO_2, H_2O and CO_2 are excess components. The chemographic projections show the significant mineral assemblages for mafic rocks at different metamorphic grades.

conditions – and partial melting takes place at the temperatures of the granite minimum in H_2O-saturated conditions (*ca.* 650°C), resulting in the formation of migmatites. In water-feel conditions, orthopyroxene may form at the expense of biotite at granulite-facies conditions (*ca.* 750°C). In quartzo-feldspathic rocks, which have undergone eclogite-facies metamorphism, plagioclase is usually replaced by jadeite, zoisite, and quartz and biotite by garnet and phengite.

Pelitic Rocks

Pelites are rich in sedimentary water ocurring as H_2O-molecules and as lattice-bound OH-groups. This enables mineral reactions to take place at very low temperatures. The mineral assemblage prior to metamorphism comprises mainly detrital clay minerals in addition to relict quartz, feldspars, and micas. Additionally authigenic Fe-(hydr)oxides, sulphates, and organic carbonaceous material are common. Clay minerals show significant structural changes already in diagenetic conditions, leading mainly to an improvement of lattice ordering. At the transition to metamorphism, white mica and chlorite are formed from clay minerals at about 200°C. These rocks typically have a slaty cleavage due to compaction and parallel orientation of the sheet silicates. Incipient metamorphism is clearly demonstrated by the production of pyrophyllite from kaolinite at about 300°C. In chlorite-rich assemblages, pyrophyllite may in turn be replaced by chloritoid and the terminal breakdown of pyrophyllite to kyanite occurs at *ca.* 400°C. Approximately at this temperature, biotite appears for the first time from K-feldspar and chlorite. In the upper greenschist facies, first garnet appears in an assemblage with biotite at the expense of chlorite and muscovite. At *ca.* 500°C, the terminal reaction of chloritoid leads to the formation of staurolite and garnet. This marks the transition to the amphibolite facies. Staurolite and biotite, formed at the expense of chlorite, garnet, and muscovite, constitute an important coexisting assemblage in the medium grade amphibolite facies at *ca.* 600°C. At the same temperature, paragonite is replaced in quartz-bearing assemblages by albite and an Al_2SiO_5-modification (typically kyanite). At about 700°C, staurolite disappears in quartz-bearing assemblages with the formation of the assemblage sillimanite, kyanite, and garnet, whereas muscovite and quartz are replaced by K-feldspar, sillimanite, or kyanite and melt. Hence metapelites of the upper amphibolite facies are poor in phyllosilicates (the surviving mica

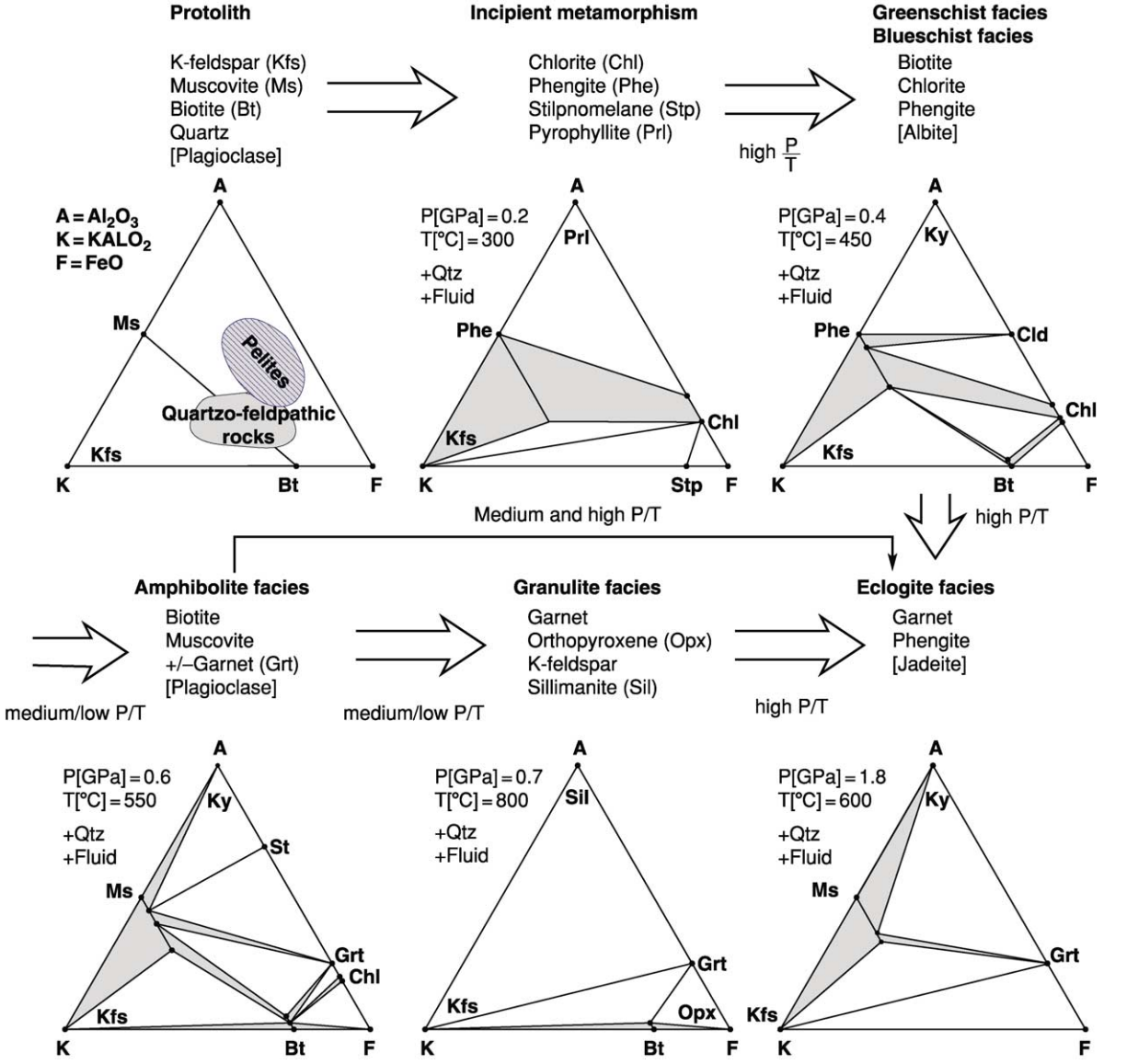


Figure 9 For the chemographic presentation of quartzo-feldspathic rocks, the AKF-system is applied. The difference to the ACF-projection is the consideration of K instead of Ca. Again feldspar, quartz, and H_2O are used as projection phases. K-feldspar, muscovite, biotite, and stilpnomelane are the only four important potassium-bearing minerals, besides the K-free Fe-Mg-Al-phases plotting at the AF-line. The chemical composition of quartzo-feldspathic rocks plot in the lower Al-poor part of the triangle and therefore these lithologies contain only few and not diagnostic assemblages. The AKF chemographic projections show the significant mineral assemblages at different metamorphic grades.

being biotite) but rich in feldspars, garnet, and sillimanite or kyanite. At low pressures, garnet and sillimanite may be replaced by cordierite in the amphibolite facies. In the upper amphibolite-facies, also an orthoamphibole (anthophyllite, gedrite) may form together with cordierite. At *ca.* 750–800°C, the formation of orthopyroxene from orthoamphibole, or from remaining biotite in quartz-bearing rocks, marks the transition to the granulite facies. At still higher temperatures, between 850 and 950°C, spinel, sapphirine, and ossumilite are diagnostic in quartz-bearing metapelites. Eclogite-facies metamorphism of metapelites can best be recognized in MgO-rich bulk chemical compositions, where talc and kyanite form the diagnostic 'white schist'-assemblage. High *P/T* gradients are also indicated by carpholite at lower greenschist facies and talc–phengite assemblages at higher temperatures (Figure 10).

Carbonate Rocks

Pure sedimentary carbonate rocks consist of either calcite, calcite and dolomite, or dolomite and have a

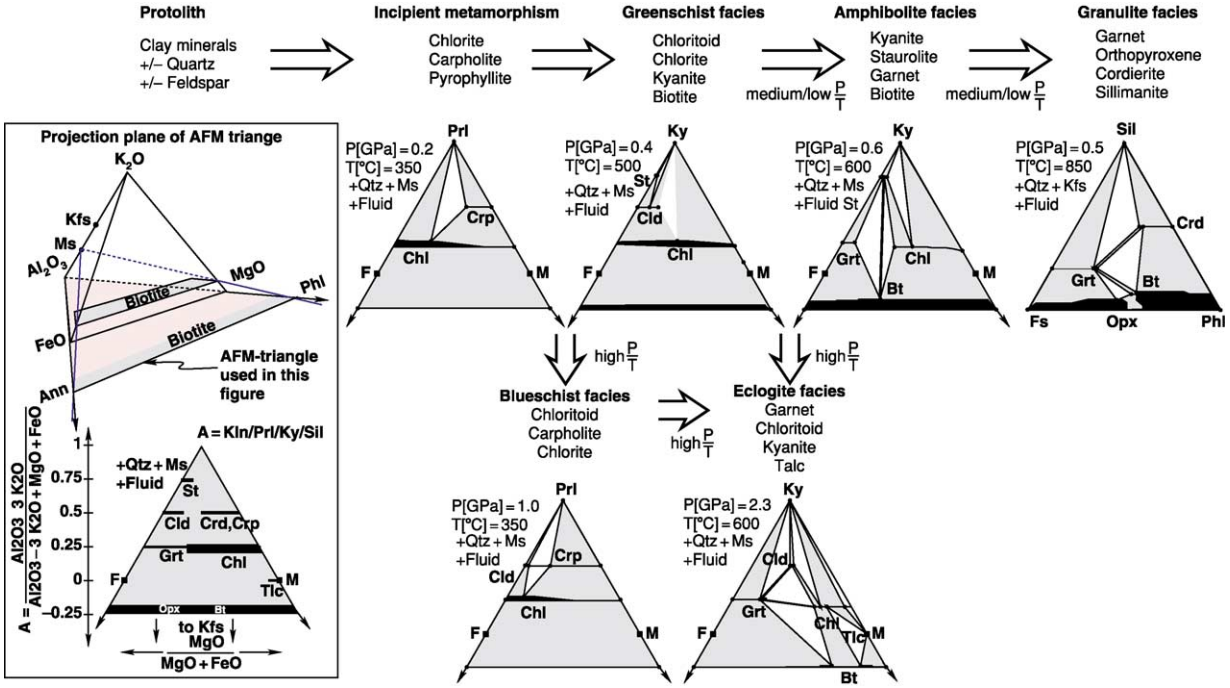


Figure 10 The chemical system applied to metapelites is identical with the AKF-system. However, an essential difference has to be taken into account: pelites plot in the Al-rich part of the AKF-triangle (**Figure 9**) and are mainly capable of producing Fe-Mg-Al-silicates plotted at the AF line. This rock composition probably contains the most key-(or index-) minerals with changing metamorphic grade. In order to show the reaction sequence in terms of the Fe-Mg-Al-silicates, it is necessary to consider FeO and MgO as separate components. Thus these four components build up an AKFM-tetrahedron and for practical use the AFM-plane (projected from muscovite or K-feldspar) is used for graphical presentation of the mineral phases.

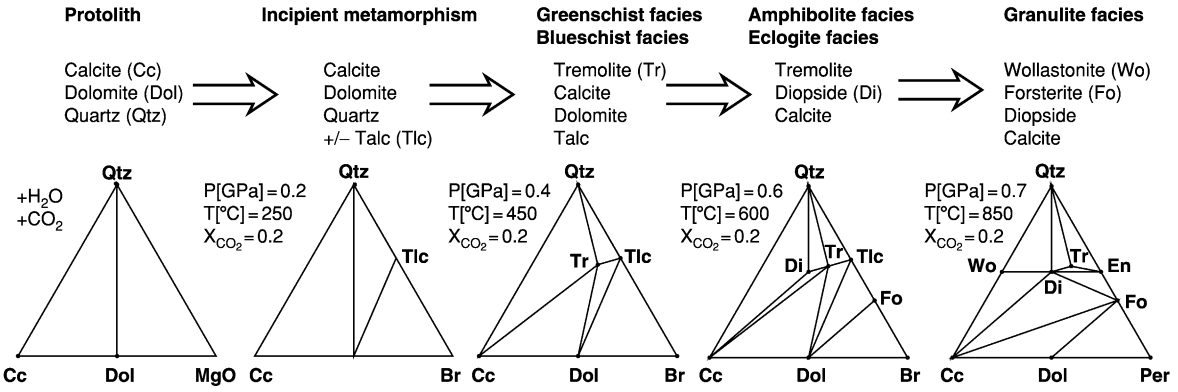


Figure 11 Possible mineral assemblages in metacarbonates with increasing metamorphic grade.

H_2O-rich pore fluid. Quartz has to be added to enable mineral reactions of the carbonates. Hence the relevant chemical system comprises CaO–MgO–SiO_2–CO_2–H_2O. (**Figure 11**). The temperature of incoming or breakdown of diagnostic assemblages in the carbonate system is dependent on the fluid composition (X_{CO2}) and a T-X_{CO2} plot is usually applied to show this effect (**Figure 4**). The prograde sequence of metamorphic index minerals in calcite–dolomite–quartz marbles affected by an intermediate *P/T*-gradient is generally: talc, tremolite, diopside, forsterite, and wollastonite. The maximum transition-temperature from talc+calcite to tremolite is about 500°C. The maximum stability of tremolite+calcite is *ca*. 700°C. The first appearance of diopside is at *ca*. 600°C when it is replacing the mineral assemblage tremolite + calcite + quartz. At *ca*. 800°C, diopside + dolomite breaks down to forsterite. At low P/T-gradients,

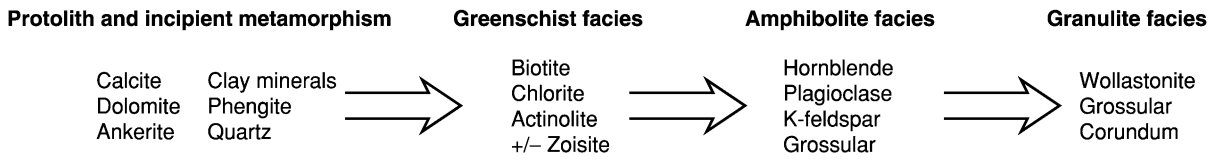


Figure 12 Possible mineral assemblages in metamarls with increasing metamorphic grade.

typical of contact metamorphism, these temperatures are reduced by about 100°C. The regular sequence observed all over the world of index minerals in metacarbonates is indicative of an internal buffering of the fluid composition during prograde metamorphism. An exception is probably the formation of wollastonite from calcite + quartz, or periclase from dolomite at high grade and low X_{CO_2}.

Marls

These rocks represent lithological and chemical mixtures between carbonate rocks and pelites, resulting in a rather complicated chemical system resembling mafic lithologies. Hence, in meta-marls, most minerals occurring in metapelites, metacarbonates, and metamafics may be present. For simplicity, meta-marls are considered here as contaminated metacarbonates where additional K + Na, Al, and Fe lead to several additional mineral phases such as plagioclase, zoisite/epidote, garnet, margarite, paragonite, muscovite, biotite, chlorite, and K-feldspar. Since carbonates and OH-bearing silicates occur, the diagnostic mineral assemblages depend on X_{CO_2}. Starting with an assemblage of ankeritic carbonates, white mica, and quartz, the mineralogy changes to biotite and chlorite at the greenschist–amphibolite facies transition (500°C). At lower amphibolite facies conditions, hornblende, plagioclase, and K-feldspar are diagnostic minerals and occur together with calcite and quartz. Hornblende, calcite, and quartz react to diopside at higher amphibolite facies conditions. Zoisite replaces plagioclase and calcite at very low X_{CO_2} at upper greenschist-to lower amphibolite-facies conditions. At slightly higher temperatures but the same low X_{CO_2}-values, plagioclase, calcite, and quartz are replaced by grossular-rich garnet (Figure 12).

See Also

Impact Structures. **Metamorphic Rocks:** Facies and Zones; PTt - Paths. **Minerals:** Definition and Classification. **Regional Metamorphism**. **Shock Metamorphism**. **Thermal Metamorphism**. **Ultra High Pressure Metamorphism**.

Further Reading

Árkai P, Sassi FP, and Desmons J (2002) *Towards a unified nomenclature in metamorphic petrology: Very low-grade to low-grade metamorphic rocks*. A proposal on behalf of the IUGS Sucommission on the Systematics of Metamorphic Rocks. http://www.bgs.ac.uk/SCMR/scmr_products.html.

Blatt H and Tracy RJ (1996) *Petrology: Igneous, Sedimentary, and Metamorphic*, 2nd edn. WH Freeman and Company.

Brodie KH, Fettes D, Harte B, and Schmid R (2002) *Towards a unified nomenclature in metamorphic petrology: Structural terms, including fault rocks*. A proposal on behalf of the IUGS Sucommission on the Systematics of Metamorphic Rocks. http://www.bgs.ac.uk/SCMR/scmr_products.html.

Bucher K and Frey M (2002) *Petrogenesis of Metamorphic Rocks*. Berlin, Heidelberg, New York: Springer Verlag.

Desmons J, Smulikowski W, and Schmid R (2002) *Towards a unified nomenclature in metamorphic petrology: High-P/T rock terms*. A proposal on behalf of the IUGS Subcommission on the Systematics of Metamorphic Rocks. http://www.bgs.ac.uk/SCMR/scmr_products.html.

Dudek A, Coutinh JMV, Desmons J, *et al.* (2002) *Towards a unified nomenclature in metamorphic petrology: Amphibolite and granulite*. A proposal on behalf of the IUGS Sucommission on the Systematics of Metamorphic Rocks. http://www.bgs.ac.uk/SCMR/scmr_products.html .

Le Maitre RW (ed.) (2002) *Igneous Rocks. A Classification and Glossary of Terms*. Cambridge: Cambridge University Press.

Philpotts AR (1990) *Principles of Igneous and Metamorphic Petrology*, 1st edn. London: Prentice Hall.

Rosen OM, Desmons J, and Fettes D (2003) *Towards a unified nomenclature of metamorphism: Carbonate and related rocks*. A proposal on behalf of the IUGS Subcommission on the Systematics of Metamorphic Rocks. http://www.bgs.ac.uk/SCMR/scmr_products.html.

Schmid R, Fettes D, Harte B, Davis E, Desmons J, and Siivola J (2001) *Towards a unified nomenclature in metamorphic petrology: How to name a metamorphic rock*. A proposal on behalf of the IUGS Sucommission on the Systematics of Metamorphic Rocks. http://www.bgs.ac.uk/SCMR/scmr_products.html.

Smulikowski W, Desmons J, Harte B, Sassi FP, and Schmid R (2003) *Towards a unified nomenclature in metamorphic petrology: Types, grade and facies*. A proposal on behalf of the IUGS Subcommission on the Systematics of Metamorphic Rocks. http://www.bgs.ac.uk/SCMR/scmr_products.html.

Stöffler D and Grieve RAF (2003) *Towards a unified nomenclature of metamorphism: Impactites*. A proposal on behalf of the IUGS Subcommission on the Systematics of Metamorphic Rocks. http://www.bgs.ac.uk/SCMR/scmr_products.html.

Wimmenauer W and Bryhni I (2002) *Towards a unified nomenclature of metamorphism: Migmatites and related rocks*. A proposal on behalf of the IUGS Subcommission on the Systematics of Metamorphic Rocks. http://www.bgs.ac.uk/SCMR/scmr_products.html.

Winter JD (2001) *An Introduction to Igneous and Metamorphic Petrology*, 1st edn. London: Prentice Hall.

Yardley BWD (1996) *An Introduction to Metamorphic Petrology*. Addison-Wesley Publication Company.

Facies and Zones

K Bucher, University of Freiburg, Freiburg, Germany

Intensity of Metamorphism

The intensity of metamorphism is mainly related to the pressure and temperature conditions prevailing during rock transformation. A rock recrystallized at 800°C in the lower crust contains characteristic minerals or mineral assemblages that reflect the high PT conditions under which the rock formed. In contrast, a shale in the shallow upper crust equilibrates perhaps at 300°C and is said to be weakly metamorphosed. It is of prime interest in geology to characterize the intensity of metamorphism and to use a reliable classification system to compare it within and between metamorphic terrains. The metamorphic facies system is such a tool for characterizing the intensity or grade of metamorphism (*see* **Metamorphic Rocks:** Classification, Nomenclature and Formation).

The distinctive grade of metamorphism of rocks from an outcrop can be evaluated and described by a large variety of techniques and methods. Because pressure and temperature are the main variables that control metamorphism, one could simple determine P and T for each collected sample by using methods of geological thermobarometry. This is, however, a tedious and costly affair and not necessary, except for special research studies. The metamorphic facies system is a much simpler system that can be used in the field and normally does not require laboratory work or even thin section observation.

The Metamorphic Facies Concept

The concept of metamorphic facies, as it is used today, is simple in principle. It makes use of the fact, that at a given pressure and temperature the chemical components of a rock are distributed among a unique set of minerals with fixed compositions. This is an inevitable consequence of chemical thermodynamics, provided the rocks reach a state of equilibrium. The unique sets of minerals that form in a reference material at one P and T can be compared with assembages of minerals that formed in other rock compositions under the same P and T. The collective sets of mineral assemblages that form under the same conditions in many kinds of rocks are called a metamorphic facies.

The reference material that is used in the metamorphic facies scheme is tholeiitic basalt (mid-ocean ridge basalt or MORB) because of its well-defined composition and global abundance. Mafic rocks of basaltic composition are extremely widespread and metamorphosed basalts (metabasalt) constitute a major rock type in orogenic belts, the oceanic lithosphere, and continental crust.

For example, at an outcrop one finds a mafic rock consisting of hornblende (Hbl) and plagioclase (Pl, e.g., andesine), that is termed an amphibolite. A metapelitic rock exposed at the same outcrop consists of kyanite, staurolite, biotite, muscovite, and quartz, and an associated band of dolomite marble contains dolomite, calcite, tremolite, and diopside. All three kinds of rocks with very diverse compositions have been metamorphosed under the same pressure and temperature conditions. They are of the same metamorphic grade.

At this outcrop, the reference material consists of Hbl + Pl and hence the set of rocks with their characteristic mineral assemblages each belong to the same metamorphic facies. The facies name could be the hornblende-plagioclase facies. For simplicity and because Hbl + Pl rocks are amphibolites, the name amphibolite facies is used for conditions under which Hbl and Pl are formed in the reference material, namely MOR basalt.

The kyanite-staurolite micaschist at the same outcrop has also been metamorphosed under amphibolite facies conditions, i.e., where basalt is transformed into amphibolite, and the micaschist is also an amphibolite facies rock. However, it contains no hornblende and no andesine. Similarly, the

tremolite-diopside marble belongs to the same amphibolite facies.

At another outcrop some 20 km from the first, one finds micaschist but no amphibolite and no marble. The micaschist contains kyanite, staurolite, biotite, muscovite, and quartz, i.e., the same assemblage as at the first outcrop. Consequently, the micaschist at this second outcrop is also an amphibolite facies rock, metamorphosed under conditions where hornblende and andesine would have formed in metabasalt if this rock were present.

Metamorphic Facies Systems

The example above illustrates the practical use and the concept of metamorphic facies. The amphibolite facies is the total of all pressure and temperature (P-T) conditions under which metabasalt would consist predominantly of hornblende and plagioclase. These PT conditions can be broadly outlined on a pressure versus temperature diagram (Figure 1).

In a next step, a series of characteristic metabasalt mineral assemblages can be used to define additional facies to the amphibolite facies described above. In fact, one can select typical assemblages in metabasalts that form at any P-T conditions rocks can be subjected to on Earth.

A decision has to be made, however, how detailed the assemblages should be for use in the metamorphic facies system. First we use, as in the example above, just two characteristic minerals, a mineral pair, occurring in metabasalt to define a series of additional facies. Other versions and additional complexities will be introduced later.

Amphibolite Facies

Mafic rocks contain hornblende and plagioclase. Hornblende is an aluminous amphibole with a composition that contains predominantly a pargasite component. Hornblende is dark in hand specimen and green to brown under the microscope. Plagioclase contains significant amounts of an anorthite component. Oligoclase is the minimum requirement, typical is andesine or labradorite.

Greenschist Facies

In the greenschist facies mafic rocks contain actinolite and albite. Actinolite is an Al-poor amphibole; it contains mainly tremolite. Actinolite is green in hand specimen and colourless to pale green under the microscope. Plagioclase is pure albite and contains no Ca. The two minerals are the major constituents of greenschists. The greenschist facies represents lower P-T conditions than the amphibolite facies, i.e., a lower metamorphic grade. Greenschist and amphibolite facies rocks form in typical regional metamorphism related to a geodynamic setting of continental collision tectonics (Figure 2). In many orogenic belts worldwide, metamorphic rocks of the greenschist and amphibolite facies predominate over all other metamorphic facies rocks.

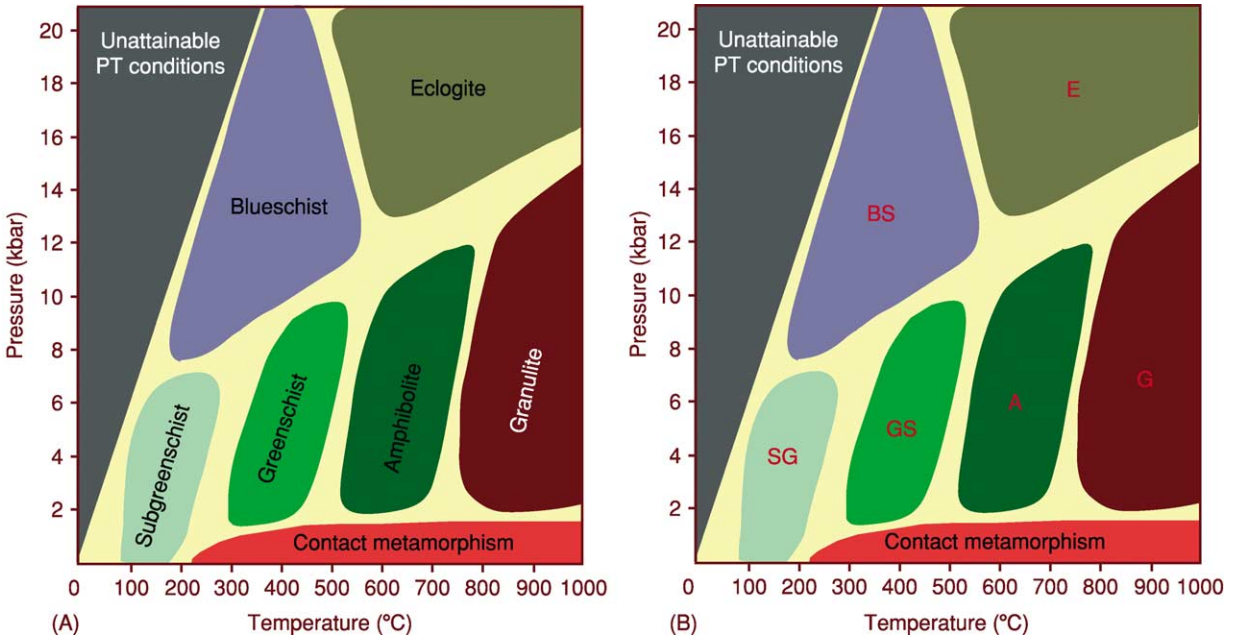


Figure 1 (A) Metamorphic facies scheme, general version (full names). (B) Metamorphic facies scheme, general version (abbreviation). SG sub-greenschist, GS greenschist, A amphibolite, G granulite, BS blueschist, E eclogite.

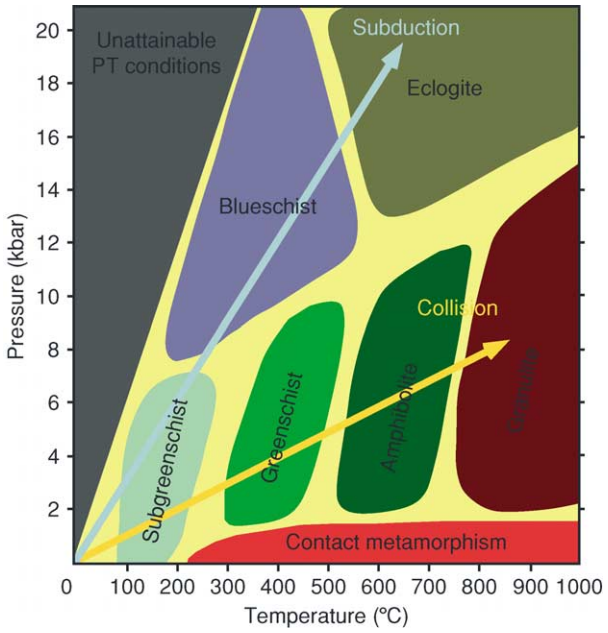


Figure 2 Metamorphic facies scheme, general version (full names). ''subduction'' = path of prograde metamorphism during subduction of oceanic lithosphere; ''collision'' = path of prograde metamorphism during continental collision.

Blueschist Facies

Mafic rocks contain glaucophane and epidote in the bluschist facies. Glaucophane is an alkali-amphibole that is dark or blue in hand specimen and colourless to blue under the microscope. Plagioclase is not normally present, but can occur as albite. Epidote or clinozoisite occurs in abundance and carries much of the calcium of the basaltic precursor rock. The two minerals are major constituents of blueschists, hence blueschist facies represent high-P low-T conditions typical of subduction zone metamorphism (Figure 2). These conditions are normally not compared to the greenschist or the amphibolite facies, in the sense of lower or higher grade rocks, because blueschists reflect metamorphism in a different geodynamic context.

Eclogite Facies

Some mafic rocks contain garnet and omphacite and are termed eclogites, hence eclogite facies. Eclogite garnet is a ternary mixture of almandine, pyrope, and grossular. Omphacite is an calcalkali pyroxene made of jadeite, diopside, and hedenbergite components. Omphacite is green in hand specimen and colourless to green under the microscope. Plagioclase is absent. Eclogite facies rocks form at high pressure in diverse geodynamic settings, ranging from subduction of oceanic lithosphere, to the formation of thickened continental crust during continent collision. Eclogites also are present at depth in stable anorogenic settings. If eclogites are formed by continuous prograde metamorphism, they are always the highest grade rocks. Although eclogite facies rocks are widespread in some orogenic belts, they typically occur in small isolated outcrops in a matrix of surrounding rocks that reflect conditions of lower grade facies.

Granulite Facies

Metamorphic mafic rocks that contain orthopyroxene (Opx) and plagioclase belong to the granulite facies. Orthopyroxene is a mixture of enstatite and ferrosilite components and is also called hypersthene. It is dark with a brownish tint in hand specimen and colourless to reddish under the microscope. Plagioclase is calcic and a labradorite composition is typical. The two minerals are present in mafic granulites. However, the Opx + Pl pair is present in granulite facies mafic rocks only at relatively low pressure. The major mineral pair in all mafic granulites is clinopyroxene (Cpx) and plagioclase. Note that this assemblage is identical to that of the unmetamorphosed basalt, however, the composition of the minerals and the texture of the rocks are different. Hence, the Cpx + Pl pair that makes up the bulk of granulite facies mafic rocks is unfortunately not exclusive to that facies.

As mentioned, the high-pressure portion of the granulite facies P-T field (Figure 1) lacks the diagnostic assemblage Opx + Pl. How does one know then, that a suite of rocks reflects high-pressure granulite facies conditions? The defining assemblage in this case is Cpx + Pl + garnet. The Cpx in mafic granulites is an augite with a low jadeite content. This is a black mineral in hand specimen and greenish under the microscope. Garnet is a ternary mixture of almandine, pyrope, and grossular components much like those in eclogite.

The granulite facies represents PT conditions higher than that of the amphibolite facies, so compared with the amphibolite facies, it represents higher grade rocks. Granulite facies rocks occur in extensive terrains, particularly in Precambrian shield areas. They are rare in Phanerozoic orogenic belts and very rare in Alpine chains. The boundary of granulite terrains to surrounding amphibolite facies rocks is usually of a retrograde nature.

Zeolite Facies

If metamorphism is weaker than the conditions of the greenschist facies, mafic rocks may contain an abundance of minerals that could be used to define metamorphic facies. Typical greenschist facies rocks require about 300°C to form from basalt. Greenschist facies rocks are typically well equilibrated and relics

from the original igneous assemblage are rare. At conditions of very low-grade metamorphism this is typically not the case. At, for example, 150°C, a number of characteristic zeolite minerals may form in vesicles and fractures of basaltic lava but the transformation is incomplete and igneous minerals are still present, although often severely altered. The alteration of igneous plagioclase and clinopyroxene is a process of hydration and water must have access to the reaction sites. Alteration by hot aqueous solutions is known as hydrothermal alteration (*see* **Thermal Metamorphism**). It may take place at the sea-floor and then be termed oceanic metamorphism or it may represent the initial stages of metamorphism in continental settings. It is rarely pervasive. Hydrothermal alteration is always accompanied by chemical transport, deposition of new minerals in the pore space of rocks and thus is not isochemical. The zeolite assemblages formed in basaltic rocks do not completely replace the original rock and are characteristic of PT conditions at the Earth's surface to about 150–200°C (Figure 3).

Prehnite–Pumpellyite Facies

At higher temperatures, but still not hot enough for greenschists to form, two characteristic minerals may be found in mafic rocks: prehnite and pumpellyite. The two minerals rarely occur together in the same rock. However, mafic rocks with prehnite or pumpellyite are typical of metamorphic conditions between the zeolite facies and the greenschist facies (Figure 3). There is a considerable overlap of mineral assemblages of these two very low-grade metamorphic facies. Pumpellyite often occurs in metabasalts together with a number of different zeolites. Prehnite and zeolites are also not exclusive. In greenschist facies rocks, however, zeolites. prehnite, and pumpellyite are absent (line 1 in Figure 5).

Sub-Greenschist Facies

Some typical features of the zeolite and prehnite-pumpellyite facies include the persistence of igneous minerals, overlap of assemblages of the two facies, widespread occurrence of metastable assemblages, and non-equilibrium textures and assemblages. Because of these features it is better to characterise the rocks at grades below that of the greenschist facies as sub-greenschist facies in many metamorphic terrains. This would comprise all rocks and all processes that transform rocks below about 300°C. Sediment diagenesis is normally excluded from the facies scheme but all hydrothermally altered igneous and metamorphic rocks can be described by either terms, sub-greenschist facies or very-low-grade metamorphism (Figure 1).

Ultra-High-Pressure Metamorphism

At very high pressures, mafic rocks contain omphacite and garnet. This is the mineral pair that makes up eclogite and defines the eclogite facies (see above). Rocks from as deep as 250 km below the surface have been found at various localities. The corresponding pressure is on the order of 50–60 kbar. Such ultra-high-pressure (UHP) conditions are normally especially pointed out in rock descriptions and separated from normal eclogite facies rocks (Figure 4).

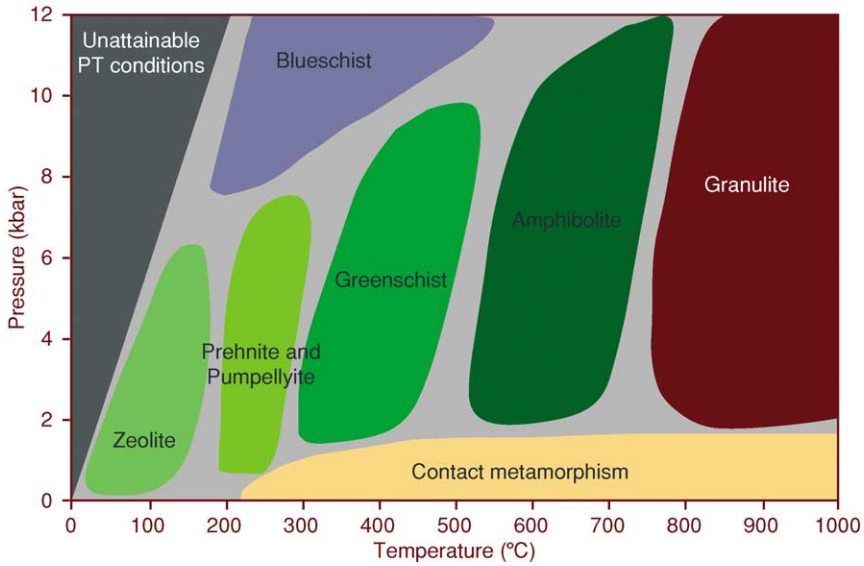


Figure 3 Metamorphic facies scheme, showing the PT fields of the zeolite and prehnite-pumpellyite facies.

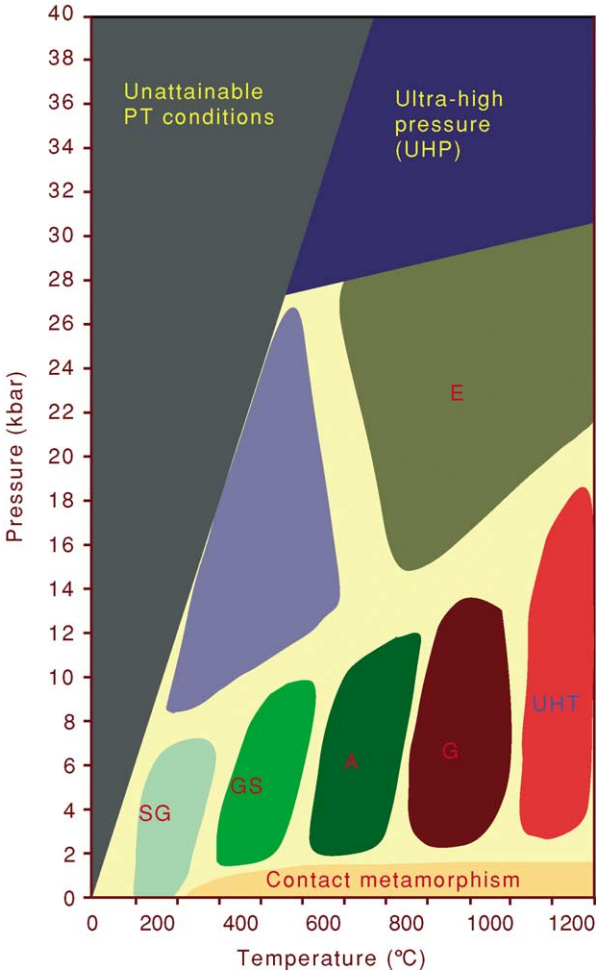


Figure 4 Metamorphic facies scheme, general version (abbreviations as in Fig. 1B). Extended P-T range showing conditions of UHP (ultra high pressure) and UHT (ultra high temperature) metamorphism.

The most diagnostic mineral of the UHP facies is coesite, indicating that UHP conditions are above the quartz-coesite transition. Coesite-eclogites are typical rocks of the UHP facies. One could also refer to a coesite facies. However, many UHP eclogites do not contain coesite. Among a number of other assemblages diagnostic of UHP facies conditions is talc + chloritoid.

Ultra-High-Temperature Metamorphism

The granulite facies includes all rocks that have been metamorphosed at high-grade conditions in the stability field of plagioclase (*see* **Ultra High Pressure Metamorphism**). Normally, granulites are associated with metamorphic temperatures in the range of 700–900°C. Increasing evidence shows that some rocks may have experienced metamorphic temperatures up to 1100°C. These ultra-high-temperature rocks may be classified as a separate facies, the ultra-high-temperature facies (UHT), analogous to the UHP facies (Figure 4).

Contact Metamorphism

Heating of rocks at low pressure in the shallow crust by intrusion of igneous rocks causes metamorphism that is restricted in areal extent, close to the igneous contact which is normally a pipe or subsurface magma chamber. The process is termed contact metamorphism and typical rocks are termed hornfels which are very fine grained metamorphic rocks resulting from rapid heating. It should be noted however, that contact metamorphic marbles can be coarse grained. The maximum temperature reached by the country rock aureole surrounding the igneous intrusion can be high and depends on the amount, nature, and composition of the magma. Basaltic magmas can reach temperatures in excess of 1200°C and contact rocks may be heated to 900°C or higher by gabbro intrusions.

In principle, contact metamorphic rocks can also be assigned to a metamorphic facies such as albite-epidote hornfels, hornblende hornfels, pyroxene hornfels, and sanidinite facies. Rocks of each of these facies would typically occur as successive zones of several tens of metres in thickness surrounding the igneous body. In the study of such contact aureoles, most researchers have used isograd mapping as the preferred tool to characterise of the metamorphic (= thermal) structure.

Volatile Components and the Role of Water

The metamorphic facies scheme relies on the premise that the stable mineral assemblage of a rock is entirely determined by pressure, temperature, and the composition of the rock. This is only true in a strict sense if the rock in question has had no chemical exchange with the surroundings, i.e., the metamorphism is isochemical. This is normally not important for the major components of a rock. However, it is a matter that has to be evaluated for volatile components in each case. The prime volatile component of rocks is H_2O. In prograde metamorphism, dehydration reactions typically create an H_2O pressure in rocks that equals the total pressure, that is H_2O is present as a free fluid phase. The presence of fluid is evidenced by the presence of ubiquitous fluid inclusions in minerals of metamorphic rocks. The rock controls its fluid phase at a given P and T.

In some geologic environments this is not the case. If rocks undergo prograde dehydration in regional metamorphism (*see* **Regional Metamorphism**), they

gradually loose their fluid phase. At a certain stage, the minerals are not in contact with a free aqueous fluid and under such circumstances hydrous minerals such as micas and amphiboles decompose at significantly lower temperatures compared with the case where free water would have been present. Consequently, high-grade mineral assemblages appear in rocks at lower grade. Lack of a free aqueous fluid is particularly typical of granulite facies rocks. Thus, the amphibolite–granulite facies transition (line 3, in Figure 5) is not only a PT boundary but is also related to decreasing H_2O pressure in the rocks. Also, in very low-grade rocks, H_2O may not pervasively wet all mineral grain boundaries in primary igneous rocks as outlined above (*see* **Metamorphic Rocks: PTt - Paths**). As a consequence, very high-grade assemblages may be found at ambient conditions, a fortunate circumstance that permits survival of the high-grade rocks at conditions of the earth surface.

Another aspect of volatile components and their effect on the facies scheme is the frequent occurrence of CO_2 and other gaseous components in metamorphic fluids. In sedimentary sequences that undergo metamorphism, carbonate-bearing strata are very common. When heated, these rocks produce CO_2 in addition to H_2O. No problems arise if the fluid produced remains in the rock that generates it. However, if the CO_2 producer exports its fluid to carbonate-free rocks, hydrous minerals may decompose at lower grade than if free pure H_2O would have been present.

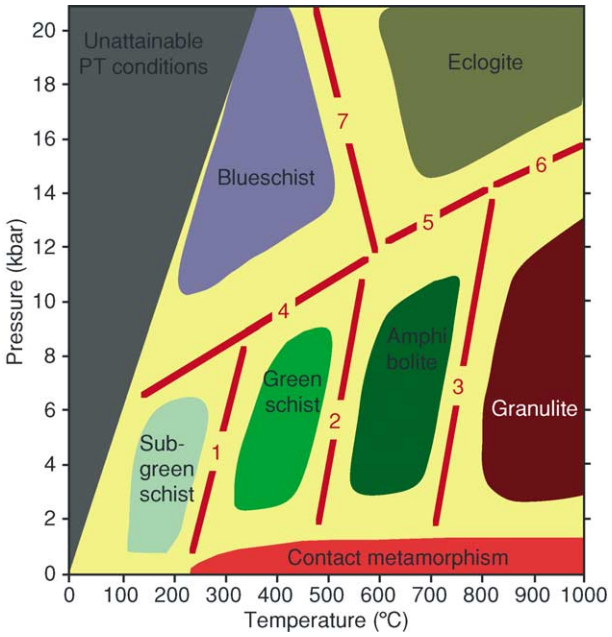


Figure 5 Metamorphic facies scheme, general version (full names) showing numbered facies boundaries as discussed in the text.

The effect on facies boundaries is similar to those cases lacking a fluid phase and it typically affects the amphibolite–granulite facies boundary (line 3, in Figure 5). The consequence of this fluid effect is that at the same temperature and pressure, granulite and amphibolie facies rocks may occur side by side depending on the H_2O pressure (or the chemical potential of H_2O in the rock).

Partial Melting and Igneous Processes

The metamorphic facies scheme (Figures 1 and 2) requires the presence of a free aqueous fluid. The facies figures have been prepared with this premise. Above, some exceptions from this requirement have been given with respect to the granulite and subgreenschist facies. An additional complication arises from the fact that in the presence of free H_2O-fluid, mafic rocks begin to melt at higher temperatures. The minimum melt temperature for metabasalts can be as low as 700°C in the middle crust. The onset of partial melting and the presence of a silicate melt at higher metamorphic grades has several consequences for the facies concept and the characterization of metamorphic grade.

H_2O and other volatiles dissolve in the silicate melt or may be expelled from the melt. The fine details of the melt-producing process has, therefore, consequences for the requirement of a free H_2O fluid in rocks. At conditions of the granulite facies, the amount of melt present depends on the H_2O pressure. The absence of migmatites (partially melted rocks) in many granulite facies terrains is strong evidence of low water pressure during metamorphism.

At conditions of the UHT facies, melt is produced even in the absence of H_2O. The absolute upper end of crustal metamorphism is given by 'dry' melting of granites (about 1150°C).

Melt produced in metamorphic rocks has the important implication that a once homogeneous precursor rock separates into a melt phase and a residuum (or restite) of different composition. Thus, isochemical conditions, the requisite basis for comparing metamorphic rocks from different grades, is no longer maintained in rocks that undergo partial melting. One has to be aware of these complications when working in migmatite areas.

Isofacial and Allofacial Conditions

It is commonly observed in metamorphic terrains that certain outcrops are of rather different metamorphic grade compared with all other outcrops in the area. Rocks belonging to very different facies may occur side by side along a road cut and the variation in

grade cannot be related to a continuous process of metamorphism that has transformed all rocks of the area. A good example of this is found along the west coast of Norway. There are a large number of 100 m-sized lenses and blocks of eclogite or of UHP rocks (coesite eclogites) in an amphibolite facies gneiss terrain. No relic assemblages of the eclogite facies are present in the gneiss. The eclogite blocks have tectonic contacts with the gneiss and are characterised as allofacial relative to them. Other examples are outcrops of garnet peridotite in the Central Swiss Alps that occur in amphibolite facies rocks. These UHP rocks are also clearly allofacial.

An the other hand, serpentinites of the Zermatt-Saas ophiolite complex of the Central Swiss Alps occur together with metabasalts and metagabbros of the eclogite and blueschist facies. The serpentinites represent hydrated mantle of the oceanic lithosphere and their assemblages antigorite + forsterite + diopside is isofacial with the eclogite facies metabasites.

The Metamorphic Facies Scheme, Additional Aspects and Facies Boundaries

Features of the amphibolite–granulite facies boundary (line 3, in Figure 5) and the lower limit of the greenschist facies (line 1, in Figure 5) have been briefly discussed.

The greenschist–amphibolite facies boundary (line 2, in Figure 5) is characterized in mafic rocks by the relatively abrupt increase of the calcium (anorthite)-content of plagioclase and by a simultaneous increase of the tschermak- and edenite-content of amphibole. In metapelites, staurolite-bearing assemblages gradually replace chloritoid-bearing assemblages.

The boundaries 4, 5, and 6 in Figure 5 all mark the disappearance of plagioclase in mafic rocks towards higher pressures. At low temperature, albite-bearing mafic rocks are replaced by assemblages with sodic amphibole (line 4, in Figure 5). Sodium is transferred from feldspar to amphibole and mica (paragonite). At higher temperature, amphibolite and granulite are directly replaced by eclogite (lines 5 and 6, in Figure 5). The transition from blueschist to eclogite facies (line 7, in Figure 5) is a very gradual boundary. In typical low-temperature eclogite, omphacite and garnet often coexist with glaucophane and clinozoisite over a fairly wide range of conditions. This means that assemblages of both facies occur in the same metabasalt. This is mostly due to the strong dependence of the stable assemblages on the redox conditions prevailing during high-pressure metamorphism.

Mineral Zones

Characterizing the metamorphic grade can also be achieved by mapping in the field the occurrence of so called index minerals in a homogeneous rock unit with constant composition over a large outcrop area. Index minerals have a PT sensitive distribution in the rock type studied. In going up-grade, the first occurrence of an index mineral can be placed as a line on a map. The line connecting all outcrops of the first occurrence of an index mineral is called a zone boundary of a mineral zone that is characterized by that index mineral. The upper grade limit of the mineral zone is defined by the mineral zone boundary of the next index mineral with increasing grade.

The classic example for the use of this simple and straightforward technique is the map of mineral zones in metapelites (micaschists) in the Scottish Highlands (see Further Reading). At the lowest grade outcrops, the micaschists contain chlorite as a diagnostic mineral. The first occurrence of biotite in the micaschists can be mapped as a line in the field. It defines the zone boundary of the biotite zone. The next diagnostic mineral that appears in the rocks is garnet. Its first appearance defines the beginning of the garnet zone. At still higher grade staurolite appears in the schist and the staurolite zone boundary marks also the upper end of the garnet zone. Note that both biotite and garnet may still be present in the rocks. Continuing upgrade, kyanite can be found in staurolite-garnet-biotite micaschists, defining a kyanite mineral zone. Finally at the highest metamorphic grade but still within the amphibolite facies, sillimanite can be found in the rocks. The complete zonal sequence of index minerals is known as the Barrovian sequence of metamorphism. It is a typical pattern of metamorphism and it has been described from many orogenic belts.

See Also

Metamorphic Rocks: Classification, Nomenclature and Formation; PTt - Paths. **Regional Metamorphism**. **Thermal Metamorphism**. **Ultra High Pressure Metamorphism**.

Further Reading

Austrheim H (1990) The granulite-eclogite facies transition: A comparison of experimental work and a natural occurrence in the Bergen Arcs, western Norway. *Lithos* 25.

Barrow G (1912) On the geology of lower Deeside and the southern Highland border. *Proceedings of the Geologists Association* 23: 268–284.

Bucher K and Frey M (2002) *Petrogenesis of Metamorphic Rocks*. Berlin, Heidelberg: Spring-Verlag.

Coombs DS, Ellis AJ, Fyfe WS, and Taylor AM (1959) The zeolite facies, with comments on the interpretation of hydrothermal syntheses. *Geochimica et Cosmochimica Acta* 17: 53–107.

Eskola P (1921) On the eclogites of Norway. *Skrifter Vidensabelig Selskab. Christiania, Mat.-nat. Kl. I* 8: 1–118.

Evans BW and Brown EH (1986) *Blueschists and Eclogites*. Geological Society of America Memoir, 423 pp. Boulder, Colorado: The Geological Society of America.

Frost BR and Frost CD (1987) CO_2, melts, and granulite metamorphism. *Nature* 327: 503–506.

Newton RC (1985) Temperature, pressure and metamorphic fluid regimes in the amphibolite facies to granulite facies transition zones. In: Tobi AC and Touret JCR (eds.) *The Deep Proterozoic Crust in the North Atlantic Provinces*, pp. 75–104. Dordrecht: Reidel.

Pattison DRM (1991) Infiltration-driven dehydration and anatexis in granulite facies metagabbro, Grenville Province, Ontario, Canada. *Journal of Metamorphic Geology* 9: 315–332.

Pattison DRM (2003) Petrogenetic significance of orthopyroxene-free garnet + clinopyroxene + plagioclase ± quartz-bearing metabasites with respect to the amphibolite and granulite facies. *Journal of Metamorphic Geology* 21: 21–34.

Winter JD (2001) *An introduction to igneous and metamorphic petrology*. New Jersey: Prentice Hall.

PTt-Paths

P J O'Brien, Universität Potsdam, Potsdam, Germany

Introduction

A pressure–temperature–time (usually abbreviated as PTt) path is, very simply put, a record of the ups and downs in temperature and pressure experienced by a metamorphic rock during its lifetime. This information is cryptically recorded in the form of different minerals present in a rock, in their chemical and isotopic compositions and degree of chemical homogeneity, and in the fabrics (size, shape, distribution, and orientation) of the different minerals. In order to utilize PTt paths of rocks for interpreting Earth history it is necessary to understand the different processes causing temperature and pressure changes, how the rates of change of these processes differ due to different geological processes, and how this is then reflected in mineralogical and textural changes in rocks. Once these processes are understood and quantified, they can be used to model and predict the changes expected in metamorphic rocks as a result of a chosen tectonometamorphic scenario. The complementary situation is the reconstruction of the magnitude and duration of metamorphic processes based on PTt paths derived from natural rock samples that were 'eye-witnesses' to one or more tectonometamorphic events, i.e., rocks that underwent heating, cooling, burial, and exhumation as a result of tectonometamorphism. In contrast to theoretical PTt-path modelling, the reconstruction of PTt paths from actual samples requires determination of the equilibration pressure and temperature conditions for mineral assemblages, the age of these assemblages, and also how mineral compositions and assemblages have changed over time as a function of evolving pressure and temperature changes: factors all fraught with uncertainty and error. For this reason the information from natural samples is far from complete but it is information from natural samples that provide key constraints and thus allow fine-tuning of predictive tectonometamorphic models.

PTt Paths: The Basics

If a metamorphic rock preserves a perfect equilibrium mineral assemblage then its history can be reflected as a single point in a PT diagram (**Figure 1A**). The fact that it was most likely collected at the Earth's surface means that the mineral assemblage is metastable and also that the route from the depths corresponding to the PT point back to the surface is not recorded. If the age of the mineral assemblage can be determined then average cooling and exhumation rates relative to today's surface conditions are available. However, if a rock does not show equilibrium, then more than one PT point can be deduced and, assuming that the sequence of reactions or changes can be determined, a PT path can be constructed (**Figure 1B**). Disequilibrium is typically evidenced by frozen, incomplete reactions where both old and new phases are present; as compositional differences between matrix minerals and the same phases present as armoured inclusions in large porphyroblasts; as compositional zoning of minerals; or as multiple growth events for the same mineral. If the age of the individual metamorphic stages in the history can be determined then all three requirements for a PTt path are met, i.e., pressure, temperature, and time changes (**Figure 1C**). However, the very fact that different stages in the PT evolution

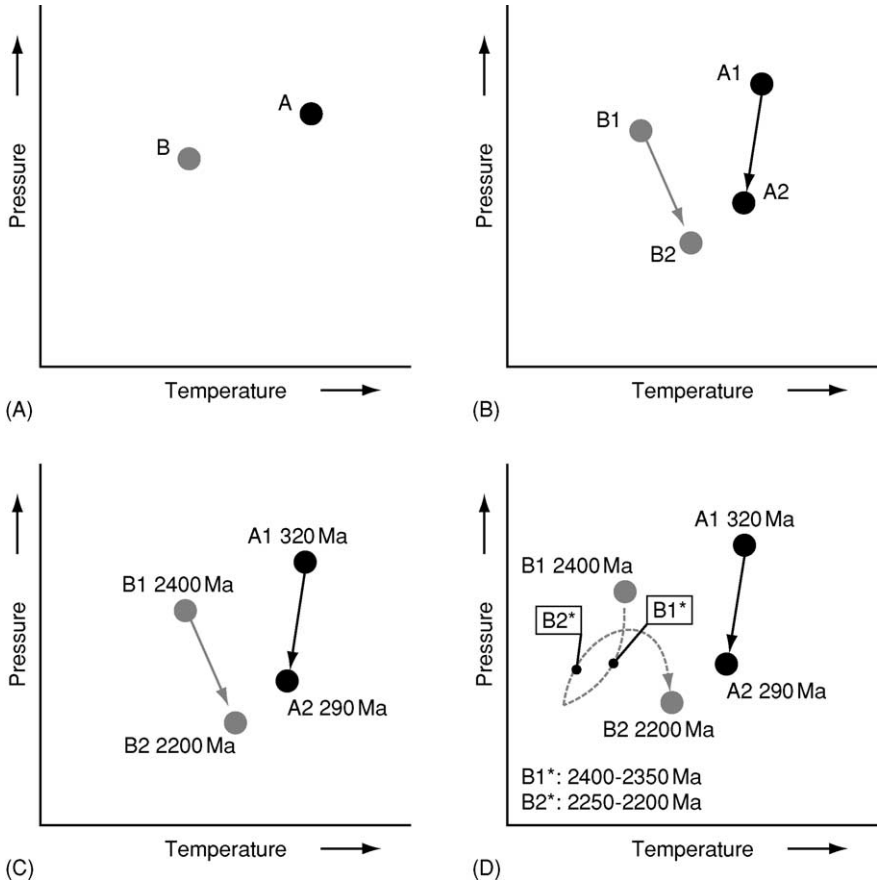


Figure 1 Schematic diagrams showing the relationship between PT point (A), PT path (B), and PTt path (C, D).

are determinable means that the rock is no longer in equilibrium, and thus the use of geothermobarometers based on equilibrium thermodynamics requires a significant degree of interpretation of textural and compositional variations. For samples that underwent several reaction stages at high temperature it is highly likely that compositional information has been modified or even lost from these earlier stages in the reaction history. Alternatively, it is by no means certain that all temperature and pressure changes of a rock's environment are actually recorded in textural and compositional information (Figure 1D). Thus the simple linking of different PT points to form a PTt path, although common practice, should not be accepted as the only solution.

The quantitative study of metamorphic rocks is essentially subdivided into three different but interrelated fields, namely formation, preservation, and exhumation. The PTt path is a historical record of these factors. The most simple factors controlling the formation of metamorphic rocks are changes in temperature and pressure. The interrelationship between rate of temperature change, rate of pressure change, and tectonic environment is the fundamental feature of a PTt path. The absolute pressure and temperature controls the thermodynamic driving force for reaction, i.e., which minerals should form and with which compositions. The actual formation and preservation of minerals is related to kinetic factors that control the rates of material transport processes, mineral breakdown or growth, and compositional homogenization. Exhumation of metamorphic rocks, leading to cooling and decompression, is a necessity if they are to be sampled at the Earth's surface. Processes such as surface erosion, tectonic extension, buoyancy, or a combination of these may all play a role and may also yield characteristic PTt-path segments. All three factors can be modelled for given tectonometamorphic scenarios so the input parameters for models, the shape of resultant PTt paths, and their relative importance will be outlined in turn.

Pressure–Temperature Controls

Pressure, as a geological parameter, is easy to understand. This is purely the average of the stresses applied to the rock. The main stress usually corresponds to

the force of the mass of rocks overlying the point of interest (lithostatic pressure) and can be determined as $P = \rho g h$ (where P = pressure in pascals, ρ = density of the overlying rock in kilograms per cubic metre, g = acceleration due to gravity = $9.81\,\mathrm{m\,s^{-2}}$, and h = depth of the point of interest in metres). Pressure in the geological literature is usually given in kilobars ($1\,\mathrm{bar} = 10^5\,\mathrm{Pa}$, $1\,\mathrm{kbar} = 10^8\,\mathrm{Pa}$), megapascals ($1\,\mathrm{Mpa} = 10^6\,\mathrm{Pa}$) or gigapascals ($1\,\mathrm{Gpa} = 10^9\,\mathrm{Pa}$). For typical rock densities (granite: $2700\,\mathrm{kg\,m^{-3}}$; basalt: $3000\,\mathrm{kg\,m^{-3}}$; peridotite: $3300\,\mathrm{kg\,m^{-3}}$), it takes a thickness of 3.8 km of granite, 3.4 km of basalt, or 3.1 km of peridotite to produce a pressure of 100 MPa. Alternatively, pressure increases at a rate of between 26.5 MPa (for granite), 29.4 MPa (for basalt), and 32.4 MPa (for peridotite) per kilometre depth. Generally, pressure deep in the Earth is said to be hydrostatic, i.e., equal in all directions, but in some situations pressures in the crust may be higher than that calculated for the depth alone due to the accumulation of directed horizontal stresses (nonhydrostatic or deviatoric stress) during collision. These stresses cannot build up to significant levels, especially at higher temperatures, because rocks will break or flow to reduce stress build-up, but they are very important for deformation and the production of oriented fabrics in rocks. This means that, within error, pressure can be correlated with depth.

In contrast, the temperature distribution in the crust is not so easy to estimate because it is necessary to know about heat sources, heat sinks, and heat transport as well as how these factors are affected during tectonometamorphism. Heat sources in the crust are either internal, external, or transported (advected). Within the crust the internal heat sources are the heat derived from the decay of radioactive elements, especially K, U, and Th, exothermic mineral reactions (typically retrograde reactions), crystallizing magmas (latent heat of crystallization), or deformation (frictional or shear heating). The radioactive elements do not fit well into the structure of mantle minerals—they are thus termed incompatible—and so over geological time they have been preferentially removed from the Earth's mantle, during melting episodes, and enriched in the continental crust, especially in its upper part. It is the distribution of these radioactive, heat-producing elements in the crust that plays an important role in the possible temperature–depth pattern in the crust before, during, and after a tectonometamorphic event. External, with respect to the crust, is heat added from the underlying mantle, whereas transported or advected heat is that introduced by intruded (magmas) or tectonically emplaced (crustal slices or nappes) hotter bodies. Heat sinks, the consumers of heat energy, are endothermic mineral reactions, the minerals themselves via their heat capacities (i.e., the amount of heat needed to be added in order to raise their temperature), and the atmosphere (whereby heat is ultimately lost from the solid Earth at the surface and can be measured as the surface heat flow). Heat transport in the crust is primarily by either conduction (relatively slow) or advection (relatively fast) with radiation being generally insignificant.

Simple Models: The Stable Geotherm

Before we can start to consider the thermal consequences of tectonometamorphic events we must firstly consider the stable temperature distribution in the crust—in its simplest form this is the stable (or steady-state) geotherm. From measurement of the actual heat flow from the Earth at the present-day surface, realistic models for the internal heat production potential (i.e., higher amounts in the upper crustal than in lower crustal rocks), and typical values for heat conduction (measured on rocks in the laboratory), it is possible to estimate the mantle heat input. Then, under the assumption that all rocks are at equilibrium and therefore that the only heat sink is the surface, it is relatively easy to calculate the temperature distribution in the crust, the so-called steady-state geotherm, in a one-dimensional model. This method works well in stable continental regions (Figure 2, right-hand side). Here the measured flow of heat at the Earth's surface, surface heat flow (Q), is around $50–70\,\mathrm{mW\,m^{-2}}$. From the heat flow alone and application of Fourier's Law

$$Q = -k\mathrm{d}T/\mathrm{d}z$$

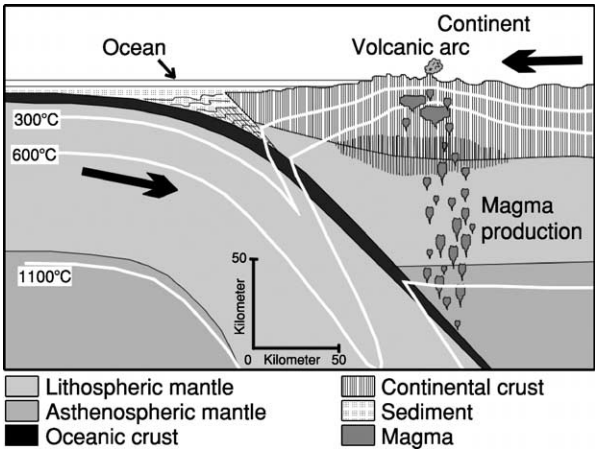


Figure 2 Simplified cross-section of an active continental margin showing the different thermal gradients in the normal crust, magmatic arc, and subduction zone domains.

where Q = heat flow ($W\,m^{-2}$ or $J\,m^{-2}\,s^{-1}$), k = thermal conductivity ($W\,m^{-1}\,K^{-1}$ or $J\,s^{-1}\,m^{-1}\,K^{-1}$), T = temperature (K), and z = depth (m). This heat flow value corresponds to a thermal gradient (rate of change of temperature with depth, i.e., dT/dz) of 22–31°C km^{-1} taking an average thermal conductivity of 2.25 $W\,m^{-1}\,K^{-1}$ (Figure 3A). Consider the environments represented by an active continental margin (Figure 2). Here the subduction of cold oceanic crust causes transfer of heat downwards to the subducting slab and so surface heat flow is lower than in the craton, $Q < 40$ m $W\,m^{-2}$, and the thermal gradient is also lower (<18°C km^{-1}). In contrast, the addition of magma in the arc regime results in a much higher heat flow (Q = 80–200 m $W\,m^{-2}$) and a correspondingly higher thermal gradient (36–90°C km^{-1}). These three different thermal regimes define distinctly different trends within the framework of the metamorphic facies scheme (Figure 3B). The subduction zone trend is one of relatively low temperatures for high pressures, the magmatic arc regime is of relatively high temperatures at low pressures, whereas the cratonic trend lies in-between. If it were possible to drill down into the crust in each of these environments then the formation conditions for the mineral assemblages in the extracted rocks, when plotted in a metamorphic facies diagram, would document these distinctly different trends. The defined trends would not be linear, however, as the irregular distribution of heat-producing elements in the crust defines thermal gradients in the crust that are curves (geotherms) and not simple linear extrapolations of these thermal gradients to depth (Figure 3C).

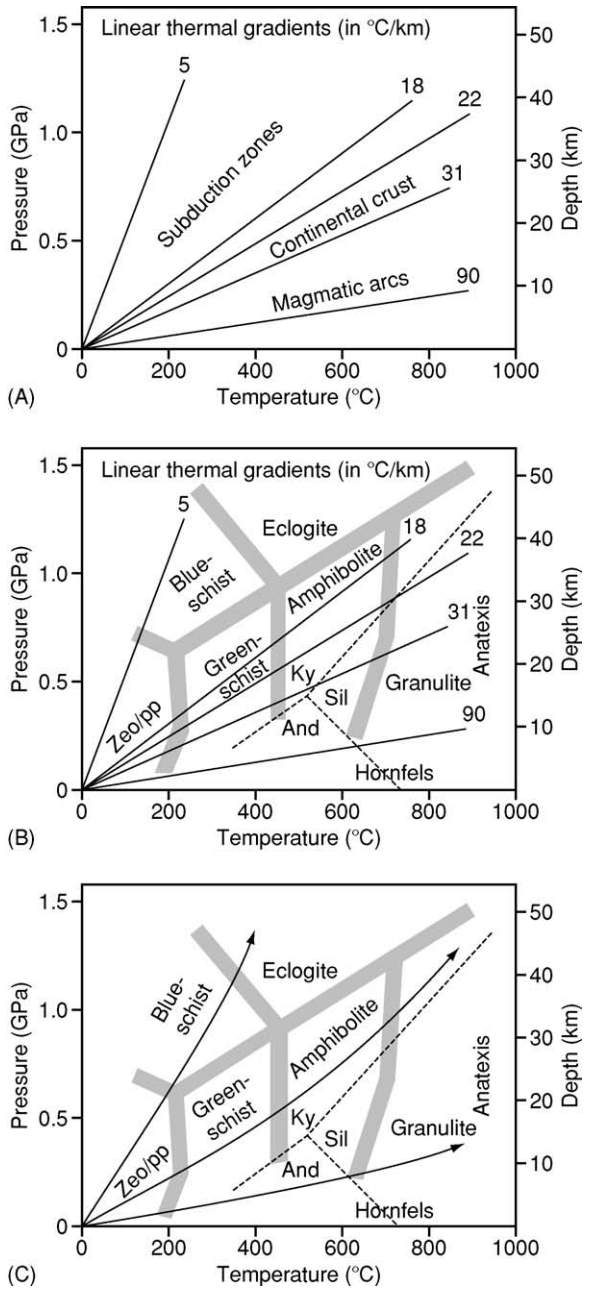


Figure 3 Simple linear temperature–depth trends for subduction, continental crust, and magmatic arc environments (A), the same gradients superimposed on the metamorphic facies grid (B), and more realistic curved geotherms (C) for the same environments.

Simple Models: Perturbation of the Stable Geotherm and Regional Metamorphism

Many plate tectonic processes result in a disturbance or perturbation of these simple temperature–depth patterns. This is because the three-dimensional distribution of heat sources and sinks is also disturbed as a result of crustal stacking, crustal extension, or magmatic intrusion. Consider modern examples of continental collision zones such as the European Alps or Himalayas. Here the present-day crust, deduced from geophysical methods, reaches a thickness twice that normally expected. Assuming a normal crustal geotherm existed before collision, the result of the thickening event has significantly changed the distribution of heat-producing elements such that the potential stable geotherm for the thickened crust is very different and the perturbed geotherm will tend to 'relax' towards this new stable state. In addition, the thickened crust is likely to be mechanically unstable and so will, by a combination of erosion and extension processes, gradually return to a stable thickness. The consequence of crustal thickening, thermal relaxation, and erosion/extension for a discrete rock 'parcel' in the crust is a change in depth and temperature over time, i.e., a PTt path.

A typical PTt evolution for a rock buried and exhumed during a collisional orogeny is depicted in Figure 4. Such PTt paths are calculated by solving the heat flow equation

$$dT/dt = \kappa(\delta^2 T/\delta z^2) + A/(\rho c) - U_z(dT/dz)$$

where T = temperature (K); t = time (s); z = depth below the surface (m); A = radioactive heat production ($\mu W\,m^{-3}$); ρ = density ($kg\,m^{-3}$); c = heat capacity ($J\,kg^{-1}\,K^{-1}$); κ = thermal diffusivity (= k (thermal conductivity)/$c\rho$) ($W\,m^{-1}\,K^{-1}$ or $J\,s^{-1}\,m^{-1}\,K^{-1}$); U_z = vertical (exhumation) rate (ms^{-1}). Solving the equation involves heat conduction, heat production, and heat convection components, in this case in one-dimension, for a specified range of parameters. The PTt path follows a single point in the crust, 'A', lying on a prethickening geotherm (dashed line), that firstly shows a marked pressure increase (the crustal thickening stage) to point 'B' on the geotherm that exists immediately after thickening. If this thickened crustal state persisted, then the resulting thermal state would be the T_∞ geotherm (Figure 4). Thermal relaxation of the post-thickening geotherm towards the T_∞ geotherm results in heating of the observed rock unit. If, as is usual, thermal relaxation is accompanied by erosion or tectonic thinning, then any rock unit would show heating to a thermal maximum ('C' in Figure 4), reached at a pressure below that of the pressure maximum, before cooling. In this case the T_∞ geotherm is never reached because the factors controlling the shape of the geotherm, such as the distribution of heat-producing elements, have been continually modified. Instead, a new geotherm will emerge once erosion and tectonic thinning has ceased. The PTt path resulting from this tectonometamorphic event has the form of a clockwise loop in a pressure–temperature diagram drawn with the pressure-axis upwards. (Note that the same path would be anticlockwise if the pressure axis is downwards.) The segment from A to C is designated the prograde path and that after C is the retrograde part of the PTt path.

It should be noted that many published PTt paths have been modelled by a one-dimensional solution to the heat production–conduction–advection equation, which only allows instantaneous thickening. A consequence for models of crustal thickening by overthrusting is the generation of a saw-tooth-shaped geotherm because the initial post-thickening geotherm is simply a segment of a normal stable geotherm stacked on top of itself (Figure 5). This perturbed geotherm then relaxes over time. In reality, and as realized in two-dimensional modelling, thermal conduction between over- and underthrusting units works to overcome the advected heat component transported in the hanging wall such that the saw-tooth geotherm never occurs. Only in a large-scale underthrusting situation, as in a subduction zone, can an overturn or inversion in the thermal gradient occur. Thus, the oft-shown instantaneous isothermal (i.e., at constant temperature) pressure increase (A to B in Figure 5), followed by an isobaric (i.e., at constant pressure) heating stage for tens of millions of years, is an artefact of one-dimensional thermal models that should be disregarded.

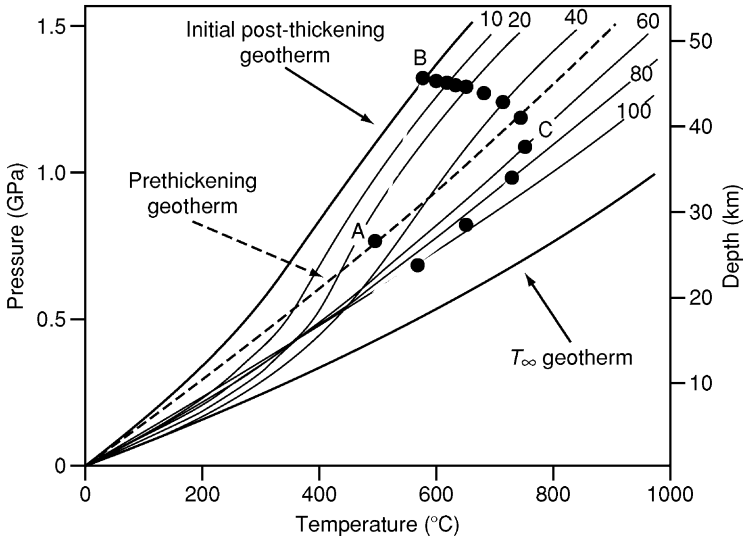


Figure 4 Crustal geotherms before and after a thickening event. The numbers indicate time, after the end of the thickening event, for development of that particular geotherm. A, B, and C mark the prethickening starting point, the pressure peak, and the temperature peak, respectively, for the observed PTt path.

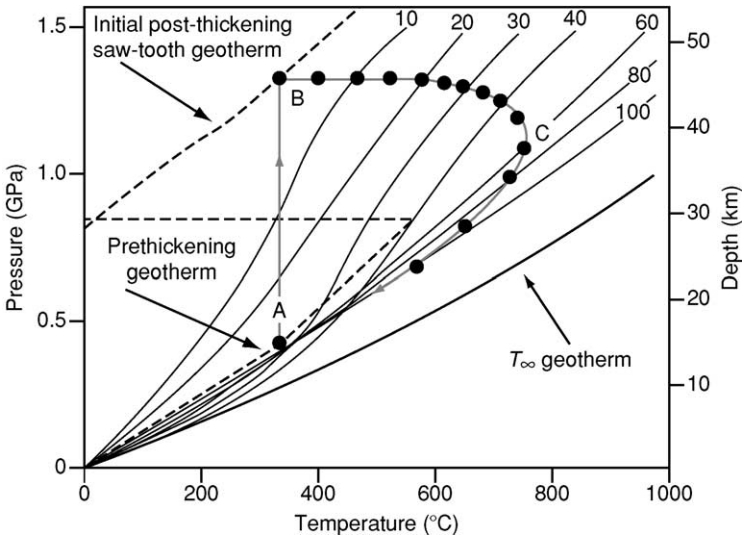


Figure 5 Schematic PTt path for crustal thickening by instantaneous overthrusting as typical for one-dimensional modelling. Point A on the prethrusting geotherm undergoes an isothermal pressure increase to B followed by isobaric heating for 20 Ma before erosion is initiated.

How can the clockwise PTt loop of Figure 4 be used as an aid to understanding the PTt evolution of natural metamorphic rocks? The PTt loop shows a wide range in PT conditions and, if superimposed on a facies diagram, would show an evolution passing through several different facies. Mineral reactions and diffusive transport are, however, thermally activated with kinetics following an Arrhenius relationship. For this reason, reaction rates increase exponentially with temperature and therefore the peak temperature is likely to be the point where most reaction occurs. Cooling after the thermal peak will see a slowing of reaction and material transport such that the peak temperature mineral assemblage is preferentially preserved. In addition, the general shape of dehydration reactions means that a higher degree of fluid release, useful for material transport during prograde reaction, will occur for a path of increasing temperature with minor pressure change. Once fluids have left the system at the thermal peak, any retrogression to hydrous assemblages during cooling will be hindered by the absence of a free fluid phase thus again favouring preservation of the peak temperature assemblage. Combining this information it is possible to predict the most likely determinable PT point for a rock that followed a standard clockwise PTt path, based on preserved mineral assemblage for the temperature peak.

If, for a single segment of crust, several rocks formerly at different depths are traced, it is possible to plot the loci of their peak temperature points (Figure 6A). These are the most probable PT points that would be determinable for rocks that had followed these particular paths and represents the sort of information available from field geology for a tilted and peneplained regional metamorphic terrane: hence the name metamorphic field gradient (sometimes also piezothermic array). Several important points can be deduced from this plot especially when additional temperature–age (Figure 6B) and depth–age plots (Figure 6C) for the same model are presented. Firstly, the metamorphic field gradient defines a PT trend that lies between that of the initial and final geotherms. However, the points defining this trend represent different ages, visible from Figures 6B and 6C, and so the curve does not represent any actual, temporary geotherm that occurred during the thermal relaxation. Thus, the sequence of preserved metamorphic rocks is not the same as would be predicted, for example, along one of the Figure 3 thermal trends. The age difference between the different preserved points is also important. If a major deformation event occurred at 25 Ma, this would be syntectonic with respect to rock B but would be pretectonic for rock A and post-tectonic for rock C. In addition, the difference in depth of the peak-temperature points (see Figure 6C) is less than the true depth difference, thus leading to an apparent thinning of the crustal section. Such thermal models for crustal thickening followed by exhumation, although relatively simple, illustrate some of the fundamental problems inherent in trying to interpret natural metamorphic sequences.

Contact Metamorphism

So far the models presented have considered only the effects of tectonically induced crustal-scale

disturbance of geotherms. These have a time-scale measured in tens of millions of years. A further possibility to disturb a geotherm is by the introduction of heat transported (advected) by magmas. This is a discrete quantity of heat, quickly emplaced as a result of intrusion, and dissipated due to conduction, mineral reaction (i.e., contact metamorphism), and perhaps also hydrothermal convection. In contrast to regional metamorphism, contact metamorphism is more restricted in range and has a duration of only a few million years. In many cases, contact metamorphism is a simple thermal process not related to erosion or tectonic thinning. In such cases, PT paths are essentially isobaric heating trends starting from an initial geotherm, quickly reaching a peak, and then cooling back to the initial geotherm. This is illustrated for the case of a 2-km wide, vertical, mafic intrusion initially at 1175°C, intruded into country rocks with an initial temperature of 300°C, solved by a conduction equation:

$$T_{(a,t)} = T_R + 0.5(T_I - T_R)\ (\mathrm{erf}[(a - x)/(2\sqrt{\kappa t})] + \mathrm{erf}[(a + x)/(2\sqrt{\kappa t})])$$

where $T_{(a,t)}$ = temperature (°C) at point x at time t; T_I = temperature of the intrusion; T_R = preintrusion temperature of the host rocks; a = half-width of the intrusion (m); x = distance from the centre of the intrusion (m); κ = thermal diffusivity of the rocks ($m^2\,s^{-1}$), in this example taken to be 1×10^{-6} $m^2\,s^{-1}$; t = time (s); and erf is the mathematical error function.

The equation is solved for points located at different distances from the intrusion and the resultant T–t-curves plotted in Figure 7. The following features should be recognized. The highest contact metamorphic temperature is reached close to the intrusion, this peak temperature is reached quickly, and the time period where the rock is close to T-max is short. Further away from the contact, the peak temperature is lower, it is reached later, and the duration of the period close to T-max is much longer. The duration of the thermal disturbance is only a few million years,

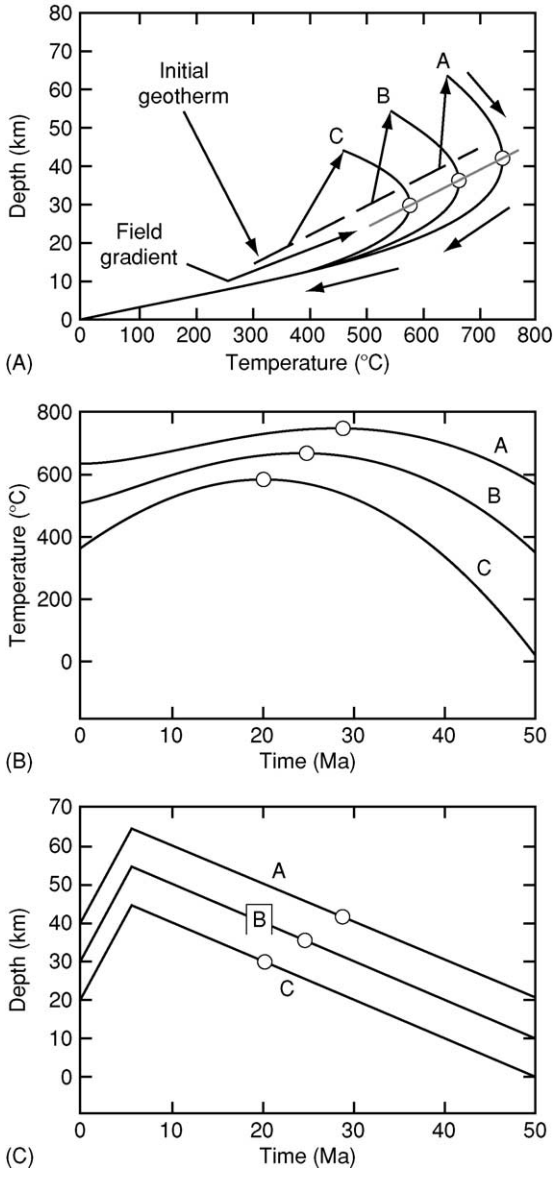


Figure 6 Schematic pressure–temperature, temperature–time, and pressure–time paths for rocks originally sitting on the stable geotherm at depths of 40 (path A), 30 (path B), and 20 (path C) km that were overthrust by a 30-km crust unit in the first 5 Ma and then underwent erosion at 1 mm a^{-1}. Note age difference for peak temperature and depth at peak temperature.

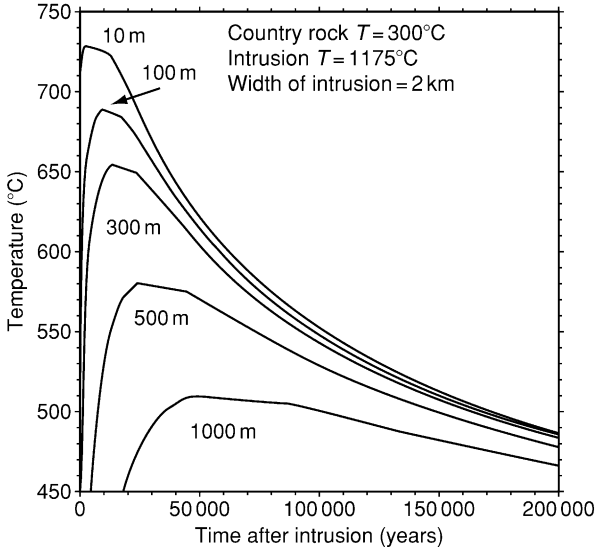


Figure 7 Temperature–time paths for positions at different distances from a 2-km-wide intrusion emplaced at a temperature of 1175°C into country rocks at 300°C (conduction model).

much shorter than the tens of millions of years seen for regional metamorphism, and the variation in *T–t* paths over a small distance is much greater. For these reasons there is a considerable variation, for a given bulk composition, in mineral assemblages as well as textures in a contact aureole.

Age Determination: The Small 't' in PTt

Qualitatively, the reaction sequence in natural rocks is often visible in the form of growth features, compositional zoning, or breakdown textures. Using simple 'rule-of-thumb' techniques these changing assemblages can be attributed to a sequential path in the metamorphic facies diagram. The sequence tells us the relative age but not the absolute age of the different stages. Did the rock experience a monophase metamorphism, a multistage reactivation during a single orogeny, or a complex polyphase evolution? This is only possible by determining the ages of the different metamorphic stages. Weakly metamorphosed rocks may still preserve fossils, thus putting a maximum age limit, or rocks may have fossiliferous deposits overlying them, thus giving a minimum age limit. More commonly, only isotopic dating of minerals within the studied rock, or from related intrusive of extrusive rocks, can yield absolute ages. The time-dependent decays of radioactive isotopes in the systems Rb–Sr, K–Ar, U–Th–Pb, Nd–Sm, or Lu–Hf are commonly used in geochronology. Important to recognize here is that minerals used for isotopic dating show different resistances to loss of radiogenic isotopes important for geochronology. Some minerals, like zircon, can survive even the highest temperatures unscathed, whereas other minerals, like micas and feldspar, only start their use as geochronometers at lower temperatures. The last few kilometres of exhumation may be recorded in the number of unhealed fission-induced traces in minerals such as apatite, titanite, or zircon. Although radioactive decay is time-dependent, the preservation of daughter isotopes or fission tracks is temperature–time-dependent because of the diffusion processes acting to homogenize the phases in question. Generally, in order to quantitatively determine a PTt path, it will be necessary to extract age information using a number of different methods on different minerals.

PTt Paths as Interpretative Tools

The shapes of certain PTt paths or path segments are often characteristic. A rock that remains in the subduction channel for both its prograde and retrograde history will generally stay cool and thus show a 'hairpin-shaped' PTt path (1 in Figure 8). The clockwise path described already is typical for many crustal collision orogens, although there has been considerable argument as to whether the shape of the retrograde path reflects erosion or extension (2 in Figure 8). For a given crustal segment an extensional path should cool quicker but the poor resolution of PTt paths derived for natural rocks rarely allows such a distinction. If exhumation occurs much faster than conductive loss to the surface, the shallow-level, early-exhumed rocks may receive a short-lived thermal pulse. In contrast, some rocks exhibit anticlockwise (for pressure upwards) PTt paths—paths requiring heat addition during burial followed by cooling accompanied by little erosion or extension (3 in Figure 8). This is a significantly different tectonometamorphic scenario only seldom reliably demonstrated. Important PTt path segments that are characteristic are isobaric heating, isobaric cooling, or isothermal decompression paths (6, 5, and 4 in Figure 8). The latter two are very commonly deduced, especially in granulites, where they are segments of incompletely-defined overall paths, and reflect distinct differences in exhumation rates. An isothermal decompression path segment may reflect a fast exhumation in response to buoyancy or extensional processes, whereas an isobaric cooling path segment may reflect the final cooling back to a stable geotherm after the crust has regained isostatic equilibrium. Some rocks may show both these characteristic paths at different times of their history. An isobaric heating path is most probably a reflection of magmatic heat addition or of mantle processes and is most commonly a minor segment of a more usual clockwise path (the spike in path 6 of Figure 8).

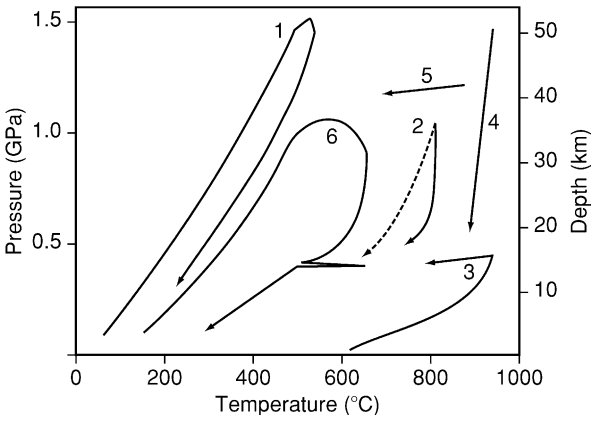


Figure 8 Schematic PT paths and path segments. (1) Subduction zone 'hairpin'; (2) erosional (solid line) or extensional (dotted line) exhumation from the same PT point; (3) anticlockwise path; (4) isothermal decompression; (5) isobaric cooling; (6) clockwise path with short-lived contact metamorphic isobaric heating–cooling segment superimposed.

Conclusions

The PTt path is the curriculum vitae of a rock. Regional or local variations in characteristic PTt paths may allow the boundaries between large-scale rock units to be defined or allow major tectonic boundaries to be identified. The difficulty is that the rocks deliver only a few 'snapshot' positions on a PTt path and do not actually fully define the path. The quantitative estimation of PT conditions, and ages, for the formation of the identified metamorphic minerals or mineral assemblages, requires both luck and judgement: luck in finding an 'informative' sample and judgement in deciding reaction sequence and which phases, or parts of phases, should be analysed chemically and/or isotopically. However, the PTt data from natural rocks deliver important boundary conditions for the theoretical mathematical modelling of orogenic processes. The actual numerical values of the parameters controlling production, transport, and consumption of heat energy, and mechanical stability driving tectonic thickening and thinning of the lithosphere, are so varied in their possible individual ranges that a definitive model for any actual orogenic event does not exist. Regardless, such models must predict: (1) the PT conditions recorded by the mineral assemblages and (2) the heating/cooling, burial/exhumation rates (if the ages or relative ages of mineral assemblages are available). In other words, the true test of the model is how it relates to reality—in this case the actual rocks that have been formed.

See Also

Analytical Methods: Geochronological Techniques. **Metamorphic Rocks:** Facies and Zones. **Plate Tectonics**. **Regional Metamorphism**. **Tectonics:** Mountain Building and Orogeny. **Thermal Metamorphism**.

Further Reading

Bohlen SR (1987) Pressure-temperature-time paths and a tectonic model for the evolution of granulites. *Journal of Geology* 95: 617–632.

England PC and Molnar P (1990) Surface uplift, uplift of rocks and exhumation of rocks. *Geology* 18: 1173–1177.

England PC and Richardson SW (1977) The influence of erosion upon the mineral facies of rocks from different metamorphic environments. *Journal of the Geological Society, London* 134: 201–213.

England PC and Thompson AB (1984) Pressure-temperature-time paths of regional metamorphism I. Heat transfer during the evolution of regions of thickened continental crust. *Journal of Petrology* 25: 894–928.

Ernst WG (1973) Blueschist metamorphism and P-T regimes in active subduction zones. *Tectonophysics* 17: 255–272.

Harley SL (1989) The origin of granulites: A metamorphic perspective. *Geological Magazine* 126: 215–247.

Spear FS (1993) *Metamorphic Phase Equilibria and Pressure-Temperature-Time Paths*. MS America Monograph. Washington, DC: Mineralogical Society of America.

Spear FS and Peacock SM (1989) *Metamorphic Pressure-Temperature-Time Paths*. Washington, DC: American Geophysical Union Short Course in Geology.

Stüwe K (2002) *Geodynamics of the Lithosphere: An introduction*. Berlin: Springer-Verlag.

Thompson AB and England PC (1984) Pressure-temperature-time paths of regional metamorphism II. Some petrological constraints from mineral assemblages in metamorphic rocks. *Journal of Petrology* 25: 929–955.

Treloar PJ and O'Brien PJ (eds.) (1998) *What Drives Metamorphism and Metamorphic Reactions?* Geological Society, London: Special Publications, 138.

METEORITES

See **SOLAR SYSTEM: Meteorites**

MICROFOSSILS

Contents

Acritarchs

K J Dorning, University of Sheffield, Sheffield, UK

Introduction

The acritarchs are a very large group of organic-walled microfossils of unknown or uncertain affinity. They are mostly unicellular, or apparently unicellular, although they may be found in clusters. The acritarchs are an informal, polyphyletic, organic-walled microfossil group, conceived as a holding category for the very large numbers of *incertae sedis* recorded in palynological assemblages that have no clear affinities. Many acritarchs are likely to be the cysts, temporary resting stages, or phycoma produced as part of the life cycle of marine planktonic algae. A few forms are recorded in apparently lacustrine, fluvial, and terrestrial environments. They were formerly regarded as part of the hystrichospheres (Hystrichophyta) until the type material and many others from the Mesozoic and Tertiary were recognized to be dinoflagellate cysts. Some others are now considered to be representatives of the Prasinophyta or Chlorophyta, although, in the Palaeozoic, they are generally studied as 'acritarchs and associated forms', because many described taxa show little or no evidence of their affinities. A few forms have morphological similarities with copepod eggs and masuelloids. The acritarchs form the main fossil record of global photosynthetic production during the Late Precambrian and Early Palaeozoic, prior to the colonization of the land by plants and consequent terrestrial photosynthesis. There are over 7000 species described, which, given the number of undescribed forms, would suggest that there are well over 10 000 acritarch species preserved in the fossil record. Acritarchs are routinely recorded in abundance in Late Precambrian, Palaeozoic, and Early Mesozoic marine sediments, and are regularly recorded, together with dinoflagellate cysts, in later Mesozoic and Cenozoic marine sedimentary sequences. They are generally studied by examining palynological preparations produced by dissolving rock samples in hydrochloric and hydrofluoric acids in specialist laboratories, although they can also be examined in rock thin sections. Most routine studies are undertaken with strew mounts using transmitted light microscopes, with ultrastructural studies using the scanning electron microscope and transmission electron microscope. Because of their small size, abundance, and diversity, they are of considerable importance in geological interpretation, including biostratigraphy, palaeoenvironmental interpretation, and geothermal alteration within depositional basin analysis studies.

Acritarch Occurrence, Preservation, and Geothermal Alteration

Acritarchs are regularly recorded in abundance from siliciclastic and carbonate marine sediments, including mudstones, silty mudstones, sandy mudstones, shales, siltstones, sandstones, and limestones. In many sequences deposited in shelf sea areas, there is a continuous record of the preserved phytoplankton productivity. Generally, they are most abundant in fine-grained sediments, because of the slower rate of sedimentation. Specimens preserved in grey silty mudstones and muddy siltstones are often of excellent preservation, reflecting the dysoxic to anoxic conditions at the sediment–water interface. Specimens recorded from sandstones and bioclastic limestones are of very varied preservation, depending on the oxygen availability in the sediment. Those recorded in highly oxic conditions are often badly degraded,

whilst some sandstones and early diagenetic limestones can preserve specimens in three dimensions, uncrushed between the mineral grains. Forms recovered from laminated dark shales deposited with slow rates of sedimentation are normally flattened, and are often associated with the growth of minute pyrite framboids that can impact on the internal and external walls of palynomorphs, including acritarchs. Recrystallization in limestones, particularly dolomites, may also distort the acritarch wall. In areas of deep burial or high heat flow, the temperature increase has the effect of changing the colour of acritarchs from transparent to pale greenish-yellow, through increasingly darker shades of brown, to grey or black, depending on the wall thickness (Figure 1).

These series of colour and transparency changes with different temperatures are affected by the wall thickness, as well as by the composition of the different complex organic compounds, comparable with the changes that can be observed with complex carbohydrates when preparing toast. Abundant long-ranging forms, including *Leiosphaeridia* and *Veryhachium*, have been used to estimate maximum rock palaeotemperatures. This colour change has little impact on the preservation of the acritarchs until temperatures associated with the transition of mudstones to slate are reached, when the cleavage that penetrates the sediments and the associated shrinkage of the organic matter tend to fragment the acritarchs. In these situations, thin sections made parallel to the bedding may be useful for identifying the acritarchs present. In the Ordovician, Silurian, and Devonian, acritarchs are often recorded in palynological assemblages together with lower numbers of chitinozoans, graptolite fragments, and scolecodonts, the jaw apparatus of annelid worms. From the Middle Triassic, forms recognizable as dinoflagellate cysts may also be present in the marine microflora; from the Middle Jurassic, dinoflagellate cysts generally dominate the organic-walled marine microflora in marine shelf sediments.

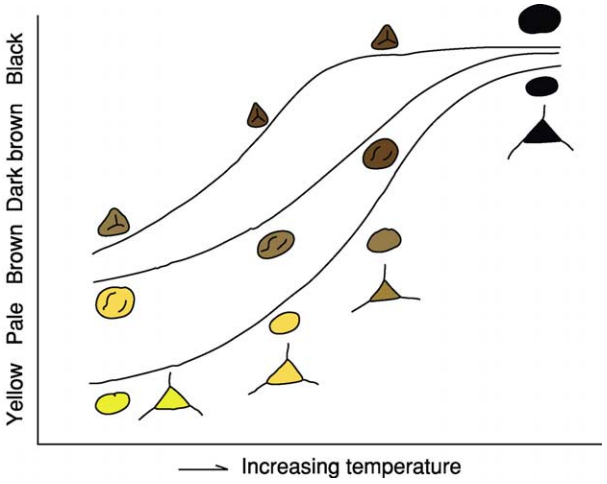


Figure 1 Diagram illustrating the change in acritarch colour with increasing temperature for *Veryhachium*, *Leiosphaeridia* and a thick-walled sphaeromorph, in comparison with a trilete spore.

Acritarch Morphology

The general morphology of most acritarchs consists of a hollow, organic-walled, enclosed cell of varied shape that may extend into flanges or processes with a wide variety of morphology. In almost all forms, the wall is very resistant to decay, and appears to be as resistant to degradation as sporopollenin. Forms that show a reflected paratabulation, an angular opening or archaeopyle, or evidence of a sulcus are generally considered to be dinoflagellate cysts. Three main morphologies can be recognized: acritarchs without processes or flanges, acritarchs with flanges but no processes, and acritarchs with processes with or without flanges (Figure 2).

For all acritarchs, the main body is known as a vesicle, with wing-like projections or alae known as flanges and large projections referred to as processes, which in many forms are spinose. Some specimens show evidence of an excystment opening, which

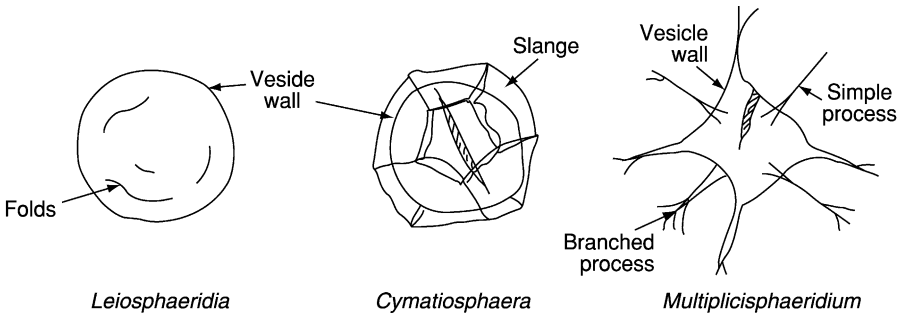


Figure 2 Morphology of *Leiosphaeridia*, a sphaeromorph acritarch, *Cymatiosphaera*, an acritarch with flanges, and *Multiplicisphaeridium*, an acritarch with processes.

enabled the cell contents of the next stage in the life cycle to be released.

The vesicle shape is highly varied, and has been used as a simple method of separating different groupings of acritarchs. The shapes observed include spherical to subspherical, ovoidal to elongate, flattened triangular to rectangular to polygonal (Figure 3).

The vesicle size varies from about 2 to 2000 μm, although many acritarchs fall into the size range of 10–50 μm.

Flanges and Processes

Some acritarchs have thin vertical projections, wings, or alae, known as flanges, projecting from the surface, which divide the vesicle surface into two or more fields. Many of these forms are considered to be prasinophytes, although this is less certain in forms in which processes support the flanges. Projections that arise from the surface, whether spinose or bulbous, are termed processes. These are distinguished from ornament at an arbitrary size of 2 μm. The process shape, number, and distribution on the vesicle surface are of particular importance in the speciation of many acritarchs. The processes may be solid or tubular, tapering or inflated, and may have a sharp to wide base at the contact with the vesicle (Figure 4). In forms with hollow processes, the interior of the process may communicate with the vesicle or be plugged by the inner vesicle wall. The processes may be simple and unbranched, or branched in a regular or irregular manner. The unbranched portion of the process is sometimes called the trunk. The first order of branching forms two or more branches, or pinnae, that are of similar or different length, and at various branching angles. The branches may continue to branch at the second, third, fourth, or more orders in a regular or irregular manner. In addition, the processes may be totally or partly ornamented or, in some forms, interconnected with a thin veil.

Vesicle and Process Wall

There are both single and double vesicle walls, with a wall thickness in the range 0.5–3.0 μm. Several different types of wall structure can be recognized. In the micrhystridian wall type, the single wall is thin, apparently homogeneous in nature, and generally appears to be of similar thickness within a species. The micrhystridian wall type is known in *Micrhystridium* and *Acanthodiacrodium*. In the tasmanitid wall type, the wall is uniform, but laminated with narrow radial pores. The wall is often thick. The tasmanitid wall type is found in *Tasmanites*. Other forms, including *Baltisphaeridium*, have walls with very narrow radial pores. The visbysphaerid wall type includes forms with a double wall, with the inner wall generally thicker and, in transmitted light, visibly darker than the outer wall. The visbysphaerid wall type is found in *Visbysphaera*. Each of the wall types may be smooth (laevigate) or ornamented with granulate, microgranulate, foveolate, costate, echinate, or striate ornament, or a combination of these.

Excystment Opening

Some acritarchs show excystment openings that allowed the cell contents to escape the cyst stage to form the next stage of the life cycle (Figure 5). In some acritarchs, this is a circular pylome, or cyclopyle, with apparently the same function as the archaeopyle in dinoflagellate cysts. Some forms have a particularly large pylome, or macropyle. In others, the vesicle splits into two equal halves by means of a median split. Others show large splits in the vesicle wall to form a wide gape, while a few forms almost split into four quarters, similar to an opening flower. Some specimens show a short split with a flap, also known as an epityche, or an irregular split in the vesicle wall, although these can sometimes be difficult to distinguish from accidental damage as a result of compaction in the sediment. Within each species, the proportion of specimens showing an excystment mechanism can be used in palaeoenvironmental interpretation.

Acritarch Clusters

Most of the organic-walled marine microflora, including the acritarchs and dinoflagellate cysts, occur as isolated individuals in palynological preparations and thin sections. Most are the benthonic cysts or resting stages of the planktonic marine microflora, although some of the cysts or phycoma were probably planktonic. If care is taken during palynological processing, some forms can be recovered as monospecific clusters. In the Late Precambrian, clusters of sphaeromorph acritarchs, including *Leiosphaeridia*, are particularly common. In the Cambrian and Ordovician, clusters of *Acanthodiacrodium*, *Cymatiogalea*, *Leiosphaeridia*, *Micrhystridium*, and *Polygonium* are not infrequently recorded, whilst, in the Silurian, clusters of *Cymbosphaeridium*, *Dilatisphaera*, *Leiofusa*, *Leiosphaeridia*, *Micrhystridium*, and *Veryhachium* are sometimes recorded. Clusters have also been observed in a few Mesozoic acritarchs and dinoflagellate cysts. Typically, all of the individuals in a cluster are identical in size. Clusters are particularly abundant in laminated sediments that lack bioturbation, which would account for their frequency in the Precambrian and their widespread distribution in the Early Palaeozoic, prior to the shelly

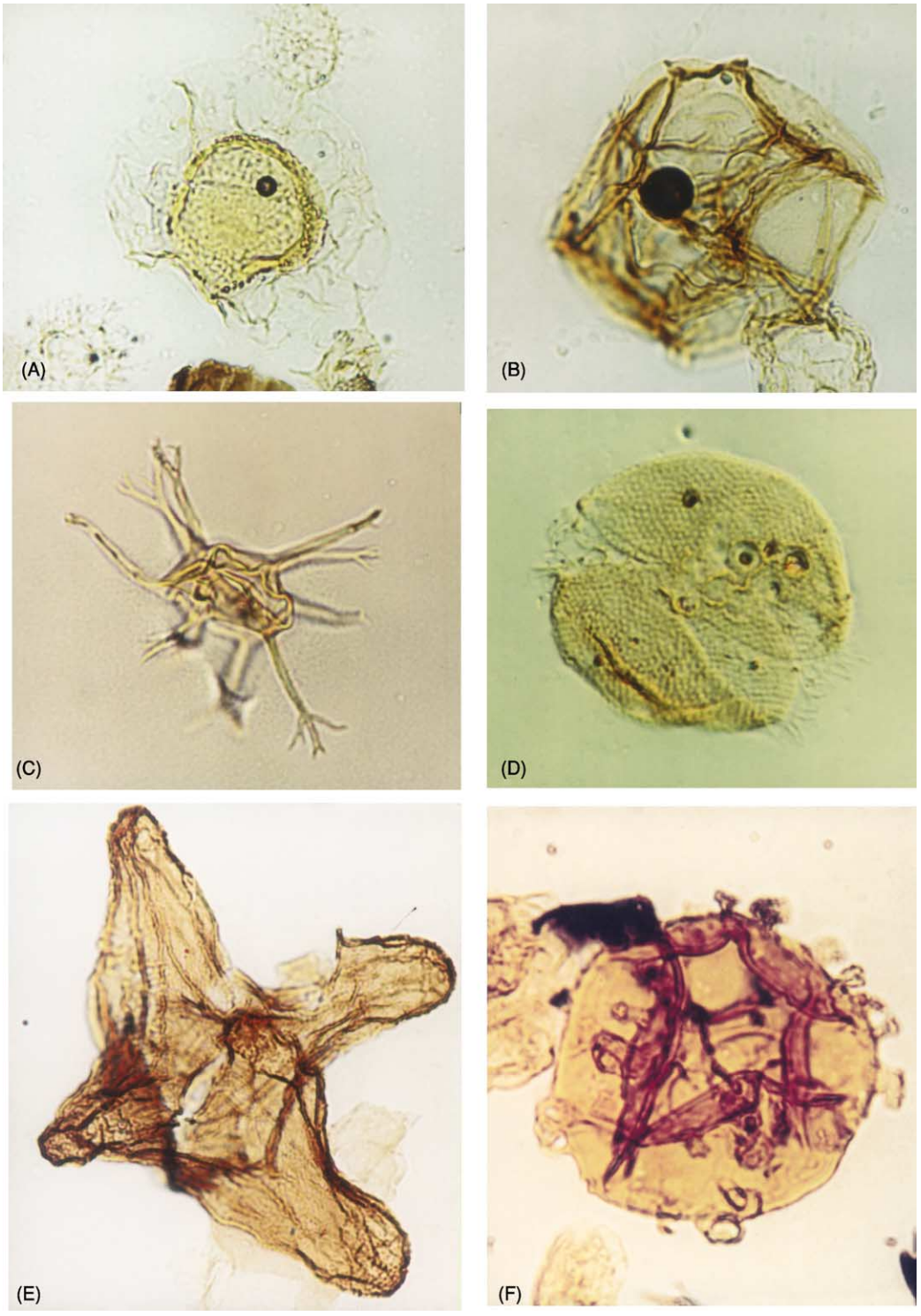


Figure 3 Colour photomicrographs of acritarchs: (A) *Pterospermella* (×1000); (B) *Cymatiosphaera* (×1000); (C) *Multiplicisphaeridium* (×1000); (D) *Helosphaeridium* (×1000); (E) *Pulvinosphaeridium* (×500); (F) *Visbysphaera* (×1000).

macrofossil migration from shallower into deeper shelf areas which caused bioturbation. Although it is possible that some monospecific clusters were derived from production in a sporangia, it is probable that others formed aggregates to make up part of the plankton snow, as a mechanism against predation by zooplankton by increasing the speed of sinking from the surface water to the sediment-water interface.

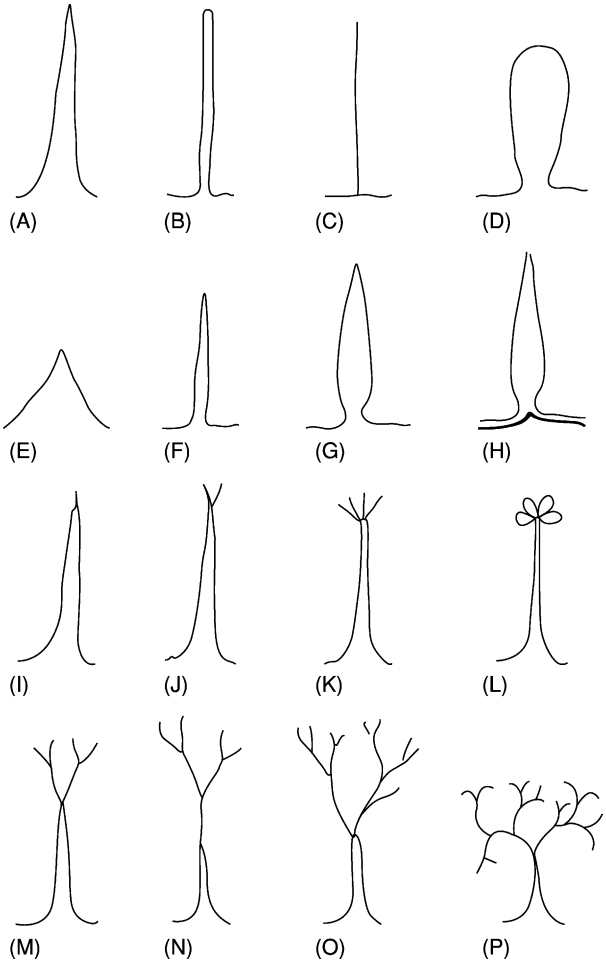


Figure 4 Process morphology: (A) tapering, distally sharp; (B) tubular, distally blunt; (C) solid, wire-like; (D) expanded; (E) wide base; (F) narrow base; (G) constricted base; (H) plugged base; (I) simple, unbranched; (J) bifurcate; (K) multifurcate; (L) distal loops; (M) regularly branched in two orders; (N) irregularly branched in two orders; (O) irregularly branched in three to four orders; (P) irregularly branched at a high angle.

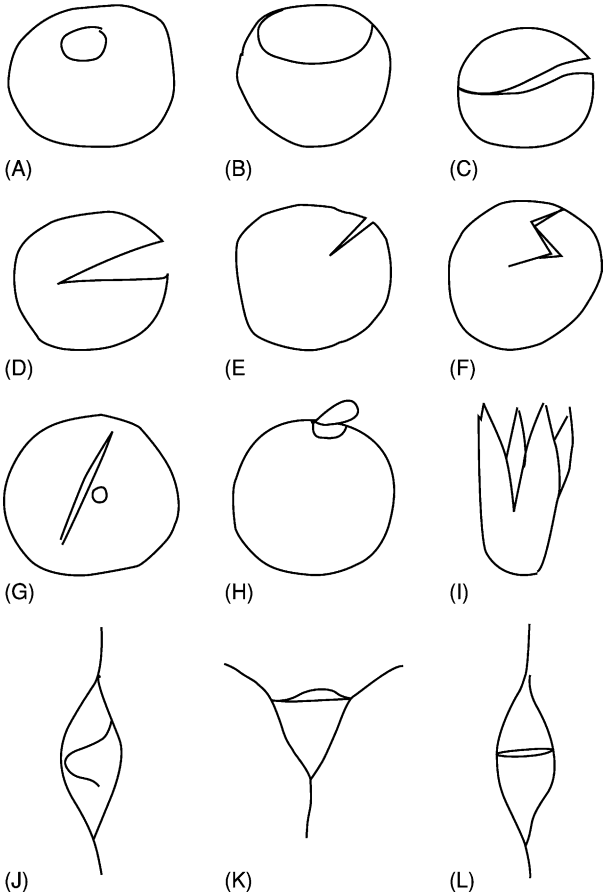


Figure 5 Excystment openings: (A) pylome; (B) macropyle; (C) split into two equal hemispheres; (D) gape split; (E) straight split; (F) irregular split; (G) straight split and small pylome; (H) hinged pylome; (I) split into four; (J) sinusoidal split; (K) flap-style epityche; (L) split into two halves.

Acritarch Classification

In general, the acritarchs are considered as form genera, and, as such, it is inappropriate to use a formal classification system. For this reason, the acritarchs are treated as *incertae sedis* microfossils, organic-walled microphytoplankton, or group Acritarcha Evitt, 1963. They have been divided, mostly on the basis of vesicle shape, into a number of informal subgroups. This system of subdivision is less than satisfactory, as there are a number of gradational forms between subgroups, particularly in forms with processes between spherical and polygonal vesicles. None of the subdivisional schemes for the acritarchs is entirely satisfactory, and many research workers list all the acritarchs, algal cysts, and colonial algae together alphabetically when describing palynological assemblages. The following classification refers to the subgroups, but also documents other groupings that may provide progress towards a more natural classification.

Acritarchs without Processes or Flanges (Figure 6)

This category includes the subgroup Sphaeromorphitae, the sphaeromorph acritarchs, together with the Schizomorphitae, Scutellomorphitae, and *Navifusa* group. The grouping is almost certainly polyphyletic, and includes forms attributable to the green algal groups Chlorophyta (Chlorophyceae) and Prasinophyta (Prasinophyceae). Some forms with an indistinct wall may be related to the Cyanophyta (cyanobacteria; blue–green algae). The first organisms that could be seen as acritarchs date from 3800 Ma, but these should probably be considered to be Cyanophyta, with the first sphaeromorph acritarchs of 20–200 μm in diameter known from 1800 Ma, following the rise in oxygen levels in the

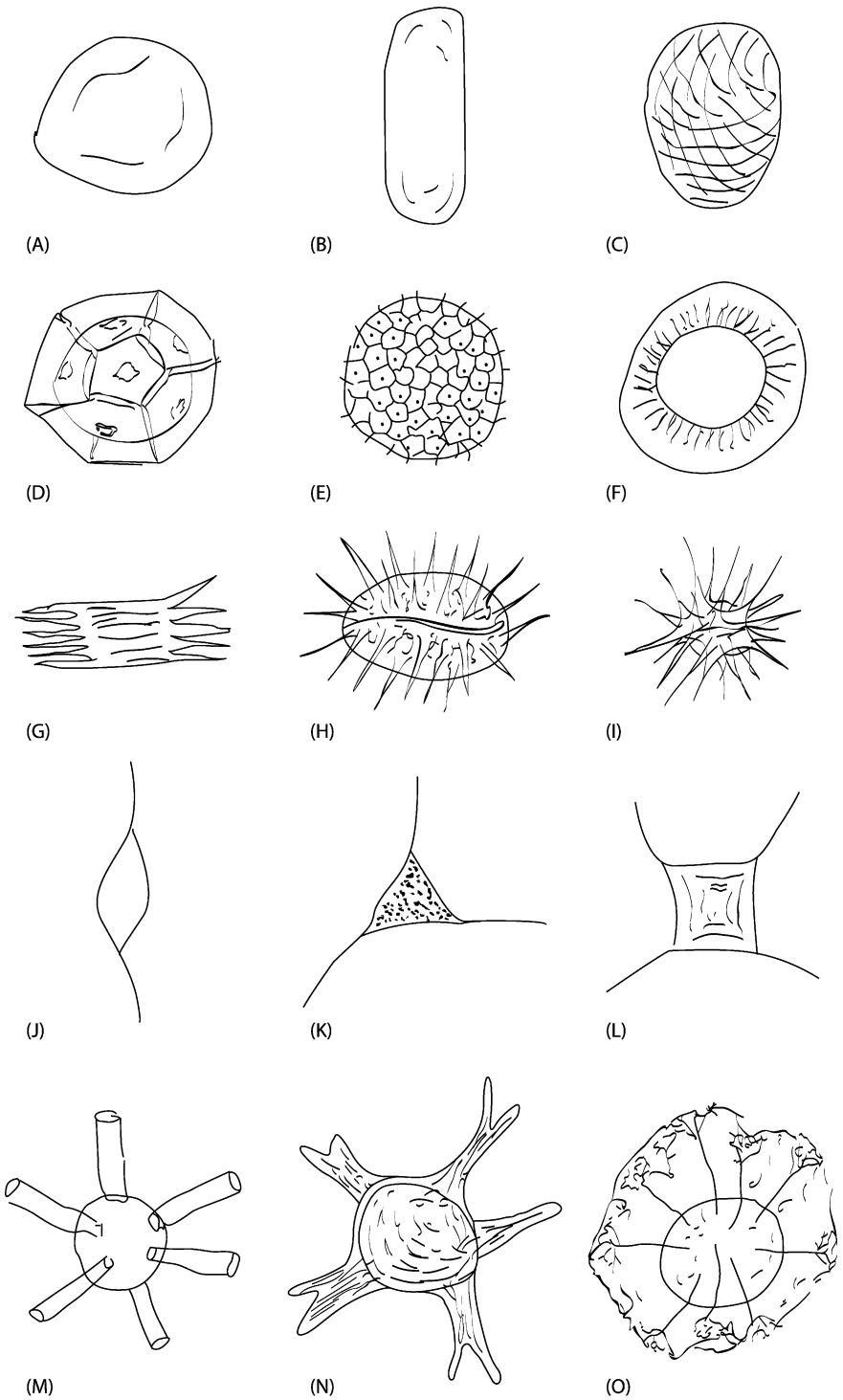


Figure 6 Acritarchs without processes or flanges: (A) *Leiosphaeridia*; (B) *Navifusa*; (C) *Moyeria*. Acritarchs with flanges: (D) *Cymatiosphaera*; (E) *Dictyotidium*; (F) *Pterospermella*. Acritarchs with processes: (G) *Acanthodiacrodium*; (H) *Salopidium*; (I) *Micrhystridium*; (J) *Leiofusa*; (K) *Villosacapsula*; (L) *Neoveryhachium*; (M) *Dilatisphaera*; (N) *Leptobrachion*; (O) *Tunisphaeridium*.

atmosphere. The sphaeromorph acritarchs have a long continuous record through the Late Precambrian, Palaeozoic, Mesozoic, and Cenozoic, and continue through the Holocene to the present day. Some sphaeromorph acritarchs, including forms referable to *Leiosphaeridia* and *Lophosphaeridium*, occur in abundance in apparently freshwater palaeoenvironments, as well as shallow shelf to open ocean

environments. The inclusion of the sphaeromorph acritarchs within the informal group Cryptarcha has not been generally adopted. The clear distinction between some thick-walled sphaeromorph acritarchs, cryptospores, and thin alete plant spores is often difficult in Palaeozoic palynological assemblages. The *Navifusa* group of elongate acritarchs includes marine forms in the Ordovician, and *Quisquilites*, a possible non-marine form, in the Devonian. The *Moyeria* group includes the discoidal acritarchs *Moyeria*, *Chomotriletes*, and *Concentricystes*, acritarchs of apparent terrestrial or fluvial affinity that may possibly have affinities with the Euglenophyta or Prasinophyta.

Acritarchs with Flanges, but without Processes (Figure 6)

This category includes many forms in the subgroups Herkomorphitae (herkomorph acritarchs) and Pteromorphitae (pteromorph acritarchs). Some species attributed to *Cymatiosphaera*, *Dictyotidium*, and *Pterospermella* are clearly related to the Prasinophyta, whilst others are of uncertain affinity. The cysts or phycoma of many prasinophytes are photosynthetic and therefore planktonic, which may explain their wide distribution in the marine realm from shelf seas to the open ocean. The size of individuals within many species of *Cymatiosphaera* and *Dictyotidium* is often rather variable, making consistent identification difficult.

Acritarchs with Processes, with or without Flanges (Figure 6)

This category includes many forms in the subgroups Acanthomorphitae (acanthomorph acritarchs), Coryphomorphitae, Diacromorphitae (diacrodian acritarchs), Netromorphitae (netromorph acritarchs), Oomorphitae (oomorph acritarchs), Polygonomorphitae (polygonomorph acritarchs), and Prismatomorphitae (prismatomorph acritarchs). Some very large acritarchs with rigid processes may have affinities with the masuelloids (Muellerisphaerida), planktonic zooplankton with an organic and phosphate wall that are mostly found in deep-water oceanic sediments.

Several groupings can be recognized in acritarchs with processes, based on wall structure and ornament, process style, and excystment method. These groupings are particularly useful in palaeoenvironmental interpretations.

- *Acanthodiacrodium* group. This group is characterized by a simple wall structure and elongate body shape. The group is particularly abundant in the Late Cambrian to Early Ordovician, where it grades into forms with four corners, attributable to *Coryphidium*.
- *Ammonidium* group. This group has an excystment structure that divides the vesicle into two equal halves. *Salopidium*, *Ammonidium*, and *Gracilisphaeridium* show increasing process complexity. The Silurian acritarchs, *Helosphaeridium* and *Percultisphaera*, have a similar excystment mechanism.
- *Cymatiogalea* group (galeate acritarchs). This group is characterized by a large cyclopyle as found in *Cymatiogalea*, *Priscogalea*, and *Stelliferidium*, and is characteristic of the Early Ordovician.
- *Diexallophasis* group. This group, which includes *Evittia* and *Tylotopalla*, is characterized by an irregular echinate ornament. A few species are found in the Late Ordovician, but it is a common component of the Silurian and Devonian assemblages.
- *Domasia* group. This group of small acritarchs with a simple wall and low number of processes of variable length includes *Deunffia*. It is abundant in the Late Llandovery to Early Wenlock.
- *Diliatisphaera* group. *Dilatisphaera* is unusual in having thin tubular processes that open distally. The wall composition is different from most acritarchs in that it is reluctant to take safranin stain.
- *Duvernaysphaera* group. This group, which includes *Quadraditum*, has an equatorial flange supported by processes. It is possible that the equatorial flange may suggest that it could be included in the Prasinophyta.
- *Eisenackidium* group. This group with a double vesicle wall and a thin flexible process wall includes *Eisenackidium*, with simple processes, and *Leptobrachion*, with branched processes.
- *Eupoikilofusa* group. This group of large elongate acritarchs includes large forms of *Leiofusa*. It is commonly recorded in the Silurian and Devonian.
- *Estiastra* group. These are very large acritarchs with a very thin wall, sometimes with dark blotches, and include *Hogklintia* and *Pulvinosphaeridium*. They are only found in Late Ordovician to Silurian tropical carbonate environments. Their large size, together with an apparently different wall composition, suggests that they might not be cysts of planktonic algae.
- *Micrhystridium–Veryhachium* group. This is a large group of small acritarchs with a simple wall and simple processes. The group, which includes some small forms of *Leiofusa*, is long ranging from the latest Precambrian to Tertiary. Representatives of this group are the main acritarchs with processes in the Permian, Triassic, Jurassic, Cretaceous, and Cenozoic.

- *Onondagella* group. This group has three different processes and a wall that is reluctant to take stain. It is possibly related to the chlorophyta, given that four specimens may be arranged to form a tetrahedral colony similar to *Deflandrastrum*, although this has not been observed.
- *Oppilatala* group. This grouping has a large gape excystment opening and includes *Oppilatala* and some species attributed to *Dateriocradus*.
- *Orthosphaeridium* group. This group with plugged bases to the processes and a straight split as an excystment opening includes *Bacispheridium*.
- *Tunisphaeridium* group. This group with solid, wire-like processes includes *Carminella*, *Geron*, and *Elektoriskos*.
- *Visbysphaera* group. This group normally has a double-walled vesicle and numerous heteromorphic processes. Specimens without the inner wall are often observed.

Biostratigraphy

The Precambrian record of the acritarchs is dominated by the sphaeromorph acritarchs. In the Late Palaeoproterozoic and Mesoproterozoic, the low-diversity assemblages are composed of laevigate sphaeromorph acritarchs. Striate vesicles appear in the Late Mesoproterozoic, with large acanthomorph acritarchs and ornamented sphaeromorphs in the Neoproterozoic. Many acritarchs appear to become extinct at the Varanger ice age, but, after this event, diverse assemblages of ornamented sphaeromorph acritarchs are seen until the latest Proterozoic, where low-diversity laevigate sphaeromorphs are dominant. Acritarchs diversified rapidly in the Early Cambrian, with the first appearance of *Micrhystridium* close to the Ediacaran–Cambrian boundary, and diverse assemblages including *Skiagia* and *Archaeodiscina* soon after. Throughout the Cambrian, Ordovician, Silurian, and Devonian, there are successions of diverse acritarch microfloral assemblages that are of particular value in the biostratigraphy of Palaeozoic successions. Formal and informal biostratigraphical zonation schemes have been proposed for sections in the Cambrian, Ordovician, Silurian, and Devonian, based on assemblage characteristics and the first appearance of distinctive acritarch taxa. Major changes of note include the occurrence of *Eliasum* in the Middle Cambrian, and the proliferation of the *Acanthodiacrodium* and *Cymatiogalea* groups, together with *Vulcanisphaera*, in the Late Cambrian and Early Ordovician in western Europe and eastern North America. Acritarch abundance was also exceptionally high in this area, with values of over 100 000 per gram of rock regularly recorded. Middle Ordovician sediments contain the distinctive acritarchs *Coryphidium* and *Frankea*. Late Ordovician assemblages are characterized by an abundance of acritarchs with plugged process bases, including *Baltisphaeridium*, *Orthosphaeridium*, and *Ordovicidium*, together with *Peteinosphaeridium*. Late Ashgill assemblages associated with the Early Hirnantian glacial events contain many small distinctive acritarchs, including *Villosacapsula* and *Pheoclosterium*, together with cryptospores. The acritarchs show a remarkably low diversity in the Early Llandovery, perhaps associated with continuing cool conditions. In Arabia and North Africa, sediments with a high total organic carbon are associated with acritarch assemblages with large numbers of *Leiosphaeridia*. Cryptospores and recycled acritarchs are regularly recorded in this interval, suggesting that relative sea-levels were still fairly low. By the Late Llandovery, acritarch diversity had recovered, and high-diversity assemblages are routinely recorded from this interval. The Late Llandovery to Early Wenlock is characterized by diverse acritarchs, including species of *Deunffia* and *Domasia* in tropical areas. Diversity continued to be high during the rest of the Silurian, with some evidence of shifts in assemblage related to cyclic alternations from warm and dry to cooler and wetter climatic episodes. Assemblages during the Devonian continued to be diverse, with species of *Poledryxium* being conspicuous, and species of *Cymatiosphaera* becoming an important component. The latest Devonian to Early Carboniferous interval shows a remarkable decline in the number of acritarchs recorded, although this is partly because there is a significant drop in the numbers of the acritarchs of medium to large size, together with many palynological assemblages being dominated by terrestrial spores. Carboniferous acritarchs are of low diversity, with small forms of *Leiosphaeridia*, *Micrhystridium*, *Unellium*, and *Veryhachium* recorded from marine sediments. There are very few records of acritarchs from Late Carboniferous sequences, but, during the Permian, the numbers recovered to some extent, with new species of *Unellium* conspicuous. The Triassic is notable for the abundance of small species of *Micrhystridium* and *Veryhachium*, which continue into the Early Jurassic, together with *Metaleiofusa* and *Multiplicisphaeridium* and dinoflagellate cysts of low diversity. Rapid diversification of the dinoflagellate cysts occurred in open marine shelf areas during the Middle Jurassic, although species of *Micrhystridium* and *Veryhachium* are often found in palynological assemblages through the Mesozoic and Cenozoic. Oceanic sediments continue to contain

acritarchs of small size, and they are notable in the Late Cretaceous pelagic sediments, including the Chalk. Acritarchs and dinoflagellate cysts were first studied in detail within flints in the Chalk, and continue to be of value in the sourcing of archaeological flint and chert. Assemblages in the Cenozoic are also dominated by dinoflagellate cysts, although the marine phytoplankton record also includes the calcareous nanoplankton (Haptophyta, Coccosphaerales) and diatoms (Bacillariophyta). Acritarchs are recorded together with dinoflagellate cysts in Holocene marine sediments, although some may include copepod eggs.

Palaeoenvironmental Distribution

In common with present-day marine phytoplankton, the acritarchs show distributions related to water masses, so that different nearshore, offshore shelf, and basinal assemblages can be recognized (Figure 7). Forms that are interpreted as benthonic cysts are most common in shelf areas, whereas forms with planktonic cysts, including photosynthetic phycoma, are widespread and common in deep-water oceanic sediments. On a global basis, some forms are apparently restricted to broad latitudinal belts, comparable to modern phytoplankton. The palaeoenvironmental interpretation of Late Precambrian acritarchs is uncertain, although it is probable that different sizes of sphaeromorph acritarch were influenced by nutrient availability. In the Cambrian, many acritarchs are widespread, although *Micrhystridium* favoured inshore areas, whereas more complex forms, including *Skiagia*, are found in offshore areas. During the Ordovician and Silurian, clear palaeoenvironmental distribution patterns become established, with distinct assemblages associated with nearshore shelf, offshore shelf, and deep-water areas. In the Late Ordovician, nearshore areas contain many *Leiosphaeridia*, with *Peteinosphaeridium* and *Dicommopalla* characteristic of shoal areas, and diverse acritarchs with *Baltisphaeridium*, *Veryhachium*, and *Polygonium* in open sea shelf areas. In the Silurian, nearshore assemblages are of low diversity and low to moderate abundance, with many small, thin-walled *Leiosphaeridia*, together with *Veryhachium* with three short processes, *Micrhystridium*, and *Diexallophasis*, together with land plant spores. The offshore shelf assemblage is of high diversity and moderate abundance, with no one taxa dominating the assemblage. Species of *Veryhachium* often have three or four medium to long processes. This pattern of increasing length of process in an offshore direction is also documented from the Early Jurassic. The assemblages associated with deep water are of low diversity and low to moderate abundance, with *Leiosphaeridia* dominant, particularly thick-walled forms. Other genera present in low numbers include *Cymatiosphaera* and *Pterospermella* that have planktonic phycoma, together with occasional *Diexallophasis*, *Micrhystridium*, and *Diexallophasis*, which are the most common acritarchs on the shelf. There is some evidence from the Silurian to suggest that acritarch assemblages may be used to predict oceanic phytoplankton productivity and palaeoclimate, as certain forms are associated with argillaceous sediments formed during cool, wet, primo episodes, with others associated with more calcareous sediments deposited during drier, warmer, secundo episodes.

During the Silurian and Devonian, distribution patterns are also recognized within carbonate environments, with distinct assemblages associated with patch reef, inter-reef and deeper non-reef areas

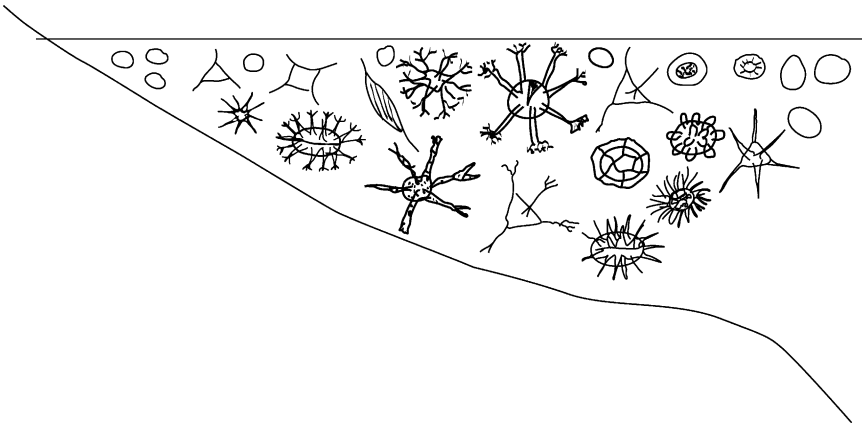


Figure 7 Diagrammatic section from a shallow to deep shelf tropical sea during the mid-Silurian, illustrating the distribution of acritarchs. Data in part from Dorning KJ (1981) Silurian acritarch distribution in the Ludlovian shelf sea of South Wales and the Welsh Borderland. In: Neale JW and Brasier MD (eds.) *Microfossils from Recent and Fossil Shelf Seas*, pp. 31–36. Chichester: Ellis Horwood Ltd.

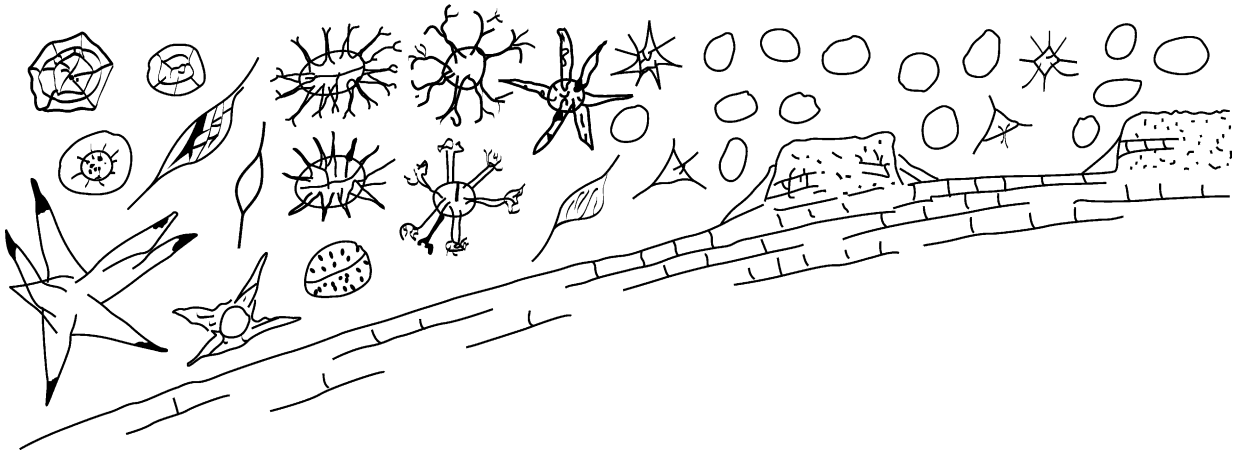


Figure 8 Diagrammatic section from inter-patch reef to off-reef shelf areas during the Late Wenlock, showing the distribution of acritarchs. Data in part derived from Dorning KJ and Bell DG (1987) The Silurian carbonate shelf microflora: acritarch distribution in the Much Wenlock Limestone Formation. In: Hart MB (ed.) *Micropalaeontology of Carbonate Environments*, pp. 266–287. Chichester: Ellis Horwood Ltd., and Dorning KJ and Harvey C (1999) Wenlock cyclicity, palynology, and stratigraphy in the Buildwas, Coalbrookdale, and Much Wenlock Limestone formations, Shropshire, England. *Bolletino della Societa Paleontologica Italiana* 38: 155–166.

(Figure 8). Reef and reef-associated sediments contain a high percentage but very low numbers of *Leiosphaeridia*. Inter-reef sediments contain much higher numbers of acritarchs, including many species with simple processes, such as *Micrhystridium*. Deeper water areas contain higher diversity shelf assemblages, with many species with branched processes, including *Oppilatala* and *Multiplicisphaeridium*. On a global scale, in the Silurian, cold-water high latitudes are characterized by assemblages with *Neoveryhachium*, large *Leiofusa*, and *Eupoikilofusa* in North Africa and Arabia, whereas warmer, low-latitude areas contain *Deunffia*, *Domasia*, and the large acritarchs *Estiastra*, *Hogklintia*, and *Pulvinosphaeridium*.

Applications

The organic-walled microplankton are the most abundant and diverse microfossil groups from the Late Precambrian through the Palaeozoic and Mesozoic, and are still important in the Cenozoic. Because the acritarchs are the principal phytoplankton in the Late Precambrian, Palaeozoic, and Early Mesozoic, they are the main biostratigraphical tool used in oil and gas exploration and production in marine sediments. Many forms have short stratigraphical ranges, which enable acritarch biohorizons and biozones to be established for the correlation of surface and borehole sections by analysing small samples of core, sidewall cores, or drill cuttings. Some acritarchs are associated with intervals of high total organic carbon, which are the main oil and gas source rock horizons. The colour of acritarchs may be used to estimate maximum palaeotemperatures, which can be employed to provide an indication of the maximum burial depth of rock sequences, in order to predict the timing of oil and gas generation. In regional studies, the distribution of acritarchs can be used to map areas of shallow to deep water as part of basin analysis. A combination of acritarch biostratigraphy, palaeotemperatures, and recycled acritarchs can be used to unravel the geological history of complex tectonic provinces.

See Also

Biosediments and Biofilms. **Microfossils:** Chitinozoa; Conodonts; Foraminifera; Ostracoda; Palynology. **Palaeoclimates**. **Precambrian:** Eukaryote Fossils; Prokaryote Fossils.

Further Reading

Aldridge RJ, Jeppsson L, and Dorning KJ (1993) Early Silurian oceanic episodes and events. *Journal of the Geological Society, London* 150: 501–513.

Commission Internationale de Microflore du Palaeozoique (CIMP) Acritarch Subcommission (from 1986) *Acritarch Newsletter*. www.shef.ac.uk/~cidmdp/ac16.doc

Dorning KJ (1981) Silurian acritarch distribution in the Ludlovian shelf sea of South Wales and the Welsh Borderland. In: Neale JW and Brasier MD (eds.) *Microfossils from Recent and Fossil Shelf Seas*, pp. 31–36. Chichester: Ellis Horwood Ltd.

Dorning KJ (1993) Group ('Phylum') Acritarcha Evitt, 1963. In: Benton MJ (ed.) *The Fossil Record 2*, pp. 33–35. London: Chapman and Hall.

Dorning KJ and Bell DG (1987) The Silurian carbonate shelf microflora: acritarch distribution in the Much Wenlock Limestone Formation. In: Hart MB (ed.)

Micropalaeontology of Carbonate Environments, pp. 266–287. Chichester: Ellis Horwood Ltd.

Dorning KJ and Harvey C (1999) Wenlock cyclicity, palynology, and stratigraphy in the Buildwas, Coalbrookdale, and Much Wenlock Limestone formations, Shropshire, England. *Bolletino della Societa Paleontologica Italiana* 38: 155–166.

Downie C (1973) Observations on the nature of the acritarchs. *Palaeontology* 16: 239–259.

Downie C (1984) Acritarchs in British Stratigraphy. *Geological Society of London, Special Report* 17: 1–26.

Downie C, Evitt WR, and Sarjeant WAS (1963) Dinoflagellates, hystrichospheres and the classification of the acritarchs. *Stanford University Publications (Geological Sciences)* 7: 1–26.

Fensome RA, Williams GL, Barss MS, *et al.* (1990) Acritarchs and fossil prasinophytes: an index to genera, species and infraspecific taxa. *AASP Contributions Series* 25: 1–771.

Jeppsson L, Aldridge RJ, and Dorning KJ (1995) Wenlock (Silurian) oceanic episodes and events. *Journal of the Geological Society, London* 152: 487–498.

Molyneux SG, Le Hérissé A, and Wicander R (1996) Palaeozoic phytoplankton. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications*, vol. 2, pp. 492–529. Salt Lake City, Utah: AASP Foundation.

Tappan H (1980) *Paleobiology of Plant Protists*. San Francisco: W. H. Freeman & Co.

Traverse A (1988) *Paleopalynology*. London: Unwin Hyman Ltd.

Chitinozoa

F Paris, University of Rennes 1, Rennes, France
J Verniers, University of Ghent, Ghent, Belgium

Introduction

Chitinozoans are strange and fascinating organic-walled microfossils. They appeared in the Early Ordovician (Tremadocian) and proliferated worldwide for 120 million years, until their extinction in the topmost Devonian (latest Famennian). The chitinozoans are exclusively recovered from marine deposits. They are usually abundant in sediments deposited under a low to moderate hydrodynamic regime. The biological significance of these microfossils and their systematic position have long been a matter of debate. The hypothesis that chitinozoans were eggs of soft-bodied metazoans is considered as the most likely.

Biological Aspects

Morphology and Structure

A chitinozoan vesicle corresponds to a small purse-, vase-, or flask-like structure formed by an organic wall, called a tegument or test. It is organized around a radial symmetry axis (Figure 1). The vesicle length ranges most frequently from 100 to a few hundred micrometres. Depending on the thickness of the wall, the weight of a single vesicle is between 10^{-3} and 10^{-5} mg. The chitinozoans are frequently linked along their longitudinal axis. In such chain-like structures, however, each vesicle does not communicate with its neighbours.

Basically, a chitinozoan comprises a chamber closed by a plug. The chamber, i.e., the bulging part of the vesicle, may be spherical, lenticular, ovoid, conical, cylindrical, bell-shaped, or claviform. A circular opening (aperture), frequently surrounded by a collarette, is located directly on the chamber or at the end of an apertural tube (neck) developing from the chamber. The plug is called an operculum (disc-like shape) when it seals the chamber directly (Figure 1B). It is named a prosome when it is situated within the neck, and corresponds to a cylinder segmented by horizontal septa (Figure 1A). The operculum and the prosome are extended anti-apperturally by the rica. This variously developed membranous expansion lines the upper part of the chamber. The vesicle wall and the plug act as a temporary protection for the inner contents.

At the junction between the flanks and the bottom of the chamber (i.e., the anti-apertural end) is the margin, whose shape and ornamentation (i.e., processes, spines, carina) are of primary importance in generic assignment. The ornaments of the vesicle display a large variety of size, structure, and shape. They may be either randomly or geometrically distributed (e.g., rows, rims, ridges) on part or on the whole vesicle. An anti-apertural mark occurs frequently at the apex of the chamber. This structure may be discrete (e.g., concentric rims, central scar, or even central pit) or more developed, forming the mucron (short membranous tube), the copula (long hollow tube), the siphon (bulb-like membranous expansion), or the peduncle (solid structure).

Investigations by transmission electron microscopy (TEM) and scanning electron microscopy (SEM) have demonstrated that the chitinozoan wall is composed

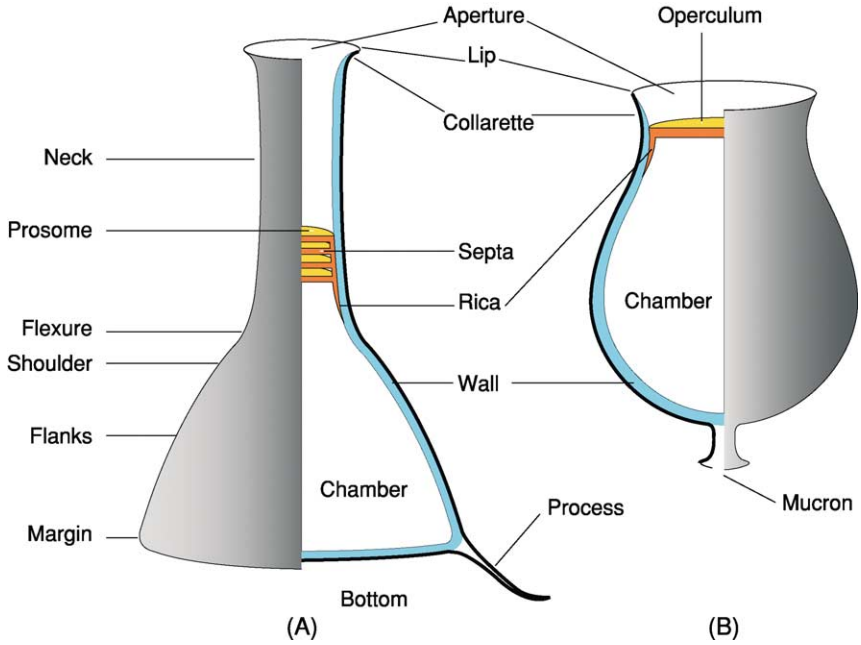


Figure 1 Diagram of a chitinozoan vesicle with morphological terminology: (A) Lagenochitinidae; (B) Desmochitinidae. Black, outer layer; blue, inner layer; orange and yellow, operculum and prosome. Adapted from Paris F (1981) Les Chitinozoaires dans le Paléozoïque du sud-ouest de l'Europe (cadre géologique – étude systématique – biostratigraphie). *Mémoire de la Société Géologique et Minéralogique de Bretagne* 26: 1–496.

of an inner and an outer layer. The actual chemical composition and molecular structure of this wall, however, are still unknown. The inner layer constitutes the frame of the vesicle (Figure 1). TEM observations have revealed neither ultrastructures, nor tiny pores. The inner layer is usually thicker than the outer layer and extends even under the central pit. Moreover, it survives longer than the outer layer when exposed to various physical (abrasion) and biological (e.g., fungi and/or bacterial attacks) degradations. No structures are observed within the chamber or on the internal surface of the wall.

The outer layer gives rise to the ornamentation through its outgrowth (e.g., verrucae, cones, granules, spines), evagination (e.g., hollow spines or processes), or folding (e.g., carina, crests, wrinkles). Its ultrastructure is highly variable (e.g., spongy, microlamellar, or perforated with microcanals). Thin-walled taxa (e.g., Ancyrochitininae) have a very thin outer layer (less than 1 μm), whose existence is mainly documented from scars on the hollow spines or processes. In thick-walled taxa (e.g., Desmochitininae), the outer layer itself can be rather thick (up to 10 μm), but fragile and easily removed. The aspect of the external surface of the vesicle is controlled by the texture of the outer layer (e.g., smooth, felt-like, spongy). The shape, size, density, and location and organization of the various ornaments of the vesicle are used as taxonomic criteria.

Intervesicle Adjustments

Probably all the chitinozoans had linked vesicles, at least at one time in their development. The less resistant connections were, however, destroyed during fossilization and/or processing. The most common vesicle adjustments are linear or chain-like structures. Planar aggregates are also known.

In linear catenary structures, the linkage between two successive vesicles occurs along the longitudinal axis. The simplest linkage, but also the most fragile and therefore less commonly observed (e.g., *Angochitina*, *Cyathochitina*; vesicles are usually free in palynological residues), resulted from close contact of the flaring lip of one individual with the convex anti-apertural part of the preceding vesicle (Figure 2A). Such an apparatus, in all probability, required a covering with mucilage. When the mucron, or the copula, is also in full contact with the operculum of the succeeding vesicle, the linkage is much more robust and the operculum frequently remains stuck to the preceding vesicle (e.g., *Cingulochitina*, *Urnochitina*) (Figure 2B). A more sophisticated stage is reinforced linkage, where important modifications of the vesicle are observed. The apex of the chamber extends towards the anti-apertural pole and covers the operculum of the succeeding vesicle. This is well illustrated in the genus *Margachitina*, where the operculum is fully part of the preceding vesicle (Figure 2C).

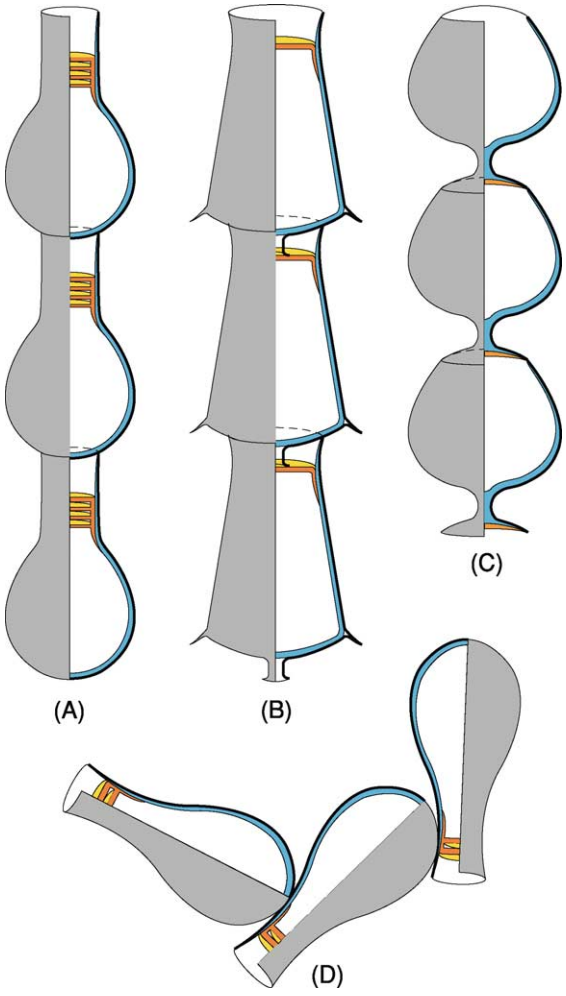


Figure 2 Schematic longitudinal cross-section illustrating the main linkage types in chitinozoans: (A) simple adherence; (B) reinforced linkage; (C) double linkage; (D) helicoidal catenary structure. Black, outer layer; blue, inner layer; orange and yellow, operculum and prosome. Adapted from Paris F (1981) Les Chitinozoaires dans le Paléozoïque du sud-ouest de l'Europe (cadre géologique – étude systématique – biostratigraphie). *Mémoire de la Société Géologique et Minéralogique de Bretagne* 26: 1–496.

Helicoidal catenary structures are less frequently observed. These chains appear normally as a coiled structure. However, due to compaction of the surrounding sediments, the original organization (several ten to several hundred linked vesicles) is frequently disturbed (folded or broken chains). Such structures have been reported in several genera (e.g., *Desmochitina*, *Bursachitina*, *Urnochitina*, *Margachitina*, *Pterochitina*). They are used for interpreting the actual biological meaning of the chitinozoans. Additional helicoidal catenary structures exist in which the linkage occurs by the adherence of an external part of the neck to part of the chamber of the following individual (e.g., *Lagenochitina navicula*). In this case, the aperture remains entirely free (Figure 2D).

Non-linear structures have been reported. This section regroups planar structures and aggregates. In planar structures, each vesicle is linked to a neighbour with an organic film. Geometrically arranged aggregates are known in *Desmochitina* species, in which vesicles are placed side-by-side, evoking an egg-laying, a cocoon, or the content of an egg capsule. In these structures, the aperture of every individual is completely free and its operculum is frequently absent. Chaotic aggregates (e.g., *Calpichitina*, *Cyathochitina)* usually correspond to individuals caught by organic matter. This arrangement, however, is not casual because it is always monospecific. When chaotic aggregates contain exclusively broken vesicles, they probably represent faecal pellets.

Teratological chitinozoans have been reported in chitinozoan assemblages, but they are extremely rare. They do not represent true intervesicle arrangements, but suggest aborted separations between two or several vesicles. For a time, they were interpreted as growing or rejuvenating stages. Biometric studies on large populations, however, support the existence of 'juvenile' chitinozoans.

Taxonomy and Classification

Progressively, a suprageneric classification, including orders, families, and subfamilies, and introducing a hierarchy amongst the morphological criteria available, has been adopted. Chitinozoans diversified rapidly throughout the Ordovician. Of the 56 usually accepted chitinozoan genera, 28% appeared in the Early Ordovician, 24% in the Middle Ordovician, and 11% in the Late Ordovician. Only seven genera became extinct at the end of the Middle Ordovician. The Ordovician biodiversification was accompanied by major morphological innovations (e.g., the appearance of the copula, carina, siphon, spiny ornaments, ornament in rows, processes). More than half of these innovations appeared before the Late Ordovician. During the Silurian, additional innovations appeared (e.g., the *Margachitina* linkage-type in the latest Llandovery; the reticulate chamber wall of *Pseudoclathrochitina*). Only a few changes occurred during the Devonian (e.g., the development of the solid and complex peduncle of *Urochitina* in the Lochkovian).

Suprageneric classification The order is the highest taxonomic subdivision in use for Chitinozoa. It is based upon the occurrence of an operculum in the Operculatifera and of the prosome (internal structure, not functional in the intervesicle linkage) in the Prosomatifera. The Operculatifera includes the unique Desmochitinidae family, whose representatives are

devoid of a neck but may have a collarette. Two families, i.e., the Lagenochitinidae (well-individualized neck) and Conochitinidae (without a shoulder, but with a deep prosome), are differentiated within the Prosomatifera. The third-order taxonomic subdivision is the subfamily, whose diagnostic criteria are regarded as reflecting evolutionary trends (e.g., differentiation of the outer layer). The Desmochitinidae include six subfamilies, the Conochitinidae seven, and the Lagenochitinidae six (Figure 3A–C).

Generic classification and specific criteria The generic assignments are based on the arrangement, location, and development of morphological features used as subfamily criteria. The outline of the vesicle, which is one of the most easily discernible elements by transmitted-light microscope, is also used. After emendations, restriction, and revision of the original diagnoses, only 56 genera from about 150 nominal ones are presently used.

Specific criteria for chitinozoans correspond to slight variations of the morphological parameters considered as generic criteria (e.g., length, density, location, shape of the ornamentation). From about 1120 species already described, only two-thirds seem to be workable taxa.

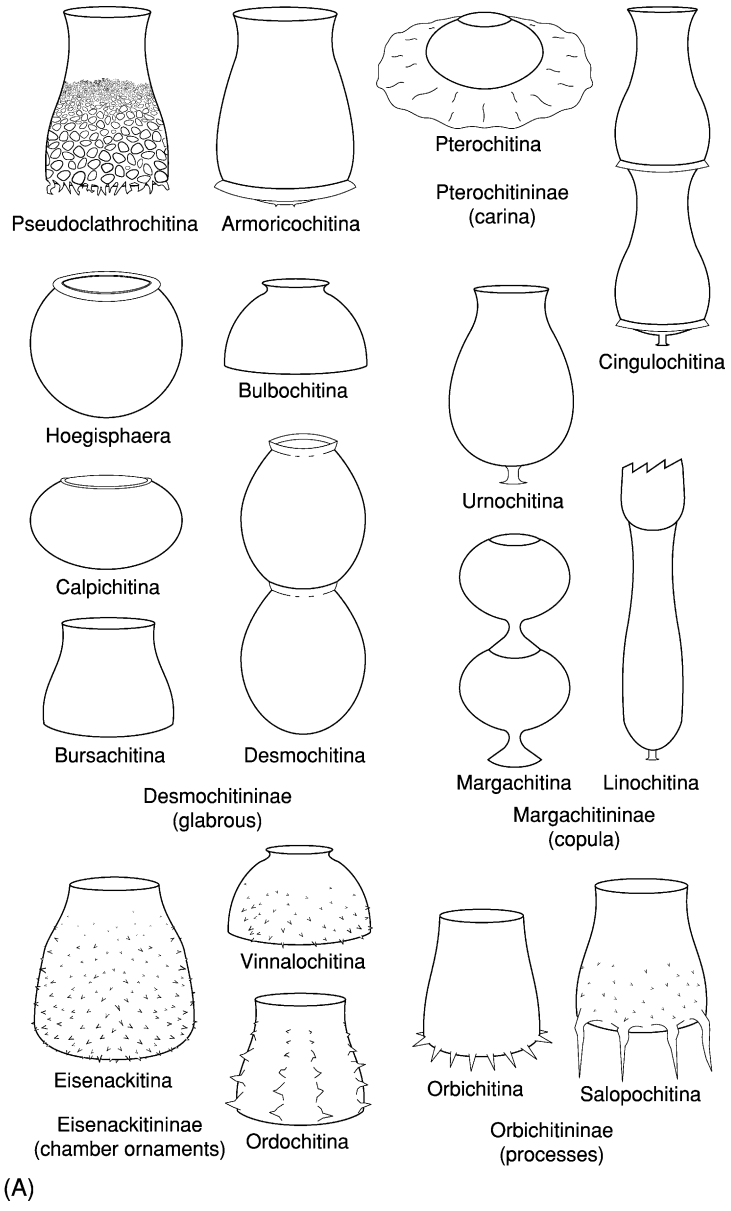


Figure 3 Outline of the most common chitinozoan genera: (A) Desmochitinidae; (B) Conochitinidae; (C) Lagenochitinidae.

Biological Interpretation and Systematic Position

The irritating question of the biological affinities of the chitinozoans has long been controversial. In general, successive proposals have rested on morphological features occurring in a small number of species belonging to the simplest shaped forms (i.e., *Desmochitina*, *Conochitina*, *Lagenochitina*). Chitinozoans have been successively assigned to protozoans (e.g., testaceans, rhizopods, tintinnids, flagellate protozoans), protists (e.g., dinoflagellates), and even to fungi ('Chitinomycetes'). However, all of these present-day groups show fundamental differences in structure, chemical composition, and environmental control from the chitinozoans. Moreover, a gap exceeding 100 million years separates the last chitinozoan occurrence from the first record of their possible, more modern analogues.

A more convincing hypothesis calls attention to the similarities between chitinozoans and eggs or egg capsules of diverse metazoans (e.g., annelids, gastropods (*see* **Fossil Invertebrates:** Gastropods), rotifers, and nematodes). The discovery of a complete coiled catenary structure preserved in three dimensions played the role of a 'Rosetta Stone' in providing the missing elements in this 'eggs theory'. This structure and the much more common fragments of chitinozoan chains are interpreted as an intraoviduct stage, and all of these catenary structures represent

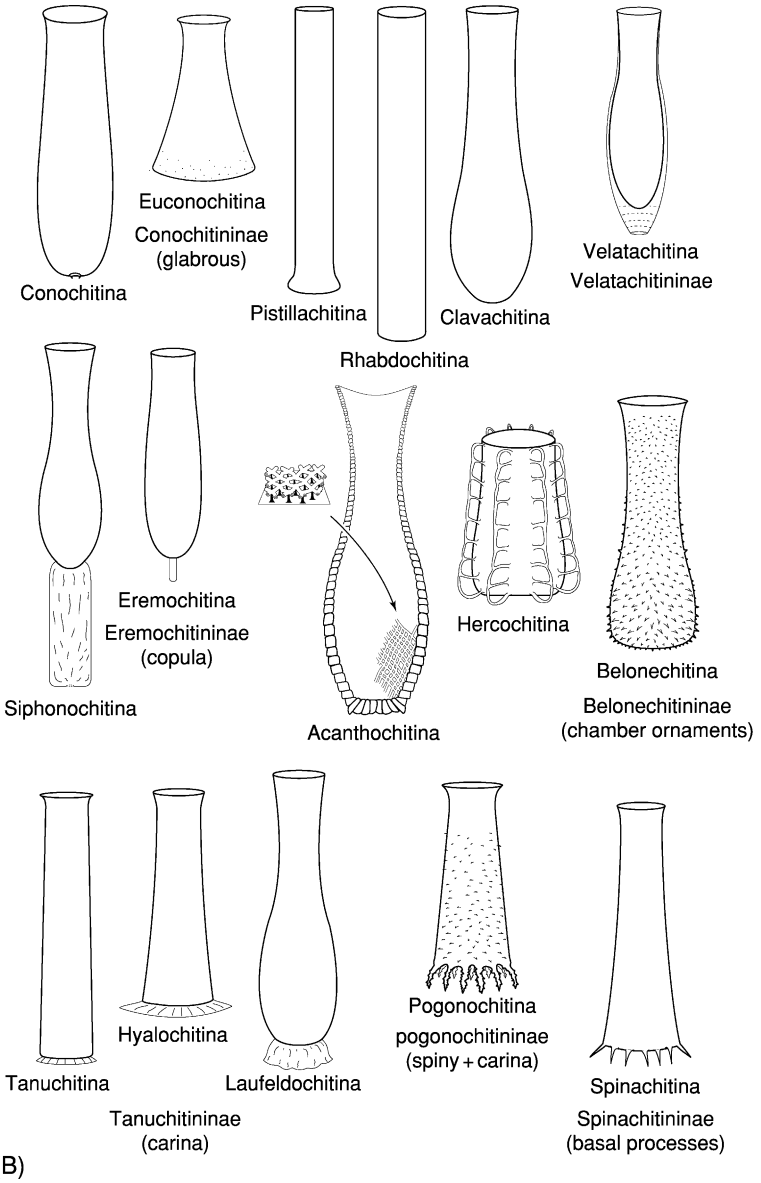


Figure 3 Continued

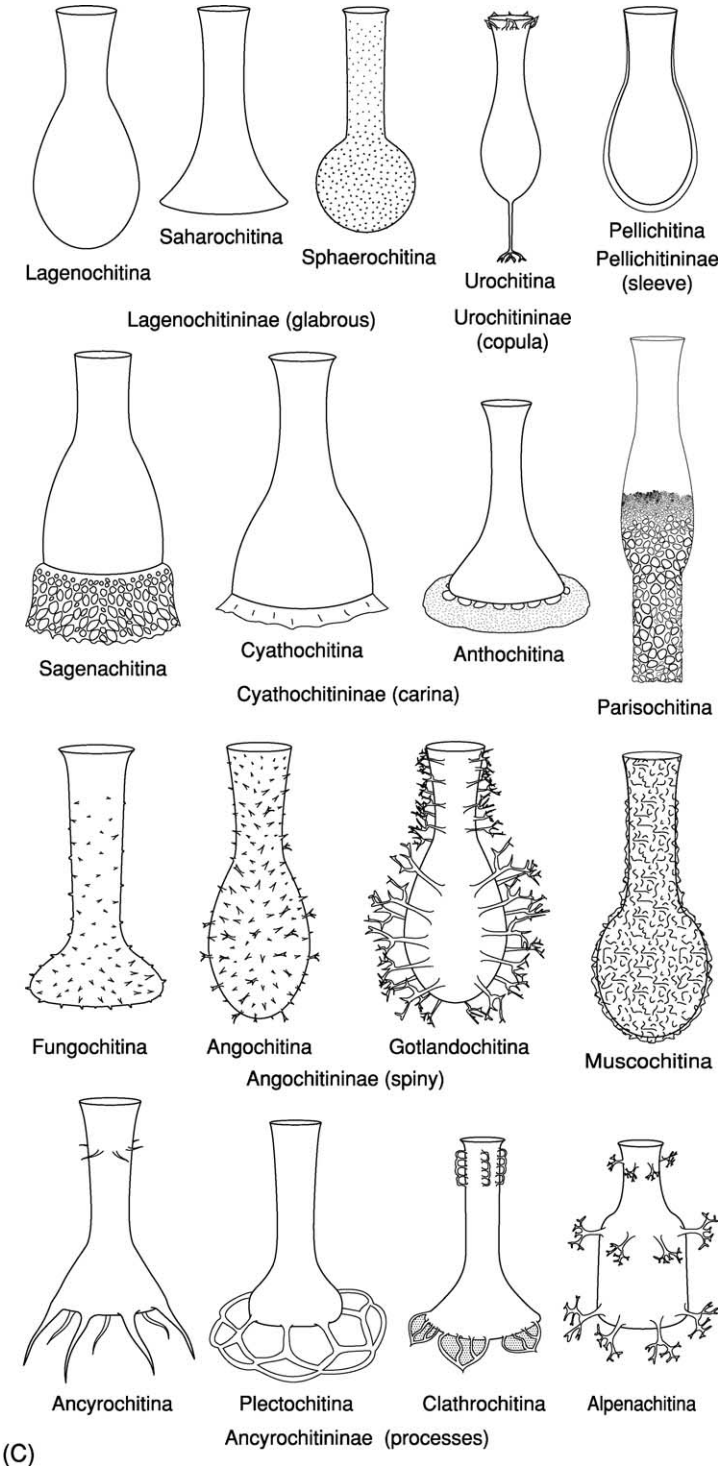


Figure 3 Continued

the release of immature eggs after the death and decay of the producing organism. The normal development of the chitinozoans included a maturation phase in the oviduct, prior to their laying in seawater, either as free isolated vesicles (e.g., *Pterochitina*, *Urnochitina*, Conochitinidae, Lagenochitinidae), as egg capsules, or as egg masses deposited through an ovipositor (e.g., *Desmochitina nodosa*). In all of these cases, each chitinozoan aperture was free. This is a prerequisite for the removal of the plug and the subsequent liberation of the chamber contents, i.e., a free larva, or a juvenile 'chitinozoan animal'.

With regard to the systematic affinities of the organisms producing the chitinozoans (i.e., the 'chitinozoan animals'), none of the body fossil groups usually recorded in Palaeozoic rocks display the same range as the chitinozoans. The tentaculitids, for instance, disappeared in the lowermost Famennian before the extinction of the chitinozoans in the topmost Famennian. The graptolites (*see* **Fossil Invertebrates:** Graptolites) appeared in the Cambrian, whereas the first chitinozoans are recorded in the Tremadocian. Moreover, the extinction of the monograptids in the Emsian had no consequences in the chitinozoan record. An extinct soft-bodied metazoan group is thus the most likely. According to the dimensions of the chitinozoans, these animals should be a few millimetres to 1 cm in length. These 'chitinozoans animals', with a worm-like soft body and a planktic or epiplanktic mode of life, were not preserved in normal conditions of fossilization. Some organic remains or discrete casts, however, probably survived fossilization in very quiet anoxic environments (e.g., 'konservat-Lagerstätten', black shale). Then, the 'chitinozoan animal' might be preserved as a carbonaceous film outlining a metazoan body with a flattened coiled chitinozoan chain inside. Attention should be paid, however, to avoid possible confusion with stomach and/or intestine contents, or with faecal pellets including chitinozoan remains.

Evolutionary Trends

If it is accepted that chitinozoan vesicles are the eggs of metazoans, their morphology is genetically controlled. This is critical for their biostratigraphical application. Amongst the examples of evolutionary trends, one of the most convincing is the progressive migration of the carina in the genus *Pterochitina*. This carina moved from an anti-apertural position in the Darriwilian (*P. retracta*) to a location below the equator of the chamber during the Llandovery (*P. deichaii*), then above this equator in the Pridoli (*P. perivelata*) and, finally, around the aperture in the Lochkovian (*P. megavelata*). Simultaneously, this carina developed significantly from the latest Ludlow to the mid-Lochkovian. Another obvious evolutionary trend is provided by a *Margachitina* lineage, from the latest Llandovery (*M. margaritana*) to the Pragian (*M. tenuipes*). Several morphological elements in this lineage present concomitant modifications affecting the chamber (from an ovoid to a lenticular shape), wall (development of thickened rims), peduncle (diameter progressively reduced), and operculum aperture adjustment (strengthening of the seal through a mortice-and-tenon assemblage system).

The widely held idea of a general shortening of the chitinozoan vesicles through time is not supported by the facts. The average smaller size of the vesicles recorded in Devonian samples is, in reality, due to an early extinction of the larger sized Conochitinidae in the Late Silurian (Figure 4). The Desmochitinidae and Lagenchitinidae themselves showed no general vesicle shortening from the Ordovician to the latest Devonian.

Chitinozoan Applications

Because of their rapid morphological changes through time (Figure 5), their records in a large variety of marine sedimentary rocks, their wide palaeogeographical distribution, and a fairly simple technical preparation, chitinozoans are amongst the most useful biostratigraphical fossil groups for Early Palaeozoic time. Chitinozoans, however, have a wider use in geology as they also provide information on palaeoenvironments, palaeoclimates, palaeobiogeography, and on the post-depositional history of the bearing rocks (e.g., diagenetic evolution, palaeothermometry, reworking events, and tectonic deformation).

Biostratigraphy

The biostratigraphical value of the chitinozoans has been progressively improved thanks to: (1) closely spaced sampling, with even bed-by-bed studies around the Global Stratotype Sections and Points (GSSPs); (2) better investigation techniques (e.g., sorting of the chitinozoans); (3) routine SEM observations; (4) more complete stratigraphical and geographical coverage of the sampled sequences and areas; (5) independent palaeontological controls provided by graptolites and/or conodonts; and (6) an improved taxonomy.

Global biozonations are now available for the Silurian and Devonian. For the Ordovician, due to palaeobiogeographical constraints, three regional biozonations are used.

The proposed chitinozoan biozones include interval range biozones and a few total range biozones. Most of the chitinozoan biozones benefit from direct chronostratigraphical calibration in the type sections of the GSSPs. Locally, and depending on the sedimentation, the chitinozoans allow high-resolution biostratigraphy, with a power of resolution approaching 0.5 million years for the most accurate biozones.

Ordovician regional biozones Because of the dispersion of the main palaeoplates and their highly contrasted latitudinal positions, three different biozonations have been defined for the Ordovician chitinozoans (Figure 6). The first represents the high-latitude assemblages from northern Gondwana regions. The second represents the assemblages from Baltica, which was located at moderate latitude

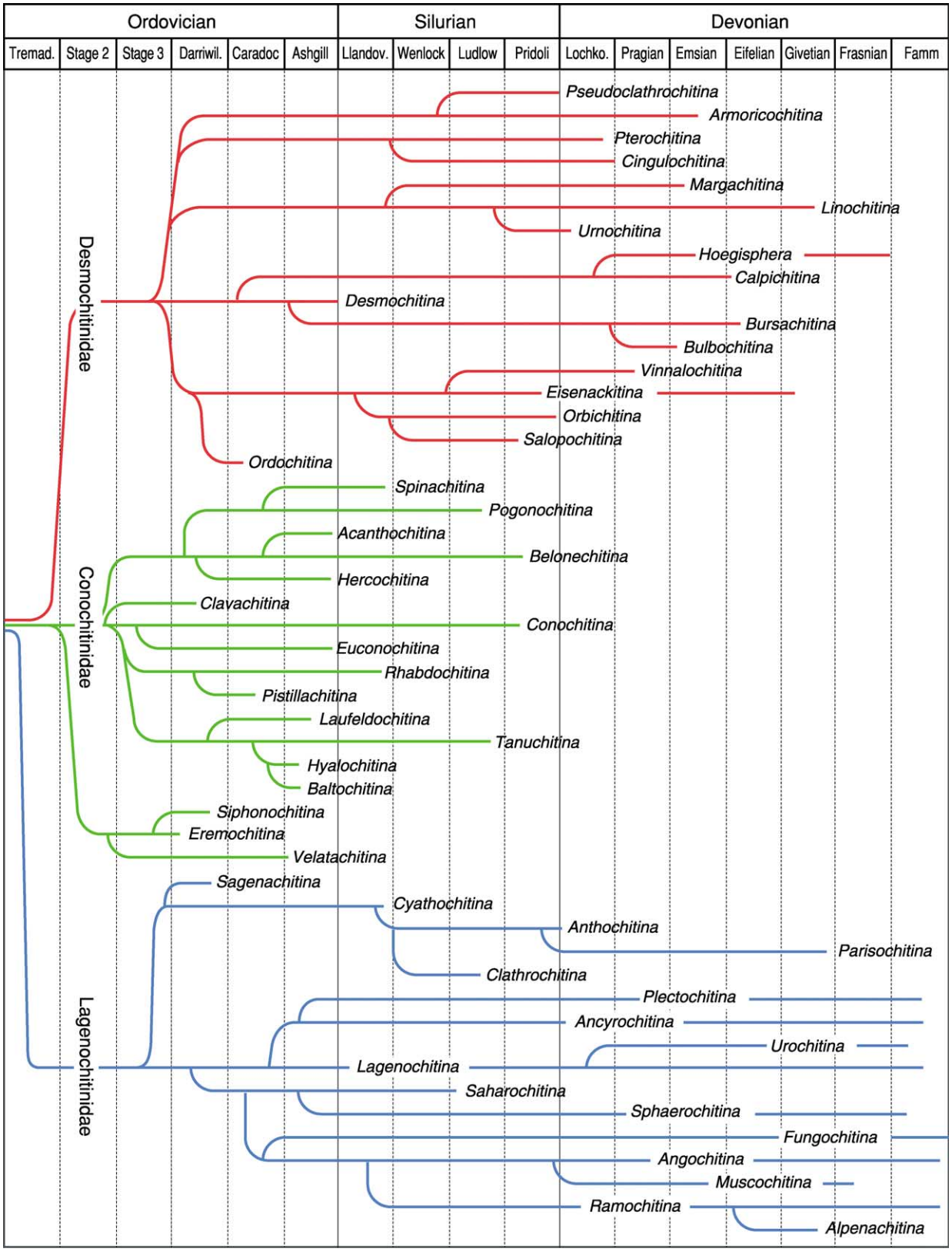


Figure 4 Morphological and temporal relationships between the chitinozoan genera.

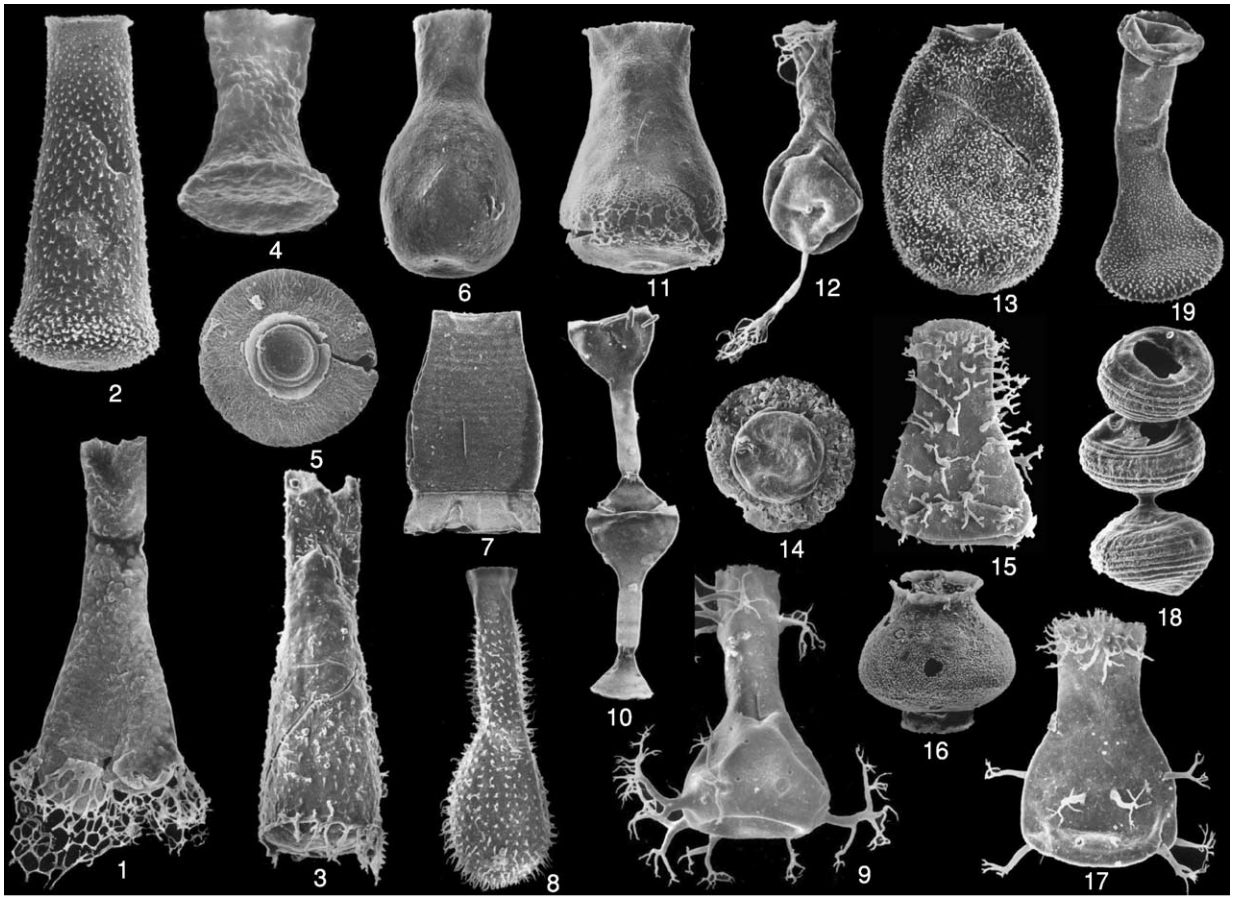


Figure 5 Scanning electron microscopy views of selected chitinozoan species. Ordovician species: 1, *Sagenachitina oblonga*; 2, *Belonechitina robusta*; 3, *Pogonochitina spinifera*; 4, *Euconochitina vulgaris*; 5, *Calpichitina lenticularis*; 6, *Lagenochitina baltica*; 7, *Armoricochitina nigerica*. Silurian species: 8, *Angochitina longicollis*; 9, *Ancyrochitina desmea*; 10, *Margachitina elegans*; 11, *Pseudoclathrochitina carmenchui*. Devonian species: 12, *Urochitina simplex*; 13, *Eisenackitina bohemica*; 14, *Pterochitina megavelata*; 15, *Ramochitina ramosi*; 16, *Bursachitina bursa*; 17, *Alpenachitina eisenacki*; 18, *Margachitina catenaria*; 19, *Fungochitina* sp.

during most of the Ordovician, and the third represents the low-latitude assemblages from Laurentia. A direct chronostratigraphical tie exists with the base of the Darriwilian in south China, and indirect calibration (through graptolites and conodonts) is available for most of the regional chitinozoan biozones.

The 25 Ordovician chitinozoan biozones presently identified in northern Gondwana are fairly regularly distributed, with five, nine, and 11 biozones in the Early, Middle and Late Ordovician, respectively. The average duration of one Ordovician chitinozoan biozone is about 2 million years. This duration is shorter (close to 1 million years) for the Middle Ordovician biozones. Eastern Avalonia, which left northern Gondwana in the Late Cambrian to earliest Ordovician, shared many chitinozoan species (including biozone index species) up to the Middle Ordovician.

A total of 22 chitinozoan biozones and subzones have been defined in Baltica, in spite of some hiatuses in the Ordovician chitinozoan record. The zonation is more accurate in the Late Ordovician, with an average duration of 1.2 million years per biozone. In contrast, only two chitinozoan biozones are so far available in the Early Ordovician of Baltica.

The Laurentian regional biozonation includes 19 biozones. Through successive improvements, the average duration of a Late Ordovician biozone in Laurentia is about 1.2 million years. The time resolution is much less precise in the Early and Middle Ordovician (average duration of about 3 million years per biozone).

Silurian global biozones Silurian chitinozoans are known from all the present-day continents, except Antarctica. The adopted Silurian global zonation includes 17 chitinozoans biozones (Figure 7). Most of these biozones benefit from direct chronostratigraphical calibration through the GSSPs of the Silurian Series in England and in the Czech Republic. Depending on the radiometric calibration accepted

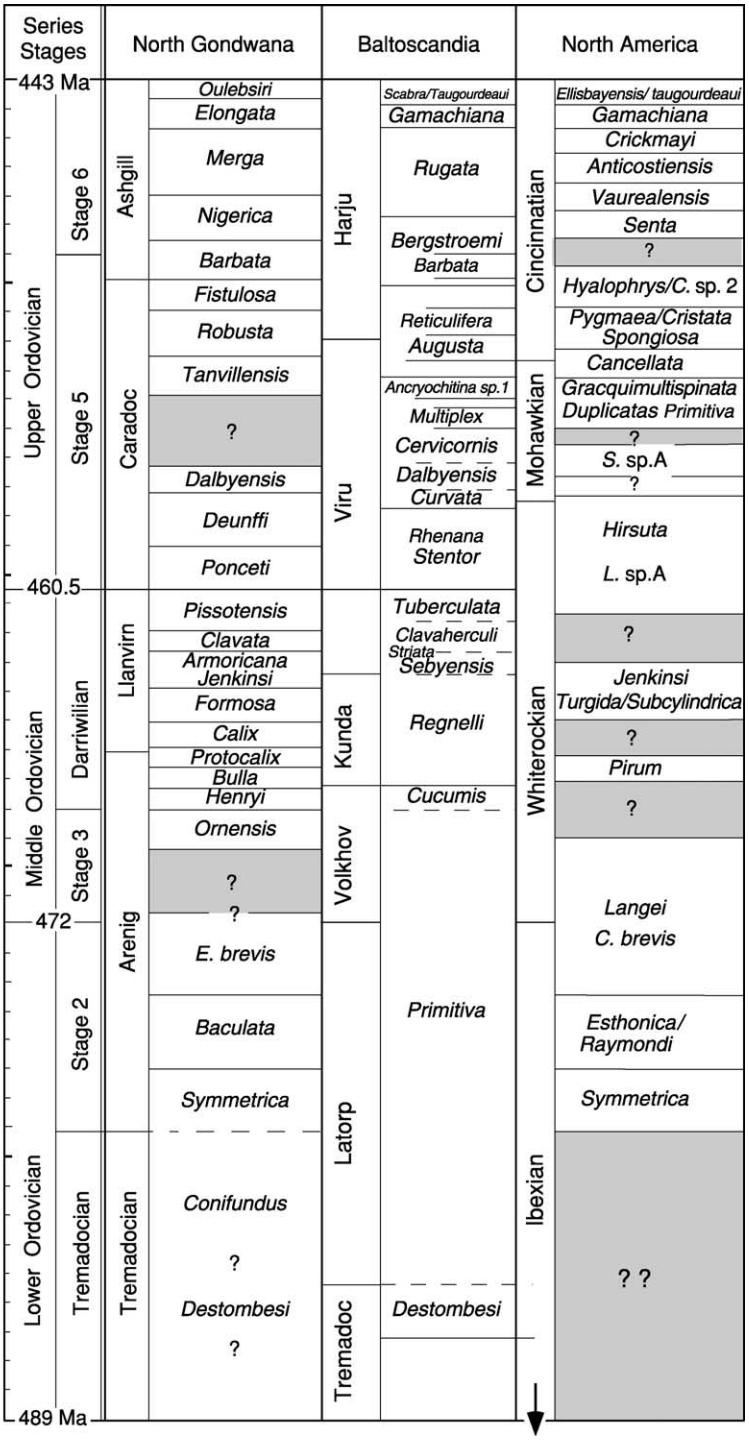


Figure 6 Ordovician chitinozoan biozones for northern Gondwana. Adapted from Webby, *et al.* (2004) Stratigraphic framework and time slices. In: Webby BD, *et al.* (eds.) *The Great Ordovician Biodiversification Event*, pp. 41–47. New York: Columbia University Press.

for the Silurian Series, the mean duration of a Silurian chitinozoan biozone is about 1.5–2 million years. A few time intervals are not yet characterized by a global chitinozoan biozone. In some cases, indirect age control of the biozone is available by means of graptolites. The present stage of the zonation will soon be improved as high-resolution biozonations become available at a regional scale (e.g., Llandovery and Wenlock of Estonia; Llandovery of Saudi Arabia).

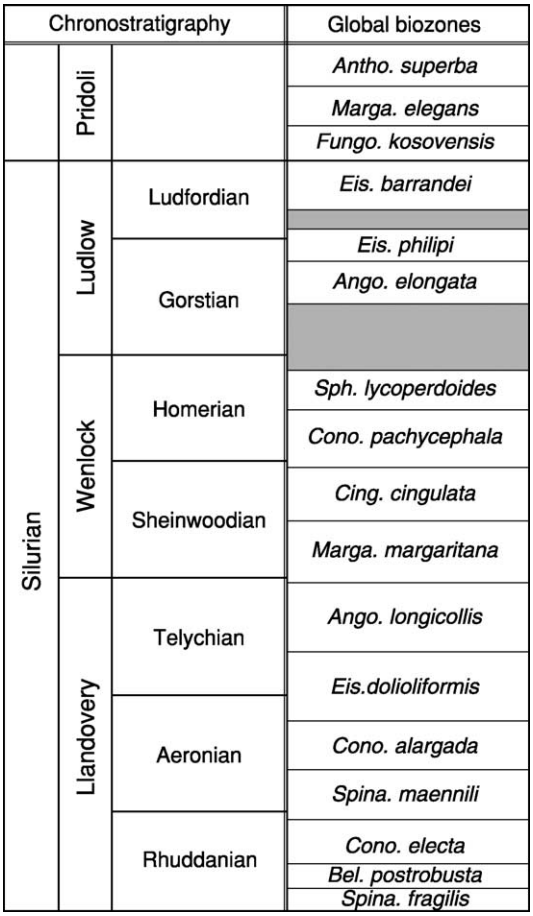


Figure 7 Silurian global chitinozoan biozones.

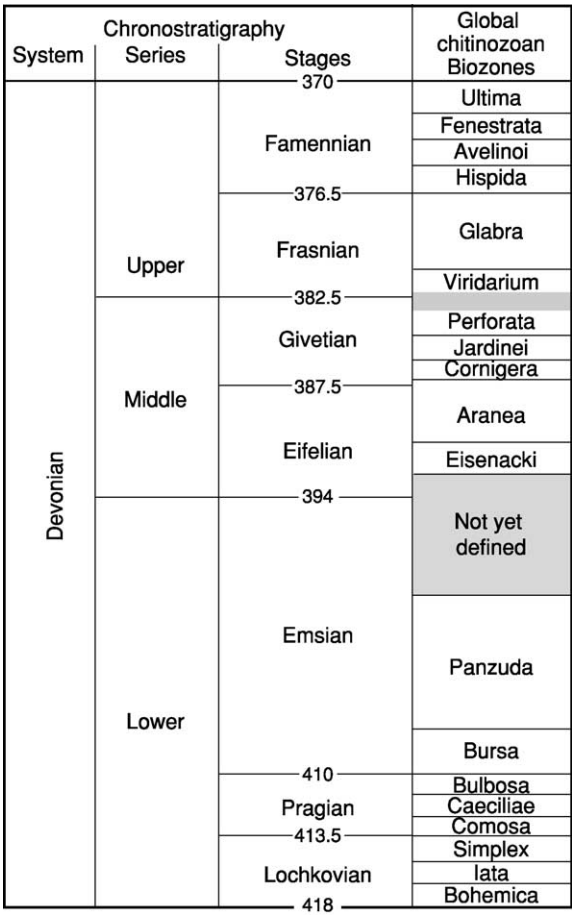


Figure 8 Devonian global chitinozoan biozones.

Devonian global biozones Most of the chitinozoan data available for the Devonian are from Gondwana (i.e., North Africa, southern and central Europe, South America, Australia) and from Laurussia. The present zonation, which benefits from a direct chronostratigraphical calibration with most of the GSSPs of the Devonian Stages (Czech Republic for the Lochkovian and the Pragian; southern France for the Frasnian and the Famennian), can be regarded as global. However, some endemic chitinozoan species have been recorded from Middle and Late Devonian strata in South America (Brazil and Bolivia). The Devonian chitinozoan zonation comprises 19 biozones, and two time intervals with no biozones yet defined (Figure 8). The average duration of one biozone is about 2.5 million years. Some high-resolution chitinozoan regional zonations are available (e.g., central Sahara, Brazil). They demonstrate that a more precise time slicing can be expected with the Devonian chitinozoans.

No undisputed chitinozoans have been recorded in the Carboniferous. The group apparently became extinct at the end of the Famennian, when *Fungochitina ultima* disappeared (Figure 4). However, this extinction may be an artefact of record as the 'chitinozoan animals' might have adopted a different mode of life (e.g., parasitism or non-marine). In both cases, dramatic changes might have occurred in the composition and structure of the wall of the chitinozoan vesicles, making fossilization no longer possible.

Palaeoenvironments

Chitinozoans disappeared at about 360 Ma. The assessment of the influence of environmental parameters on these microfossils therefore rests exclusively on indirect sedimentological and palaeoclimatological evidence, and on the ecology of associated organisms. Very little has been published on the environmental control of chitinozoans. Some authors have postulated a benthic life mode. However, many observations favour a planktic or epiplanktic mode of distribution for the chitinozoans as: (1) chitinozoans, which are exclusively marine microfossils, occur in

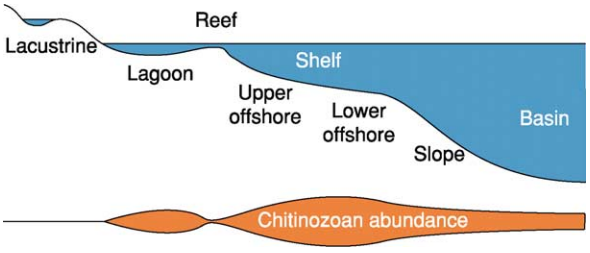


Figure 9 Diagram of the abundance of chitinozoans along a continent basin transect.

shallow-water deposits as well as in shelf and slope sediments; (2) chitinozoans are present and often very abundant (up to several thousand specimens per gram of rock) in marine anoxic or epoxic deposits devoid of benthic or endofaunas (e.g., Silurian black shales); (3) distal deposits (outer shelf and slope) yield abundant chitinozoans (several hundred to several thousand specimens per gram of rock; Figure 9), whereas acritarchs and/or miospores are rare or virtually lacking; (4) the geographical distribution of many chitinozoan species is frequently much wider than that of benthic or neritic organisms; and (5) some chitinozoan species extended throughout all the palaeoclimatic belts. Thus, the environmental control prevailing for pelagic organisms should also affect the chitinozoan vesicles, as well as the 'chitinozoan animals', if the 'egg theory' is accepted. This control is assessed through a variety of parameters, such as: (1) chitinozoan abundance (expressed as the number of specimens per gram of rock); (2) chitinozoan diversity (number of species and genera recorded per sample); (3) the relative frequency of the represented taxa; and (4) the chitinozoan ratio in palynological residues containing marine phytoplankton (acritarchs) and to land-derived elements (miospores, tracheid fragments). All of these numerical data are closely related to the production of the chitinozoans (nutrients, temperature), predation pressure, oxidation of organic matter, hydrodynamic constraints, and the fossilization processes.

Palaeobiogeography

Chitinozoans display a rather discrete provincialism compared with benthic Palaeozoic faunas. Most of the chitinozoan genera and many species have a worldwide distribution, comparable with the palaeogeographical distribution of graptolites or conodonts. However, the peculiar palaeogeographical context prevailing during the Ordovician (i.e., wide dispersion and contrasting latitudinal positions of the main palaeoplates; a very large ocean occupying most of the northern hemisphere) favoured a weak provincialism of the chitinozoans at a generic level, as illustrated by the restricted distributions of *Eremochitina*, *Velatachitina*, *Pogonochitina*, *Siphonochitina*, and *Sagenachitina*. These genera have been considered as typical northern Gondwana taxa but, probably due to the southern hemisphere gyre, some of these genera are now also reported in south China. *Hercochitina* was long believed to be an endemic genus from Laurentia, but recent investigations have revealed its occurrence in the Middle Ordovician of the northern Gondwana regions. As the record is being completed, the number of endemic genera amongst the Ordovician chitinozoans is decreasing. At species level, the geographical control is more effective. This has led to three different chitinozoan biozonations for low (Laurentia), medium (Baltica), and high latitudes (northern Gondwana), respectively. A few species only, usually corresponding to long-ranging and poorly discriminated taxa, display a more or less global distribution (e.g., *Cyathochitina campanulaeformis*, *Desmochitina minor*, *Belonechitina micracantha*, and *Lagenochitina baltica*). It is worth noting that the specific diversity of the Ordovician chitinozoan is slightly higher in the subequatorial assemblages (about ten species per sample) than in the high-latitude ones.

The Silurian and Devonian were times of convergence of the main palaeoplates. This is reflected by the distribution of the chitinozoan species. The weaker geographical control during these epochs allowed the appearance of chitinozoan biozones with global value. However, some regions displayed sporadic endemism at species level. The most obvious case concerns the assemblages recorded in the very large marine gulf which extended temporarily over westernmost Africa, Brazil, and Bolivia. These assemblages include a number of endemic species in the Early Silurian (e.g., *Pogonochitina djalmai*), Middle Devonian (e.g., *Ancyrochitina biconstricta*), and Late Devonian (e.g., *Sommerochitina langei*). These endemic species, however, coexisted with more widespread taxa, indicating communication with the global ocean.

Miscellaneous

Chitinozoans registered certain chemical parameters of the environment. The $\delta^{13}C$ values measured on carefully sorted chitinozoan vesicles show that the carbon of their wall is in equilibrium with the carbon reservoir of the ocean. Therefore, chitinozoans can be used for an accurate record of the $\delta^{13}C$ variations through time. At high latitudes, where carbonates are missing, they represent an alternative carbon source as they have an almost continuous record.

Like other organic-walled microfossils, chitinozoans registered the thermal events affecting the host sediment. A temperature increase led to a progressive darkening of their wall. This corresponds to a new molecular equilibrium (i.e., maturation of the organic matter by a progressive increase in the carbon ratio). Such changes can be measured through the reflectance of the vesicle wall, which provides a precise geothermometer for the early diagenetic to very low-grade metamorphic history of the host sediment.

Because the organic wall of the chitinozoan vesicles remained elastic for a long time, they were susceptible to collapse during compaction of the sediments, especially when the interstitial water escaped from argillaceous deposits. Vesicles preserved in full relief indicate an early lithification prior to compaction in, for example, cherts, limestones, phosphates, and ironstones. This phenomenon can be used to detect reworking in argillaceous deposits.

See Also

Biozones. **Fossil Invertebrates:** Graptolites; Gastropods. **Lagerstätten**. **Palaeoecology**. **Palaeozoic:** Ordovician; Silurian; Devonian.

Further Reading

Achab A (1988) Mise en évidence d'un provincialisme chez les chitinozoaires ordoviciens. *Canadian Journal of Earth Sciences* 25: 635–638.

Achab A (1989) Ordovician chitinozoan zonation of Québec and western Newfoundland. *Journal of Paleontology* 63: 14–24.

Combaz A, Calandra F, Jansonius J, *et al.* (1967) *Microfossiles Organiques du Paléozoïque. Les Chitinozoaires (2): Morphographie.* Paris: Centre National de la Recherche Scientifique.

Eisenack A (1972) Beitrage zur chitinozoen forschung. *Palaeontographica A* 140: 117–130.

Jansonius J (1970) Classification and stratigraphic application of Chitinozoa. In: *Proceedings of the North American Paleontologists' Convention,* 1969, vol. G, pp. 786–808. Lawrence.

Jenkins WAM (1970) Chitinozoa. *Geoscience and Man* 1: 1–21.

Kozlowski R (1963) Sur la nature des Chitinozoaires. *Acta Palaeontologica Polonica* 8: 425–449.

Laufeld S (1974) Silurian Chitinozoa from Gotland. *Fossils and Strata* 51: 1–130.

Miller MA (1996) Chitinozoa. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications,* vol. 1, pp. 307–336. Dallas, TX: American Association of Stratigraphic Palynologists' Foundation.

Paris F (1981) Les Chitinozoaires dans le Paléozoïque du sud-ouest de l'Europe (cadre géologique – étude systématique – biostratigraphie). *Mémoire de la Société Géologique et Minéralogique de Bretagne* 26: 1–496.

Paris F (1990) The Ordovician chitinozoan biozones of the northern Gondwana Domain. *Review of Palaeobotany and Palynology* 66: 181–209.

Paris F (1996) Chitinozoan biostratigraphy and palaeoecology. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications,* vol. 2, pp. 531–552. Dallas, TX: American Association of Stratigraphic Palynologists' Foundation.

Paris F, Grahn Y, Nestor V, and Lakova I (1999) A revised chitinozoan classification. *Journal of Paleontology* 73: 547–568.

Paris F and Nõlvak J (1999) Biological interpretation and paleobiodiversity of a cryptic fossil group: the 'chitinozoan-animal'. *Geobios* 32: 315–324.

Paris F, Winchester-Seeto T, Boumendjel K, and Grahn Y (2000) Toward a global biozonation of Devonian chitinozoans. *Courier Forschung-Institute Senckenberg* 220: 39–55.

Verniers J, Nestor V, Paris F, *et al.* (1995) A global Chitinozoa biozonation for the Silurian. *Geological Magazine* 132: 651–666.

Webby, *et al.* (2004) Stratigraphic framework and time slices. In: Webby BD, *et al.* (eds.) *The Great Ordovician Biodiversification Event,* pp. 41–47. New York: Columbia University Press.

Conodonts

R J Aldridge, University of Leicester, Leicester, UK

Introduction

Conodonts are a group of extinct marine animals that were entirely soft-bodied, except for an apparatus of tooth-like elements, composed of calcium phosphate and situated in the mouth and/or the pharynx. These elements normally became scattered in the sediment on the seafloor after the death and decomposition of the animals that bore them. The elements are usually microscopic, 0.2–2 mm in size, but rare larger specimens reach up to 25 mm in length. They display a range of morphologies, from coniform shapes through denticulated bars and blades to highly ornamented plates (**Figure 1**). The true conodonts have a stratigraphical range of Late Cambrian to Late Triassic and are important biostratigraphical indices for

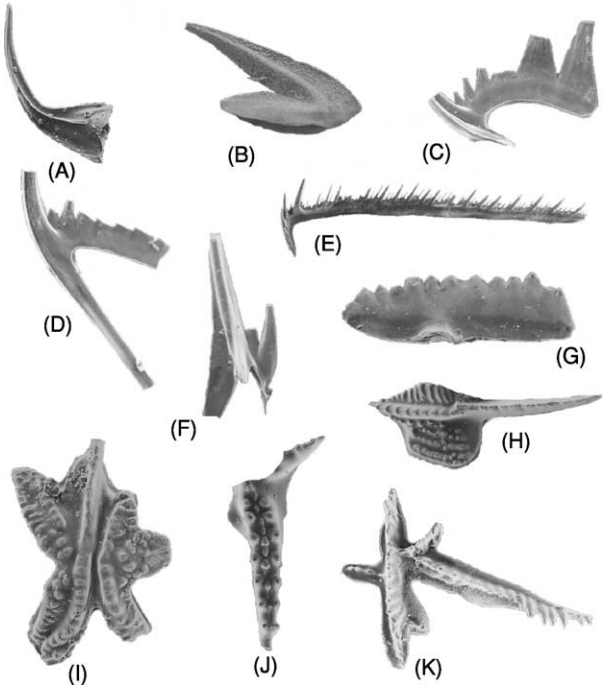


Figure 1 A vignette of conodont elements to illustrate the variety of morphologies. (A) Coniform element of the genus *Dapsilodus*, ×25, Silurian. (B) Coniform element of the genus *Drepanoistodus*, ×40, Ordovician. (C) Ramiform (alate S) element of *Gamachignathus*?, ×45, Silurian. (D) Ramiform (dolabrate S) element of *Gamachignathus*?, ×45, Silurian. (E) Ramiform (bipennate S) element of *Idiognathodus*, ×30, Carboniferous. (F) Ramiform (tertiopedate S) element of *Gamachignathus*?, ×40, Silurian. (G) Pectiniform P element of *Ozarkodina*, ×20, Silurian. (H) Pectiniform P element of *Gnathodus*, ×25, Carboniferous. (I) Pectiniform P element of *Aulacognathus*, ×20, Silurian. (J) Pectiniform P element of *Icriodus*, ×25, Devonian. (K) Pectiniform P element of *Eoplacognathus*, ×25, Ordovician.

marine strata of basal Ordovician to end-Triassic age. In addition, their colour is related to the temperatures that have been subsequently experienced by the rocks in which they are found and can be used in the investigation of the thermal history of sedimentary basins.

Fossils that preserve evidence of the soft tissues of conodonts are extremely rare. Research on these specimens, and on the internal microstructure of the phosphatic elements, has provided evidence that conodonts were chordates, although hypotheses regarding their precise relationships are the subject of continuing debate. The conodont apparatus itself was used to capture and process food, and most evidence suggests that the conodonts were active, macrophagous vertebrates.

Anatomy of the Animals

Conodont fossils preserving features of the soft tissues of the head and trunk are known from only two places in the world. The Upper Ordovician Soom Shale of Cape Province, South Africa, has yielded a single specimen that displays structures of the head and of the anterior portion of the trunk (Figure 2), as well as several tens of specimens that preserve paired head structures. The Lower Carboniferous Granton Shrimp Bed of Edinburgh, Scotland, has provided ten specimens, one of which is more or less complete from head to tail; the others all show a portion of the trunk, together with features of the head or of the tail (Figure 3).

The soft tissue features of the Scottish animals have been preserved through replacement by calcium

Figure 2 Conodont specimen with preserved soft tissues, genus *Promissum*, Upper Ordovician Soom Shale Member, Sandfontein, Cape Province, South Africa. Specimen C721b, Geological Survey of South Africa, Pretoria; scale bar, 10 mm. The conodont apparatus is at the lower right of the specimen with the position of the eyes preserved in white above and just to the right of the apparatus.

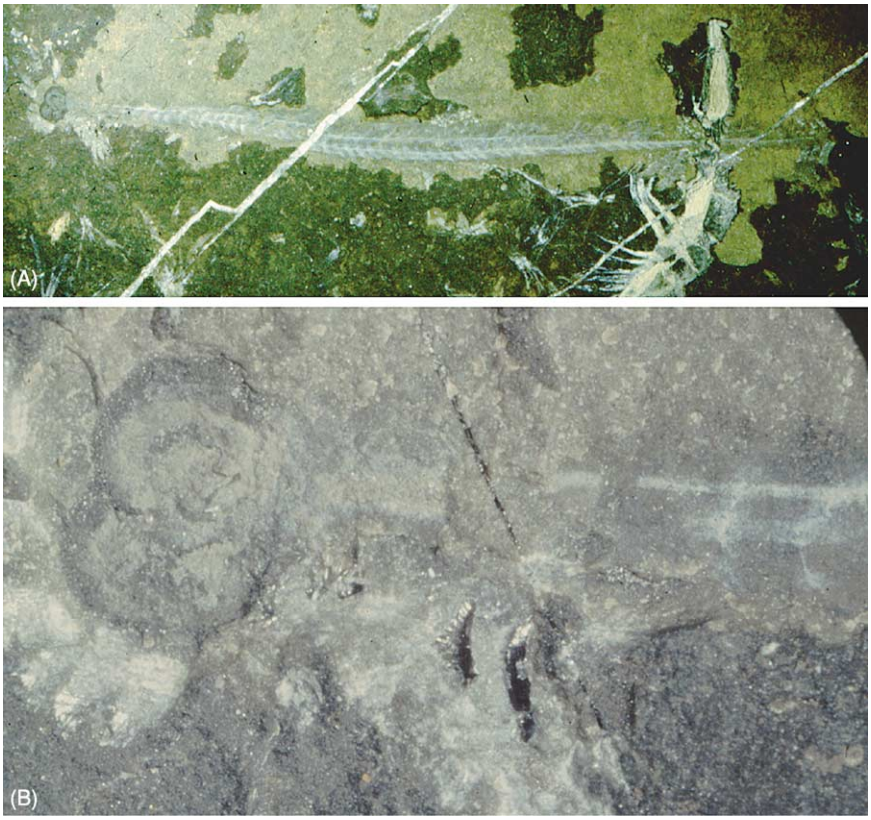


Figure 3 Conodont specimen with preserved soft tissues, genus *Clydagnathus*, Lower Carboniferous Granton Shrimp Bed, Edinburgh, Scotland. Specimen RMS GY 1992.41.1, Royal Museums of Scotland, Edinburgh, UK. (A) Complete specimen, anterior to left, ×2.3. (B) Anterior portion, showing the feeding apparatus (partly buried in the rock matrix), the eye capsules, and the anterior portion of the trunk, ×15.

phosphate, which forms a film on the bedding surface. The replacement was probably selective, and not all tissues and structures have been preserved. However, the animals can be seen to be elongate, 40–55 mm in length and less than 2 mm in height. They have a short head with a pair of ring-like structures that have been interpreted as representing the eyes. The skeletonized feeding apparatus, comprising 15 elements of various morphologies, is situated below and just behind the eyes; there is little evidence preserved of the soft tissue that housed the apparatus. The body traces show a distinct pair of subparallel longitudinal lines, which appear to represent a collapsed notochord, a precursor of the backbone. V-shaped muscle blocks (myomeres), with the apex anterior, are evident along the entire length of the trunk; it is not impossible that the myomeres were originally W-shaped, but that only the central V is preserved. Two of the specimens have an asymmetrical bilobed caudal fin with supporting fin rays. The fin rays are unbranched, set closely together, and there is no evidence for supporting musculature; the bilobed portion appears to occur on the dorsal margin. Two specimens show a possible dorsal nerve cord that stops short of the eyes, and the most complete individual shows faint traces of possible auditory capsules and gill pouches behind the eyes.

The specimens from South Africa are unusually large, and belong to a different order of conodonts, the Prioniodontida, from that represented by the fossils from Scotland (the Ozarkodinida). The most complete shows the anterior portion of an animal, with eyes, a feeding apparatus, and 100 mm of preserved trunk; the soft tissues are preserved through replacement by clay minerals. As in the Scottish specimens, the apparatus occupies a position ventral to and slightly posterior to the eyes; in this case, it comprises 19 individual elements up to 10 mm in length. The trunk shows a sequence of V-shaped myomeres, with the line of the notochord probably represented by a longitudinal gap in clay mineral replacement at the apices of the muscle chevrons. Under high magnification, high-fidelity preservation of original muscle fibres and fibrils can be seen. There is also a black organic patch within the mid-trunk region that may be the remains of a visceral organ. The body is nearly 20 mm in height; if the linear proportions are

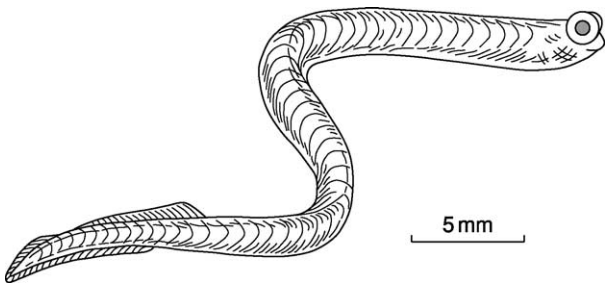


Figure 4 Reconstruction of a conodont animal, based on the specimens from Scotland and South Africa.

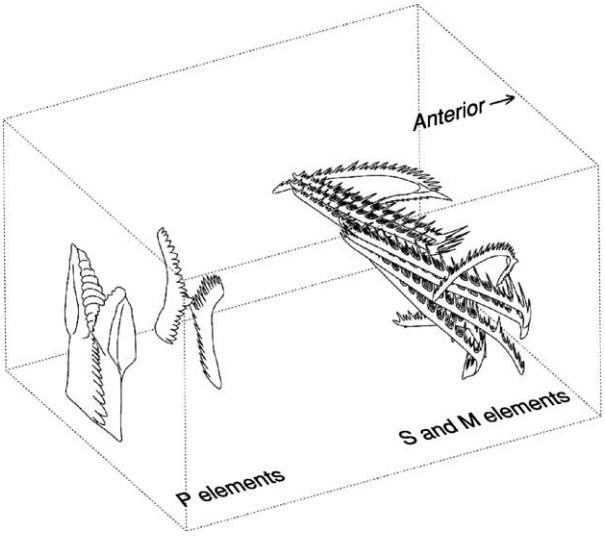


Figure 5 Architecture of an ozarkodinid conodont apparatus. Reproduced with permission from Aldridge RJ and Purnell MA (1996) The conodont controversies. *Trends in Ecology and Evolution* 11: 463–468.

comparable with those of the Scottish animals, then the entire length of this individual may have been as much as 400 mm.

Several tens of specimens from the Soom Shale preserve just the apparatus and the eyes. From these it can be determined that the eye capsules formed outwardly expanding cups, not much larger than those of the Scottish specimens. In a few specimens, including the most complete, the position of the eyes is occupied by fibrous tissue, which probably represents extrinsic eye musculature.

The evidence from these rare fossils permits a broad reconstruction of the morphology of a living conodont (Figure 4). A single, very poorly preserved, specimen with soft tissue is also known from the Silurian strata of Wisconsin, USA. This individual is from an order of conodonts, the Panderodontida, that possessed an apparatus of coniform elements, and the indistinct remains hint at a rather broader, flatter morphology for this type of animal.

Morphology and Internal Structure of the Elements

The more primitive conodonts possessed oro/pharyngeal skeletal apparatuses that consisted of an array of coniform elements. These elements vary within and between apparatuses in characteristics such as the degree of curvature, presence or absence of surface striations, and the development of costae. More derived conodont groups possessed elements of more complex morphology, with greater morphological distinction within each apparatus. Animals belonging to the most derived conodont order, the Ozarkodinida, bore a bilateral apparatus of 15 elements, which have been differentiated into two sets: a rostral array of nine denticulate, ramiform 'S' elements flanked by a pair of pick-shaped 'M' elements; behind these are two pairs of robust 'P' elements (Figure 5). The prioniodontid apparatuses from the Soom Shale differ in having four pairs of P elements that are situated above the S elements and behind the M elements, but the ozarkodinid 15-element pattern seems to apply to the vast majority of complex conodont apparatuses.

Conodont elements are composed of two structures: the crown and the basal body. The crown is heavily mineralized and relatively coarsely crystalline, and its upper surface may be elaborated into complex denticulated processes or ornate platforms. The basal body is finely crystalline with a higher content of organic material. The two components grew by the external apposition of layers of calcium phosphate. In many derived conodont elements, the basal body is either loosely attached or unmineralized and the crown has an open basal cavity. The tip of the basal cavity marks the origin of growth of the element.

In categorizing the morphology of individual conodont elements, the curvature of the cusp is taken conventionally as pointing towards the 'posterior', although this does not always correspond to the true biological orientation of the element. The cusp is defined as the denticle immediately above the tip of the basal cavity, and it is often the most prominent denticle of an element. Ramiform, or bar-like, elements are categorized by the disposition of processes (Figure 6), which radiate from the cusp and are commonly denticulate, but may be adenticulate: alate elements are bilaterally symmetrical, with a posterior process and a lateral process on each side; tertiopedate elements are asymmetrical with a long posterior process and a lateral process on each side; digyrate elements are asymmetrical, with a lateral process on each side, but normally with a very weakly developed

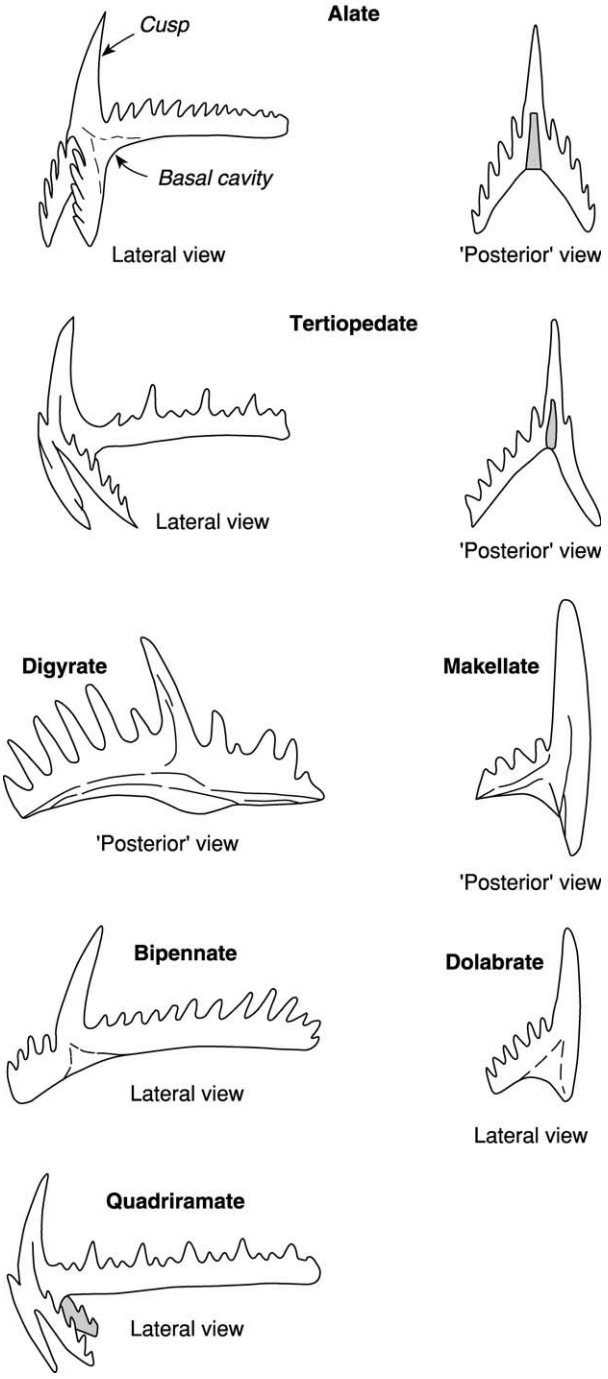


Figure 6 Morphological categories of ramiform conodont elements.

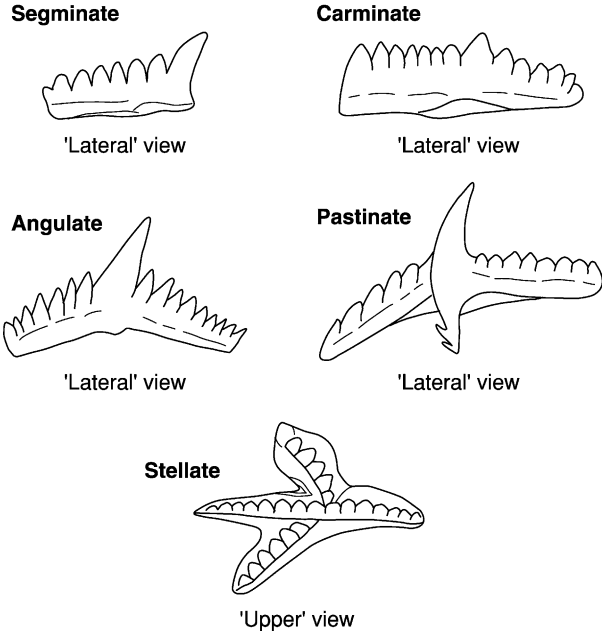


Figure 7 Morphological categories of pectiniform conodont elements.

posterior process; makellate elements have one lateral process and a downwardly directed anticusp on the other side; bipennate elements have anterior and posterior processes; dolabrate elements have only a posterior process, but there may be an extended anticusp below the cusp; quadriramate elements have four processes. Apart from the alate elements, each element is asymmetrical in itself, but left and right forms can be recognized that represent mirror-image pairs within the apparatus.

The P positions in complex apparatuses are often occupied by pectiniform elements, a term that encompasses blade-like morphologies and more complex platform or blade-plus-platform shapes. Pectiniform elements are accommodated in five major categories (Figure 7): segminate elements have only one process, conventionally termed anterior; carminate elements have anterior and posterior processes and are essentially straight in lateral view; angulate elements have anterior and posterior processes and are arched in lateral view; pastinate elements have three primary processes, anterior, posterior, and lateral; stellate elements have at least four primary processes.

The crown of conodont elements is composed of lamellar tissue, which may contain areas of opaque tissue known as white matter. The lamellar crown tissue is composed of crystallites that are typically a few micrometres in length, forming layers that are bounded by incremental growth lines. White matter is more finely crystalline, and is rendered opaque by the inclusion of small pores; it was deposited synchronously with the lamellar tissue. The basal body is formed of basal tissue, which normally shows incremental growth lines, but may be internally globular and/or tubular or may lack both globules and tubules. Most post-Devonian taxa lack biomineralized basal tissue.

Some authors have compared the lamellar crown tissue of conodont elements to the enamel of vertebrate teeth and scales, and the basal tissue to dentine. Other commentators have questioned whether these conodont and vertebrate tissues are truly homologous. White matter is certainly a type of tissue unique to conodonts.

Biological Affinity of Conodonts

The presence of chevron-shaped myomeres and a notochord indicates that conodonts belong to the chordates. Gill slits have not been recognized to date, but they would lie in the antero-ventral region where soft tissues have not been preserved in the fossils. Within the chordates, the presence of paired sensory organs (eyes), extrinsic eye musculature, a ray-supported caudal fin, a bilateral feeding apparatus, phosphatic skeletal elements, enamel, and dentine all suggest an assignment to the vertebrates, rather than to the more primitive cephalochordates, exemplified by the living lancelet *Branchiostoma*. The termination of the notochord anteriorly before the feeding apparatus is also a vertebrate feature. Some authors, however, prefer to place conodonts in the non-vertebrate chordates, pointing out the simplicity of the V-shaped myomeres and sometimes disputing the presence of eye musculature, enamel, and dentine.

Cladistic analyses, carried out using characteristics of conodont soft and hard tissues (Figure 8), have placed conodonts either as a sister group to the lampreys or, most recently, as more derived than both hagfishes and lampreys, the most primitive of living vertebrates. It has also been shown that this phylogenetic position does not rely on the interpretation of conodont hard tissues as homologues of enamel and dentine; if enamel, dentine, and extrinsic eye musculature are scored as absent, cladistic analysis still places the conodonts as more derived than the hagfishes. The weight of current evidence, therefore, indicates that the best hypothesis for conodont affinity is that they are crown group vertebrates (i.e., nested within the clade delimited by extant vertebrates). Their position as more derived than lampreys also means that they are stem group gnathostomes (jawed vertebrates), albeit more primitive than the array of extinct armoured agnathans (e.g., heterostracans, thelodonts, osteostracans) known from the Cambrian to Devonian.

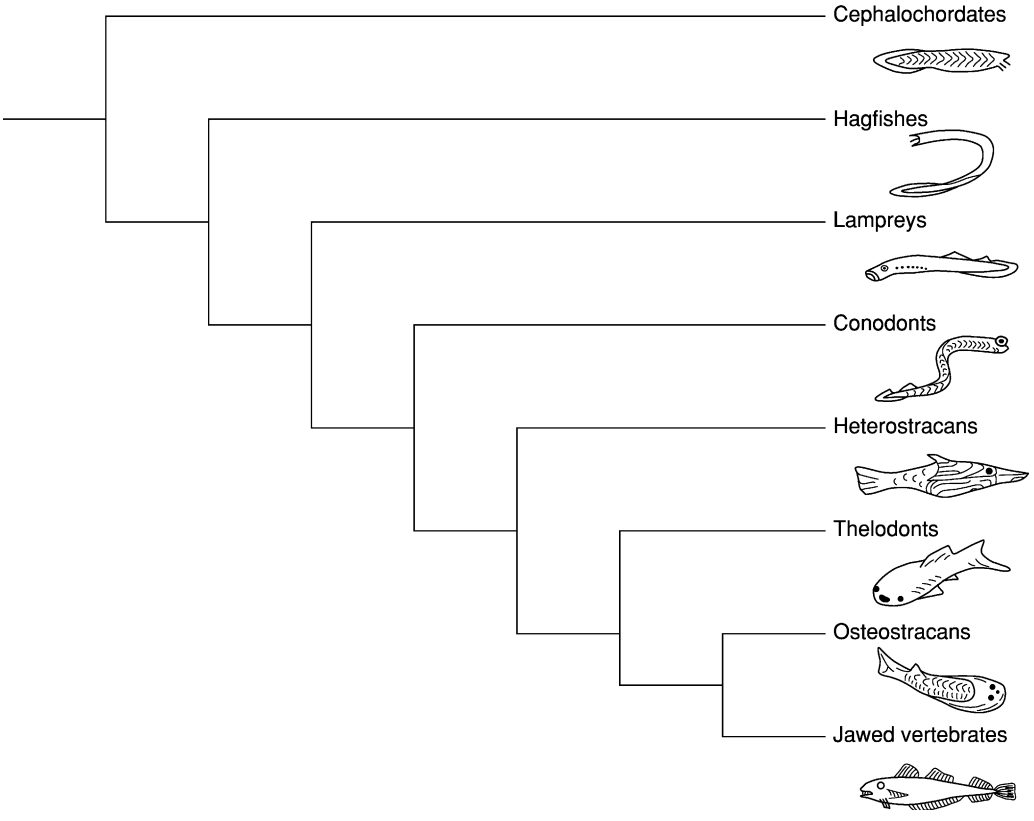


Figure 8 Cladogram of relationships between representative chordate groups, showing the most parsimonious position of the conodonts. Simplified from Donoghue PCJ, Forey PL, and Aldridge RJ (2000) Conodont affinity and chordate phylogeny. *Biological Reviews* 75: 191–251.

Classification

The classification of conodonts is in a phase of evolution. As the apparatus structures of various taxa and the homologies between elements become better understood, relationships are being increasingly studied using cladistic and other methods, but these have yet to be reflected in a firm classificatory scheme. The most widely used classification currently has conodonts assigned to seven orders, although it is realized that some of these are paraphyletic.

1. Order Proconodontida – conodonts with apparatuses mostly comprising coniform elements with deep basal cavities and smooth surfaces.
2. Order Belodellida – conodonts with apparatuses typically of coniform elements with thin walls, very deep cavities, and smooth surfaces; the elements commonly bear distinct costae, which may develop fine serration or denticulation.
3. Order Protopanderodontida – conodonts with apparatuses of longitudinally striated coniform elements that lack lateral furrows.
4. Order Panderodontida – conodonts with apparatuses of longitudinally striated elements with lateral furrows; the elements are coniform or bear denticles on the concave ('posterior') margin.
5. Order Prioniodontida – conodonts with P positions of the apparatus occupied by pastinate coniform or pastinate pectiniform elements; the more derived taxa have complex apparatuses, with strong differentiation between the morphologies of pectiniform P and ramiform M + S elements, and sometimes with the development of platform morphologies, especially in the P positions.
6. Order Prioniodinida – conodonts with apparatuses typically of ramiform elements, with those in the P positions characteristically digyrate; the denticles on all elements are commonly discrete and peg-like.
7. Order Ozarkodinida – conodonts with complex apparatuses with strong differentiation between the P and M + S elements; the P elements are pectiniform, consisting of a carminate pair and an angulate pair; platform development is common, especially on the carminate elements; the M and S elements are ramiform.

Architecture and Function of the Conodont Apparatus

Reconstruction of the architecture of conodont apparatuses relies on exceptionally preserved fossils, where evidence of the original geometrical relationship between the component elements is preserved. Such fossils include those with associated soft tissues, plus a larger number of, still very uncommon, undisturbed apparatuses preserved on bedding planes. When the conodont animal died and the soft tissues decayed, the elements were generally released to become scattered on the seabed. Scavengers and burrowers will have further dispersed them, and they were commonly transported by currents and waves before finally becoming incorporated into the sediment. In normal circumstances, therefore, it is impossible to recognize the various elements that came from an individual animal. However, in some deposits, in which decomposition of the conodont carcass took place in very quiet conditions, the skeletal apparatus has simply collapsed onto the bedding surface. These 'natural apparatuses' are normally found in shales or fine-grained limestones, and they preserve a two-dimensional representation of the original three-dimensional architecture. Skeletons that have collapsed onto bedding planes in different orientations can be used to reconstruct the original configuration, in much the same way that a building can be envisaged from drawings of the elevation and plan. Some additional information comes from clusters of elements that have become fused together diagenetically within the sediment. However, direct evidence of apparatus architecture is still only available for a small number of conodont genera. The best known apparatuses architecturally are those of a number of ozarkodinid taxa and that of the giant prioniodontid, *Promissum*, from the Soom Shale. There is also direct evidence of the three-dimensional structure of a few prioniodinid and panderodontid genera.

In the ozarkodinid apparatus (**Figure 5**), the pairs of P elements are orientated with their long axes perpendicular to that of the animal; the carminate pair is behind the scissor-like angulate pair. The rostral array of S elements is positioned with the long denticulate processes rising caudally in relation to the body axis and tilted towards the centre. The pair of M elements is above the S array, with the cusps directed downwards and inwards.

The apparatus is generally considered to be a feeding device, and functional interpretations have focused on two hypotheses. One model suggests that conodonts were microphagous suspension feeders, in which the rostral S and M elements supported a ciliated structure that captured food particles that were passed back to the tissue-covered P elements to be gently crushed and ingested. The alternative is that conodonts were macrophagous, with the S and M elements actively grasping food which was then sliced and crushed by the P elements. For this mode, the elements would have been exposed during function, and there is a need to reconcile this with the appositional accretion

of the lamellar tissue; this may be effected by periodic soft tissue cover.

The recognition of microwear features on a variety of conodont elements (Figure 9) provides strong evidence that they performed an active tooth-like function. This is supported by evidence from undisturbed ozarkodinid apparatuses in which the platform elements are constructed to occlude so precisely that there would have been no available space for an intervening cover of soft tissue. Internal discontinuities within the lamellar crown tissue can be interpreted as resulting from episodes of function alternating with episodes of growth, with worn surfaces being covered by newly secreted phosphate during the phases of dormancy.

There is less evidence for the function of conodont apparatuses consisting entirely of coniform elements. They could plausibly have been used to grasp prey, but direct evidence of surface wear is currently sparse.

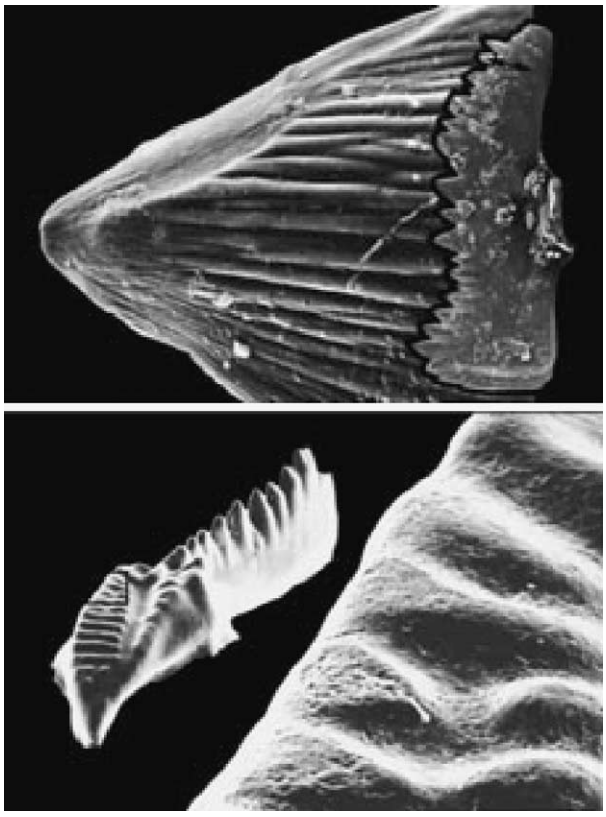


Figure 9 Microwear on the surfaces of conodont elements. Top photographs: P element of *Ozarkodina*; complete element to right, ×33; to left, close-up of highlighted single denticle showing wear facet, ×1150. Lower photographs: P element of *Idiognathodus*, ×25; close-up of pitted wear facets on oral surface, ×200. Reprinted with permission from *Nature* (Purnell MA (1995) Microwear on conodont elements and macrophagy in the first vertebrates. *Nature* 374: 798–800). © 1995 Macmillan Magazines Limited.

If conodonts were the most primitive of stem group gnathostomes, as suggested above, the functional interpretation of their skeletal apparatus assumes major significance in the assessment of the origin of the vertebrate skeleton. More primitive chordates than the conodonts all lacked a biomineralized skeleton. The evidence that conodont elements fulfilled a tooth function, albeit without jaws, therefore implies that the vertebrate skeleton first arose, in the conodonts or their ancestors, as an oro/pharyngeal raptorial feeding system.

Many conodont species are globally widespread and broadly independent of facies in their distribution, suggesting that they were free-swimming pelagic animals. Other taxa show a more restricted distribution and a closer relationship to facies belts, and these were more probably nektobenthic or even benthic in habit.

Conodont Evolution and Biostratigraphy

The ancestry of the true conodonts ('euconodonts') discussed above appears to lie within a group of taxa represented by scattered phosphatic sclerites referred to the Paraconodonta. Paraconodonts lack an enamel crown, but are formed of a tissue that has been compared with dentine. No complete apparatuses of paraconodonts have been discovered, but fused clusters of small numbers of elements have been found that indicate that some taxa, at least, had multielement apparatuses with some morphological differentiation between the component elements. Paraconodonts range from the Middle Cambrian into the Ordovician.

The earliest euconodonts appeared in the Late Cambrian, where they are predominantly represented by coniform proconodontid elements. There was a major radiation of conodonts in the Ordovician, with the appearance and diversification of protopanderodontid and prioniodontid apparatuses in the earliest Ordovician, followed by belodellid, panderodontid, prioniodinid, and, finally, ozarkodinid taxa. No new orders appeared after the Ordovician, and the diversity of genera seen during this period is not matched again in later conodont history. However, conodonts are important biostratigraphical indices from the base of the Ordovician to their disappearance at the top of the Triassic.

The most important group biostratigraphically for Ordovician strata is the prioniodontids, which were highly diverse throughout the period and evolved rapidly. However, the prioniodontids were severely depleted by the end-Ordovician extinction event and were reduced again at the end of the Llandovery Epoch (Early Silurian); the last representatives of the order disappeared at the end of the Devonian. During

the Silurian, ozarkodinids flourished and there was a major diversification of this order in the Devonian. Conodont biozonal schemes from the Silurian to the Early Triassic rely heavily on ozarkodinid taxa. Prioniodinids remained conservative for much of their history, but diversified in the Triassic, with several lineages developing platform P elements; this order outlasted all the others and prioniodinids provide the basis for Middle and Upper Triassic conodont biozonation. In general, taxa with coniform elements had declined by the end of the Silurian, with only a very few coniform genera recorded above the Devonian. Conodont diversity was low in the Permian, but conodonts were relatively unaffected by the end-Permian mass extinction; however, they declined again in the latest Triassic and succumbed at the very end of that period.

Conodont Elements as Indices of Thermal Maturation

If conodont elements are heated, organic material within the lamellar crown tissue breaks down, and carbon fixing produces a continuous colour change from pale amber through brown to black; at even higher temperatures, elements become grey, then opaque white, and then clear. This colour alteration is dependent on temperature and time, but, where temperatures have been maintained for the order of millions of years, differences in the length of time make only a small difference to the colour achieved. A colour alteration index (CAI) has been established for such long-term heating, calibrated to temperature ranges, with CAI 1–5 covering the range from pale amber to black (<50–350°C) and CAI 6–8 covering metamorphic temperatures. This index has been widely applied in investigations of the thermal history of sedimentary basins containing strata of Cambrian to Triassic age, and provides information on thermal cut-offs in oil and gas exploration.

See Also

Biozones. **Evolution**. **Mesozoic:** Triassic. **Palaeozoic:** Ordovician. **Petroleum Geology:** Exploration. **Stratigraphical Principles**.

Further Reading

Aldridge RJ and Purnell MA (1996) The conodont controversies. *Trends in Ecology and Evolution* 11: 463–468.

Donoghue PCJ, Forey PL, and Aldridge RJ (2000) Conodont affinity and chordate phylogeny. *Biological Reviews* 75: 191–251.

Epstein AG, Epstein JB, and Harris LD (1977) Conodont color alteration – an index to organic metamorphism. *United States Geological Survey, Professional Paper* 995: 1–27.

Purnell MA (1995) Microwear on conodont elements and macrophagy in the first vertebrates. *Nature* 374: 798–800.

Sweet WC (1988) *The Conodonta: Morphology, Taxonomy, Paleoecology and Evolutionary History of a Long-Extinct Animal Phylum. Oxford Monographs on Geology and Geophysics 10*. New York and Oxford: Clarendon Press.

Foraminifera

M A Kaminski, University College London, London, UK

Introduction

The term 'Foraminifera' is derived from the Latin *foramen*, little hole and *ferre*, to carry or bear. They are a phylum of predominantly marine heterotrophic testate protozoans with tubular mitochondrial cristae and granuloreticular pseudopodia emanating from one or more openings in their tests. With over 50 000 known fossil and living species, the foraminifera constitute a diverse and geologically long-ranging group of organisms found in virtually all marine habitats. Morphologically, they form a heterogeneous and perhaps polyphyletic group with a fossil record that begins in the latest Precambrian, though forms with organic or unmineralised tests probably existed earlier. The group as a whole is characterised by the presence of an organic, agglutinated, or secreted biomineralised test partially enclosing the amoeboid body. The test may be single-chambered, pseudocolonial, pseudochambered, or multichambered with interconnected chambers added as the cell grows. Openings between the chambers [foramens] allow the cytoplasm to flow freely, while one or more apertures enable the cell to communicate with its external environment. A pseudopodial network may arise from a single apertural opening forming a distinct pseudopodial trunk or radiate in all directions from numerous openings or pores in the walls of some calcareous taxa. Pseudopodia exhibit bidirectional streaming, and are used for locomotion in free-living taxa, to anchor attached forms to the

substrate, to capture and ingest food items, or to build new chambers and growth or reproductive cysts. Reproduction is remarkably complex, and typically involves an alternation of generations between a multinucleate diploid [agamontic] asexually reproducing stage and a mononucleate haploid [gamontic] sexually reproducing stage. The latter gamontic individuals may also reproduce asexually by budding or fission, and sexually by the release of flagellated gametes. The two generations may produce morphologically distinct tests, with the gamont possessing a relatively large [megalospheric] first chamber or 'proloculum', and the agamont possessing a much smaller [microspheric] proloculum, but often a more complex and larger adult test. This dimorphism is most evident in the more advanced groups. Foraminifera are heterotrophic, and food items include other protozoans, algae, ciliates, small crustaceans, or other foraminifers. Some planktonic and larger benthic species inhabiting the photic zone harbour endosymbionts; some benthic forms burrow, ingesting sediment to crop bacteria, while others can make use of dissolved organic matter or phytodetritus.

The Rank of the Foraminifera

The discovery by Dujardin (1835) that Foraminifera are protozoans rather than cephalopods led d'Orbigny in 1852 to classify the group as a class with six orders based on chamber arrangement, and a seventh for single-chambered forms. Subsequent to d'Orbigny's original classification, later workers variously regarded the group to be of lower taxonomic rank. The widely accepted classification of Loeblich and Tappan regarded the group as a class with 12 major groups treated as orders. The class rank is maintained in the second edition of the *Illustrated Guide to the Protozoa*. However, over the last three decades, Protozoologists in both Russia and North America have assigned the group to a higher rank. Margulis first elevated the Foraminifera to the rank of a phylum, a rank that is maintained in her popular textbook *Five Kingdoms*. In his expanded classification of the Kingdom Protozoa, Cavalier-Smith first regarded the Foraminifera as a subphylum of the phylum Reticulosa (= Granuloreticulosa of earlier authors), but in his latest revision Cavalier-Smith (1998) quoted cytological evidence that removes the naked athalamids from that phylum. As a result, Cavalier-Smith removes the Granuloreticulosa/Reticulosa from his classification and elevates the foraminifera to the status of a phylum. In Russia, foraminiferal workers were quick to embrace the idea of a higher rank for the Foraminifera, with Mikhalevich regarding the group as a subphylum, later elevated to the status of phylum. This is the rank used in the monumental volume *Protista: Handbook on Zoology* published by the Russian Academy of Sciences (Alimov, 2000), which adopts the classification of Mikhalevich. The *International Working Group on Foraminiferal Systematics* [IWGFS] of the Grzybowski Foundation now accepts the phylum rank adopted by the protozoological community.

Classification of the Foraminifera

The two most widely accepted classifications of the Foraminifera are currently based on the morphology of the foraminiferal test. Important features include the composition and layering of the wall, the presence or absence of pores, the overall test shape, the mode of coiling, the shape of chambers and the presence of any internal structures such as pillars, the number and position of the apertures or openings, and any apertural modifications such as lips or teeth. Two schools of thought have recently emerged, both of which utilise morphological criteria with different weightings.

Since the mid-nineteenth century, wall ultrastructure has been regarded as a prime criterion for classification at a higher level, when Carpenter first subdivided the Foraminifera into two suborders (Perforata and Imperforata) based on the presence or absence of perforations in the test wall. In his classification, Carpenter also took into account the composition of the wall and remarked "The imperforate sub-order may be divided into three very natural groups, according as the nature of the envelope is membranous, porcellanous, or arenaceous; and thus we have the families Gromida, Miliolida, and Lituolida". In 1876, TR Jones raised the status of the 'arenaceous' and 'porcellanous' forms to that of groups with equal rank to the perforate forms. Jones' idea of grouping the agglutinated, imperforate porcellanous, and perforate 'hyaline' forms into separate higher-order categories was used in many later classifications. In their first comprehensive classification in 1964, Loeblich and Tappan used wall composition and microstructure as the defining character for the hierarchy of the foraminiferal groups, recognizing five suborders: (i) the organic-walled Allogromiina; (ii) the agglutinated-walled Textulariina; (iii) the calcareous microgranular Fusulinina; (iv) the porcelaneous Miliolina; and (v) the calcareous perforate Rotaliina. The most recent update by Loeblich and Tappan continued to use wall composition (including features such as mineralogy, layering, or nature of the cement in the case of agglutinated forms), to define 14 foraminiferal orders, which is here extended to 16 groups based on new research. If the class rank for the foraminifera adopted by North American workers is

Figure 1 Summary of compositionally-based groups of the Foraminifera, (modified after Loeblich and Tappan (1992), Sen Gupta (1999) and Kaminski (2004).) The 'orders' of Loeblich and Tappan (1992) are here regarded as informal groupings pending revision of their status. The agglutinated Trochamminida and the Carterinida were included within the Lituolida by Kaminski (2004).

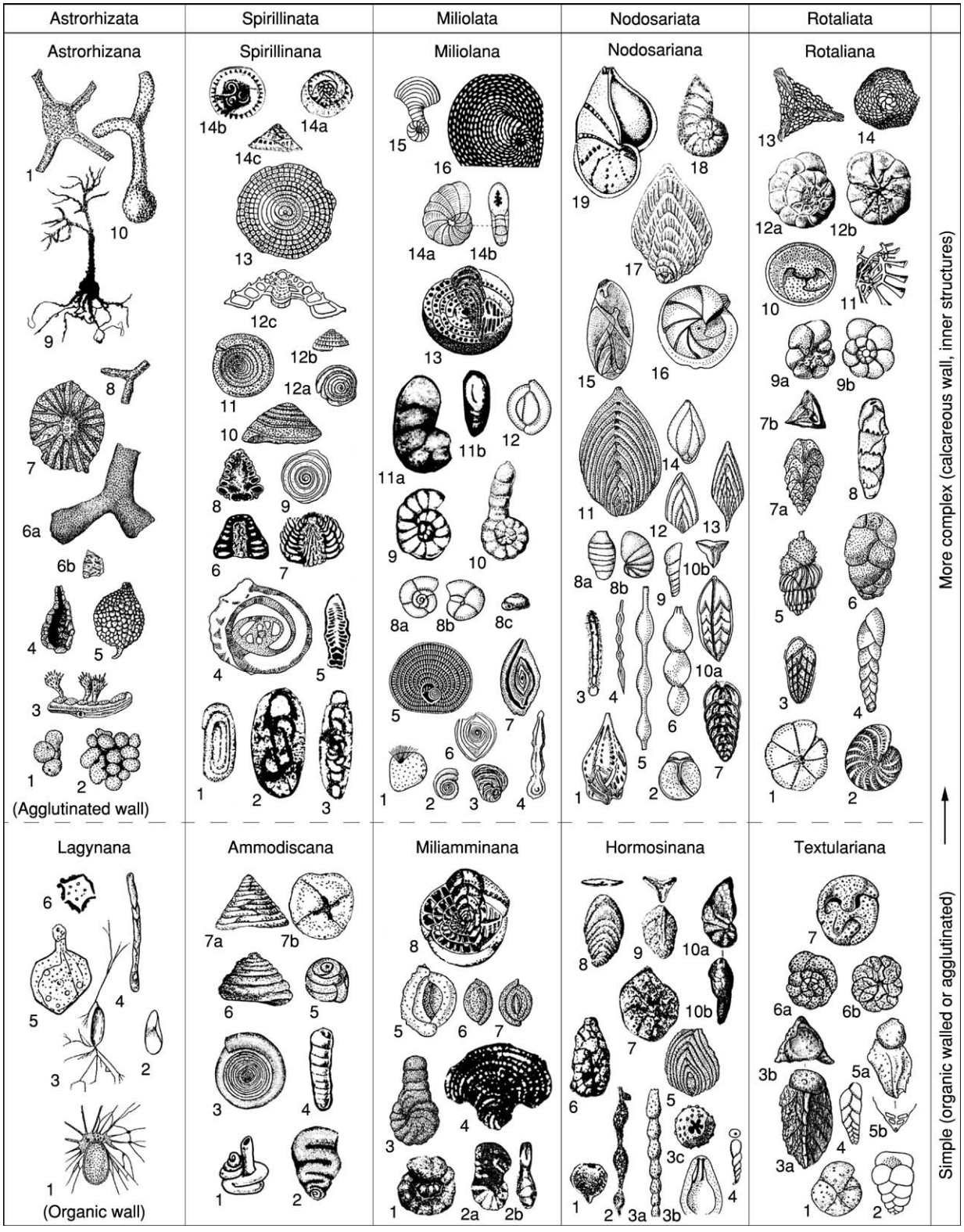


Figure 2 The main phyletic lines of benthic foraminifera (classes); with primitive and advanced subclasses and representative genera (modified after Mikhalevich (2004)).

Class Astrorhizata. Subclass Lagynana: 1. *Diplogromi*a, 2. *Lagynis*, 3. *Phainogulmia*, 4. *Xenothekella*, 5. *Heterogromia*, 6. *Microcometes*. Subclass Astrorhizana: 1. *Sorostomasphaera*, 2. *Thuramminopsis*, 3. *Halyphysema*, 4. *Lagenammina*, 5. *Astrammina*, 6a,b. *Schizammina*, 7. *Masonella*, 8. *Rhabdammina*, 9. *Notodendrodes*, 10. *Astrorhiza*, 11. *Saccorhiza*.

Continued

Figure 3 Fossil foraminifera, showing diversity of form of the chambered test (or shell). Left to right, they comprise the genera Ammotium, Bolivina, Planulina, Cyclammina, and Pseudorotalia, respectively. The first and fourth are agglutinating foraminifera, making their test from mineral grains 'glued' together with an organic cement. The other three are calcareous, secreted by the foraminifer itself. Magnifications between ×25 and ×150.

retained and the foraminiferal wall structure and composition is used at the highest taxonomic level, the eight main systematic groupings within the Foraminifera may be defined at the rank of a subclass. However, in view of the elevation of the Foraminifera to the status of a phylum, the rank of these groups will probably change in the near future. Pending a decision on their rank by the IWGFS, the main wall-structure groups are treated here as informal categories. The main groups are the allogromiids, agglutinates, fusulinids, miliolids, silicoloculinids, robertinids, and other minor perforate aragonitic groups, the spirillinids which display a unique crystal stucture, and the perforate calcitic groups consisting of rotaliids, buliminids, and globigerinids (Figure 1). This subdivision, based on wall structure, is not strictly a phylogenetic scheme, and even separates some groups based on habitat (i.e., planktonic and benthic groups) in spite of their identical wall structure.

Alongside the work of Loeblich and Tappan, a conceptually different version of the higher systematics of the foraminifera ('the Russian school') was pioneered by Fursenko and Rauser-Chernousova, and developed in detail by Mikhalevich and others. According to these authors, the starting point for the systematic hierarchy is the overall test morphology rather than wall structure. The general 'bauplan' of the foraminiferal test is regarded to be associated with the function and habitat of the whole organism. The mode of coiling and apertural characteristics of the test are considered to be the most conservative feature of the test, and the overall test shape is given priority in those cases where the test composition and/or microstructure has undergone modification within evolutionary lineages. Therefore, the composition and ultrastructure of the foraminiferal wall has subordinate significance compared with test morphology. Mikhalevich uses the following criteria to define the main foraminiferal classes: 1. the number of chambers, 2. the chamber shape, 3. pathways of chamber formation resulting in the nature of chamber conjunction, 4. the shape of the test and the predominant mode of coiling, 5. the shape and position of the main aperture, 6. the development of any inner apertural structures, 7. the development of integrative apertural systems, 8. the presence or absence of additional apertures, 9. the presence of additional skeletal plates, and 10. the presence or absence of canal systems. According to the Mikhalevich classification, the foraminifera are composed of a fewer number of larger groups of equal rank, and consist of the classes Astrorhizata, Spirillinata, Miliolata, Nodosariata, and Rotaliata (the latter also includes the planktonic Globigerinana as a subclass). The more advanced multichambered classes have both secreted calcareous and agglutinated subclasses (Figures 2 and 3). In each case, the agglutinated subclass (or in the case of the Astrorhizata, the organic-walled Lagynana) is regarded to be the ancestral group. Within each subclass, there are numerous examples of strikingly isomorphic pairs of

Class Spirillinata. Subclass Ammodiscana: 1. *Ammovertellina*, 2. *Ammovertella*, 3. *Ammodiscus*, 4. *Turritelella*, 5. *Repmanina*, 6. *Arenoturrispirillina,* 7a,b. Tetrataxis.
Subclass Spirillinana: 1. *Miliospirella*, 2. *Glomodiscus*, 3. *Archediscus dubitabilis*, 4. *A. karreri*, 5. *Cylindrotrocholina*, 6. *Howchinia*, 7. *Lasiotrochus*, 8. *Babelispirrillina*, 9. *Coronipora*, 10. *Trocholina*, 11. *Spirillina*, 12. *Spirotrocholina* (a,b. views of the test, c. view of the canal in axial section), 13. *Annulopatellina*, 14. *Paleopatellina.*

Class Miliolata. Subclass Miliamminana: 1. *Recurvoides*, 2a,b. *Charentia*, 3. *Lituola*, 4. *Alzonella*, 5. *Dentostomina*, 6. *Sigmoilopsis*, 7. *Ammomassilina,* 8. *Reticulinella.* Subclass Miliolana: 1. *Squamulina*, 2. *Cornuspira*, 3. *Cornuspiroides*, 4. *Gheorgianina*, 5. *Discospirina*, 6. *Cornuloculina*, 7. *Spirophtalmidium*, 8. *Fisherinella*, 9. *Zoella*, 10. *Spirolina*, 11a,b. *Danubiella*, 12. *Quinqueloculina*, 13. *Neoalveolina*, 14a,b. *Dendritina*, 15. *Laevipeneroplis*, 16. *Parasorites.*

Class Nodosariata. Subclass Hormosinana: 1. *Saccammina*, 2. *Hormosinella*, 3a,b,c. *Nodosinum,* 4. *Adelungia*, 5. *Pseudopalmula*, 6. *Nouria*, 7. *Agardhella*, 8. *Flabellammina*, 9. *Triplasia,* 10a,b. *Ammomarginulina.* Subclass Nodosariana: 1. *Lagena*, 2. *Parafissurina*, 3. *Syzrania*, 4,5. *Grigelis*, 6. *Nodosaria*, 7. *Multiseptida,* 8a,b. *Lingulina* (a. microspheric, b. megalospheric forms), 9. *Marginulina*, 10a,b. *Tristix*, 11. *Kyphopixa*, 12. *Dyofrondicularia*, 13. *Flabellina*, 14. *Polymorphina*, 15. *Laryngosigma*, 16. *Lenticulina,* 17. *Planularia*, 18. *Hemicristellaria*, 19. *Saracenaria.*

Class Rotaliata (the third subclass, Globigerinana, is not shown). Subclass Textulariana: 1. *Haplophragmoides*, 2. *Minouxia*, 3a,b. *Gaudryina*, 4. *Pseudobolivina*, 5a,b. *Clavulina*, 6a,b. *Asterotrochammina,* 7. *Tiphotrocha.* Subclass Rotaliana: 1. *Bermudezinella*, 2. *Elphidium*, 3. *Brizalina*, 4. *Bolivinellina*, 5. *Euuvigerina*, 6. *Sporobulimina*, 7a,b. *Reussella*, 8. *Pseudobuliminella*, 9a,b. *Discorbis*, 10. *Neoconorbina*, 11. *Rotalia* (detail of canal system), 12a,b. *Ammonia*, 13. *Baculogypsinoides* (horizontal section of megalospheric individual showing canals), 14. *Eulinderina.*

agglutinated and calcareous genera. The Mikhalevich classification also makes fuller use of the Linnean hierarchy of classes, subclasses, orders, suborders, superfamilies, etc.

Further Reading

Cavalier-Smith T (1998) A revised six-kingdom system of life. *Biological Reviews* 73: 203–266.

Kaminski MA (2004) The year 2000 classification of the agglutinated foraminifera. In: Bubik M and Kaminski MA (eds.) *Proceedings of the Sixth International Workshop on Agglutinated Foraminifera. Grzybowski Foundation Special Publication.* 8: 237–255.

Lee JJ (1990) Phylum Granuloreticulosa (Foraminifera). In: Margulis L, Corliss JO, Melkonian M, and Chapman DJ (eds.) (1990) *Handbook of Protoctista.* Boston: Jones and Bartlett.

Lee JJ, Pawlowski J, Debenay JP, *et al.* (2000) Class Foraminifera. In: Lee JJ, Leedale GF, and Bradbury P (eds.) *An Illustrated Guide to the Protozoa,* second edition, pp. 877–951. Society of Protozoologists. Lawrence Kansas: Allen Press.

Loeblich AR and Tappan H (1964) Part C. Protista 2. Chiefly 'Thecamoebians' and Foraminiferida. In: Moore RC (ed.) *Treatise on Invertebrate Paleontology,* p. 900. Lawrence Kansas: The Geological Society of America and the University of Kansas.

Loeblich AR and Tappan H (1987) *Foraminiferal Genera and their Classification.* New York: Van Nostrand Reinhold.

Loeblich AR and Tappan H (1992) Present status of Foraminiferal Classification. In: Takayanagi Y and Saito T (eds.) *Studies in Benthic Foraminifera.* Tokyo: Tokai University Press.

Loeblich AR and Tappan H (1994) Foraminifera of the Sahul Shelf and Timor Sea. *Cushman Foundation for Foraminiferal Research Special Publication* 31: 661.

Margulis L (1974) Five-kingdom classification and the origin of evolution in cells. *Evolutionary Biology* 7: 45–78.

Margulis L and Schwartz K (1988) *Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth,* 2nd edn. New York: W.H. Freeman and Co.

Mikhalevich VI (1980) Sistematika i evolyutsiya foraminifer v svete novykh dannykh po ikh tsitologii i ul'trastrukture. *Trudy Zoologicheskogo Instituta Akademii Nauk SSSR* 94: 42–61.

Mikhalevich VI (1998) Makrosystema Foraminifer. *Izvestiya Akademii Nauk, Seriya Biologicheskaya* 1998(2): 266–271.

Mikhalevich VI (2000) Typ Foraminifera d'Orbigny, 1826. In: Alimov AF (ed.) *Protisty: Rukovodstvo po Zoologii,* pt. 1, pp. 533–623. St. Petersburg: Nauka Publishers.

Mikhalevich VI (2004) On the heterogeneity of the former Textulariina (Foraminifera). In: Bubik M and Kaminski MA (eds.) *Proceedings of the Sixth International Workshop on Agglutinated Foraminifera,* 8, pp. 317–349. Grzybowski Foundation Special Publication.

Sen Gupta BK (1999) Systematics of modern Foraminifera. In: Sen Gupta BK (ed.) *Modern Foraminifera,* pp. 7–36. Dordrecht: Kluwer Academic Publishers.

Ostracoda

D J Horne, University of London, London, UK

Introduction

Ostracods, small crustacean arthropods (*see* **Fossil Invertebrates:** Arthropods) characterized by a bivalved carapace that can totally enclose the body and appendages, have an excellent fossil record by virtue of their small size and calcite shells (valves). Their bodies have reduced trunk segmentation and up to eight pairs of specialised limbs that are protruded from the gaping valves for locomotion, feeding, and reproductive activity. Adult ostracods are typically 0.5–2.0 mm long; some interstitial forms, however, are as small as 0.2 mm, some freshwater species grow up to 8.0 mm, and the pelagic marine myodocopan *Gigantocypris* reaches 32 mm. They are one of the most diverse crustacean groups (there are estimated to be more than 20 000 living species, of which only about 8000 have been described) and have a 500 million-year fossil record, from the Ordovician onwards, with more than 65 000 fossil species described. They are all essentially aquatic, inhabiting both marine and non-marine environments, although some taxa are adapted to a semi-terrestrial life. They have many applications in palaeoenvironmental analysis, palaeoclimatology and biostratigraphy.

Classification

The Class Ostracoda is divided into two subclasses, the Myodocopa and the Podocopa, each comprising three orders (Table 1). Of the Myodocopa only the Cladocopina and Thaumatocypridoidea (both Halocyprida) have strongly calcified valves and good fossil records, while those of most Myodocopida and Halocypridoidea (Halocyprida) are weakly calcified and hence less common as fossils. The three podocopan

Table 1 A summary classification of the Ostracoda

Class ***Ostracoda***
Subclass **Podocopa**
Order **Palaeocopida**
Suborder **Beyrichicopina**
Including superfamilies Beyrichioidea, Tetradelloidea, Eurychilinoidea, Aparchitoidea, Primitiopsoidea
Suborder **Binodicopina**
Including superfamilies Bollioidea, Aechminoidea, Drepanelloidea, Nodelloidea, Limbatuloidea
Suborder **Kirkbyocopina**
Including superfamilies Kirkbyoidea, Puncioidea
? Suborder **Eridostracina** (may not be ostracods)
Order **Platycopida**
Suborder **Platycopina**
Including superfamilies Leperditelloidea, Kloedenelloidea, Cytherelloidea
Suborder **Metacopina**
Including superfamilies Healdioidea, Thlipsuroidea
Order **Podocopida**
Suborder **Bairdiocopina**
Superfamily Bairdioidea
Suborder **Sigilliocopina**
Including superfamilies Sigillioidea, Bairdiocypridoidea
Suborder **Cypridocopina**
Including superfamilies Macrocypridoidea, Pontocypridoidea, Cypridoidea
Suborder **Darwinulocopina**
Including superfamilies Carbonitoidea, Darwinuloidea
Suborder **Cytherocopina**
Including superfamilies Quasillitoidea, Cytheroidea, Terrestricytheroidea
Subclass **Myodocopa**
Order **Myodocopida**
Including superfamilies Cypridinoidea, Cylindroleberidoidea, Sarsielloidea, Cyprelloidea, Bolbozooidea
Order **Halocyprida**
Suborder **Entomozocopina**
Superfamily Entomozooidea
Suborder **Halocypridina**
Including superfamilies Thaumatocypridoidea, Halocypridoidea
Suborder Cladocopina
Superfamily Cladocopoidea
? Order Leperditicopida (may not be ostracods)

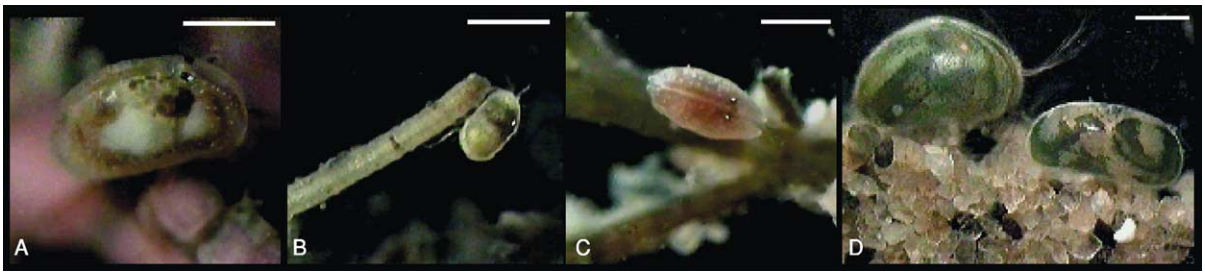


Figure 1 Some living ostracods. A–C: marine-brackish cytheroidean podocopids; D: freshwater cypridoidean podocopids. Scale bars approx. 1 mm.

orders are the Platycopida, the ubiquitous Podocopida (the most diverse group of Ostracoda at the present day; Figure 1) and the Palaeocopida (diverse and widespread in the Palaeozoic, but now extremely rare). The Cambrian Bradoriida and Phosphatocopida, once included within the Ostracoda, are now considered to be separate bivalved arthropod groups. The ostracod affinities of some other groups, such as the Eridostracina and Leperditicopida, are currently being questioned.

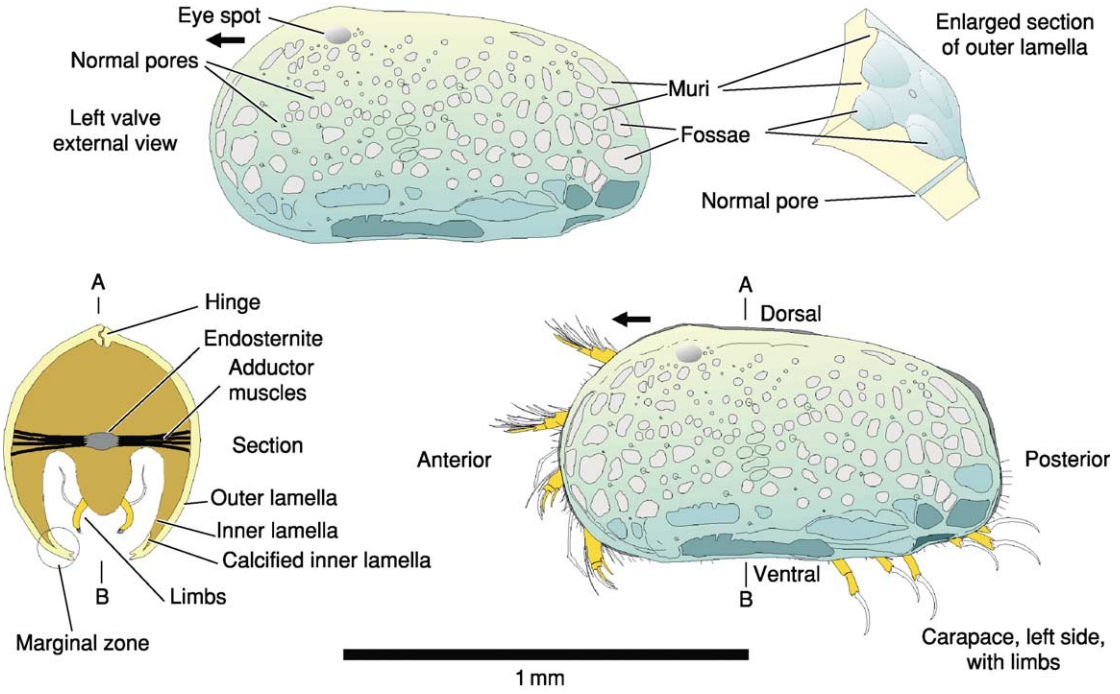


Figure 2 Morphology of a typical ostracod: external carapace morphology and cross-section of a cytheroidean podocopid. Arrows indicate anterior direction in this and subsequent figures.

Morphology and Biology

The ostracod carapace consists of two valves formed by lateral folds (duplicatures) of the epidermis, completely enclosing the body and limbs (Figures 2 and 3). In podocopans the valves, comprising outer and inner lamellae, are mineralized with low-magnesium calcite; in fossil material, only the calcified inner lamella indicates the duplicature, since the uncalcified part is unlikely to be preserved. Many myodocopans are weakly or not at all calcified, and their ultrastructure differs from that of podocopans; furthermore, fossil valves of myodocopans are often secondarily calcified and calcareous nodules can form as postmortem artefacts in the valves of specimens caught alive; the probable equivalent of the podocopan calcified inner lamella is termed the infold. In the Podocopa one valve is usually larger and overlaps the smaller valve along part or all of its margin, while in the Myodocopa the valves are usually more symmetrical but may show posterior overlap. A few taxa show a narrow gape even when the valves are fully closed.

Although typically 'bean-shaped' or 'mussel-shaped', ostracod carapaces are extremely varied in shape and ornament (Figures 4–6). Sexual dimorphism is common; males may be larger or smaller than females and sometimes inflated posteriorly to accommodate the relatively large copulatory appendages, while in podocopid taxa with brood care the female is larger and more inflated. The valves are closed by means of adductor muscles attached to the inner surface of the calcified outer lamellae, where distinctive scars are formed that are a useful taxonomic character, especially at superfamily level; other muscle scars on the inner surfaces of valves are associated with various appendages. Dorsally the flexible cuticle connecting the two valves has been referred to as a ligament, but it probably plays no part in the opening of the valves, which is achieved by relaxation of the adductor muscles, hydrostatic pressure, and/or appendage movements. The valves of some taxa (notably cytheroidean podocopids) have a dorsal hinge structure of interlocking grooves and bars or teeth and sockets, also a useful taxonomic character. The outer lamella is pierced by normal pore canals which may terminate as simple pores through which sensilla protrude externally, sieve pores (also bearing sensilla) or exocrine pores (without sensilla; associated with moulting). The marginal zone, formed by the coincidence of the calcified inner lamella with the outer lamella, contains many taxonomically useful characters; the former may be fused throughout its width to the inner surface of the outer lamella, or the two lamellae may separate and diverge inwards, forming a space known as a vestibulum. Sensilla are often densely spaced along the free margins of the valves,

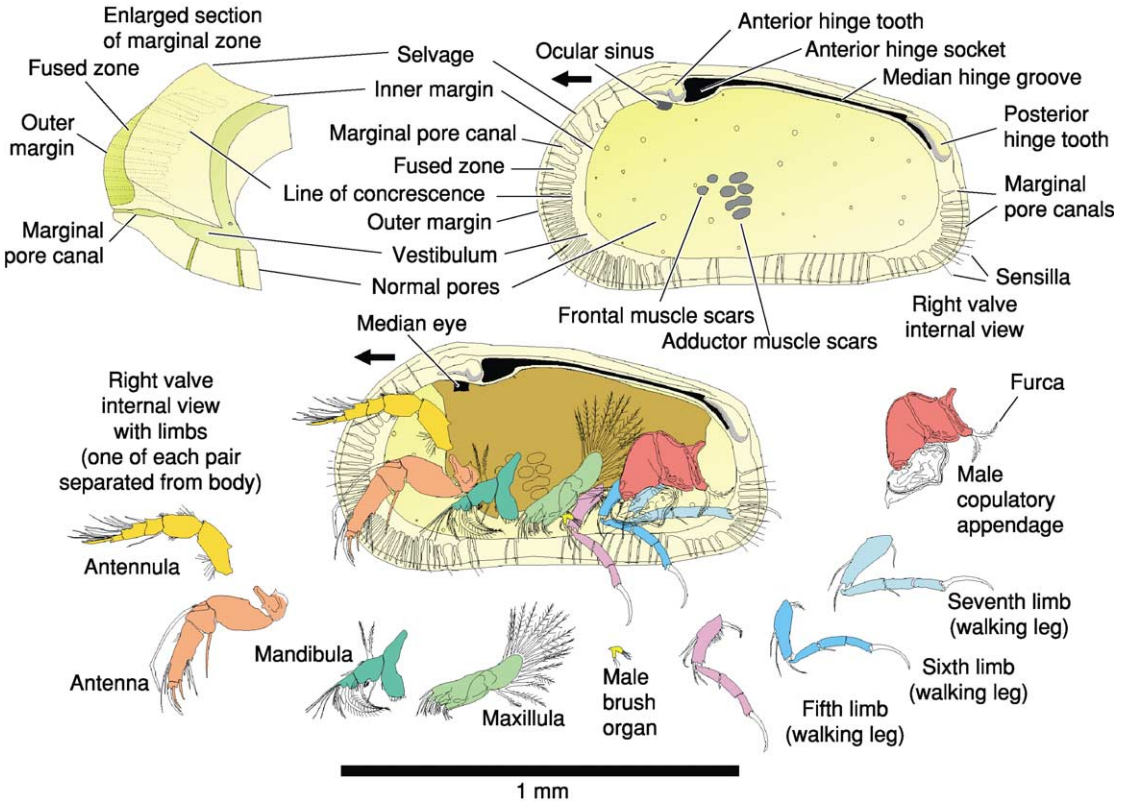


Figure 3 Morphology of a ostracod: internal valve morphology and limbs of a cytheroidean podocopid.

where the nerve-bearing canals leading to the pores pass through the fused zone, forming marginal pore canals. The calcified inner lamella may bear a ridge called a selvage, running sub-parallel to the outer margin; this forms an effective seal when the valves are closed, and there may be other, less prominent ridges called lists.

Many podocopan and myodocopan ostracods have a median or naupliar eye consisting of three optic cups which may be fused or separated, the left and right cups being inserted into ocular sinuses in the corresponding valves. Externally the calcite valves may show a clear eye-spot, a lens, or even a prominent eye tubercle. Most Myodocopida additionally have paired compound lateral eyes. The Halocyprida and some podocopan taxa are blind.

The paired limbs of adult ostracods are, from front to back, the antennules, antennae, mandibles, maxillulae, fifth limbs, sixth limbs, and seventh limbs; in some taxa (e.g., Cladocopina) posterior limbs may be absent. The eighth limbs are only present in the rare Puncioidea (the only living palaeocopids) but the copulatory appendages of other groups may have originated partly or wholly as modifications of these limbs. Additionally, a furca (a pair of caudal rami) is situated near the posterior termination of the body; in males of some taxa, there is also a pair of brush organs which may be vestigial limbs. The limbs are highly adapted to a variety of special functions, including locomotion, feeding and reproduction, and many of them bear sensory setae. The four or five head limbs are attached to the cephalon, comprising forehead, upper and lower lips, and hypostome. Most ostracods show little trace of post-cephalic segmentation other than the presence of paired appendages, but the evidence in some taxa suggests 10 or 11 trunk (thorax plus abdomen) segments (plus a posterior telson) in the Podocopa and 4 to 7 in the Myodocopa. Ostracod limbs are only fossilised under exceptional circumstances, but such occurrences are extremely valuable for resolving problems of classification and phylogeny.

Like all arthropods, ostracods grow by moulting. Podocopid ontogeny usually consists of 9 instars (moult stages): 8 juvenile and 1 adult, designated A, A-1, A-2, etc., in descending order of size. The first juvenile instar is already enclosed by two (non- or very weakly calcified) valves and has 3 pairs of limbs: antennules, antennae, and rudimentary mandibles. The genus *Manawa* (Palaeocopida: Puncioidea) is unusual in that the earliest instars have a single shield-like carapace, and only the later instars are bivalved.

Limbs are added progressively through ontogeny, accompanied by changes in carapace shape. In many taxa, the features of the marginal zone and hinge are only fully developed in adults, while in juvenile valves they are weakly developed or absent. The ornament of juveniles is usually a subdued version of that of corresponding adults, but some taxa show a marked change in ornament at the final moult. Myodocopan ontogeny consists of 4 to 7 juvenile instars and a single adult instar, the first juveniles having 5 or 6 limbs (or limb rudiments).

Most marine ostracods (myodocopans and podocopans) reproduce sexually (a few may be parthenogenetic). Males of some myodocopids produce bioluminescent courtship displays. Three reproductive modes are recognized in non-marine ostracods: fully sexual, exclusively parthenogenetic (e.g., darwinuloideans, termed 'ancient asexuals' since they are believed to have reproduced without sex for 200 million years) and mixed, for example cypridoideans with widespread parthenogenetic populations but geographically restricted sexual populations. Males of some Cypridoidea have the longest spermatozoa, relative to body size, of any animal (up to ten times the length of the adult carapace), which are kept coiled inside the duplicature and during copulation are passed through a muscular pump called a Zenker's Organ.

Most Podocopida deposit their eggs singly or in clusters, but brood care of the eggs and early instars in the posterior brood space of the adult females is known in several cytheroidean families and all darwinuloideans. Female platycopids brood eggs (usually only 4–8) but not juvenile instars. All myodocopids have brood care, but it is rare in halocyprids, most of which release eggs directly into the sea. In the Palaeozoic, palaeocopid beyrichicopine females had distinctive anteroventral brood pouches known as cruminae (Figure 9), in which fossilized juveniles have sometimes been found.

In Podocopa, life cycles vary from a few months to as long as 4 years. Shallow marine cytheroideans typically have a single generation per year, development taking place mainly in spring and summer, with delayed development of eggs or instars during the winter months; some species, however, manage 4 or 5 generations during the warmer part of the year. Many non-marine cypridoideans, especially those living in temporary ponds, have short life cycles of only a few weeks, but their desiccation-resistant eggs can remain viable for years or even decades. Total life span in Myodocopa can be as little as 1 month, or up to 4 years.

Ecology

The great majority of marine ostracods are benthonic or nektobenthonic. The only pelagic ostracods belong to the almost exclusively marine Myodocopa (Figure 4); they include active predators (e.g., on copepods) and scavengers. Pelagic Cypridinoidea

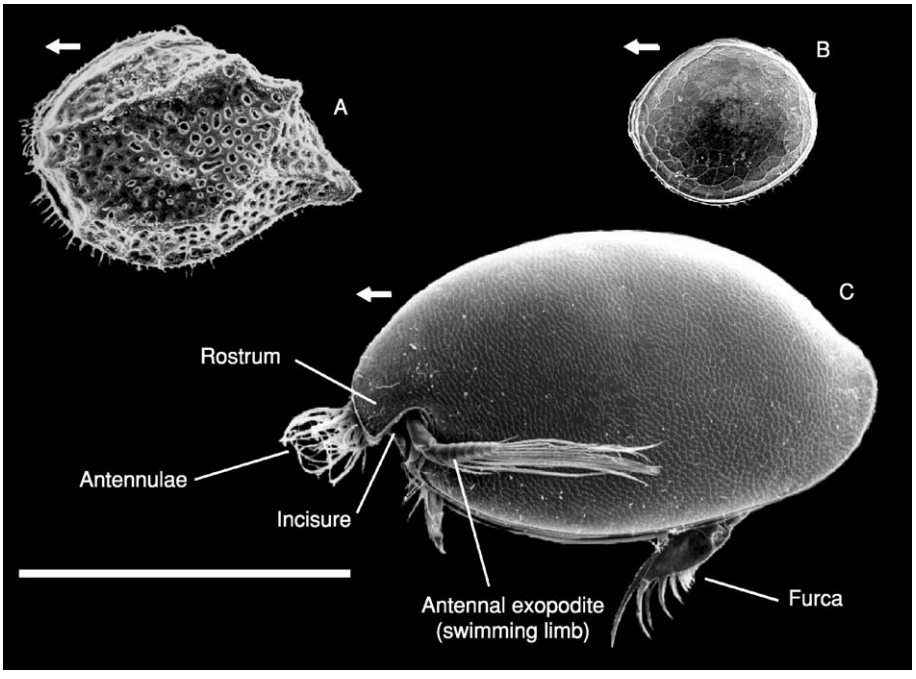


Figure 4 Shells of some Recent marine myodocopan ostracods. A: Myodocopida, Sarsielloidea; B: Halocyprida, Cladocopina; C: Myodocopida, Cypridinoidea (with limbs). Scale bar = 1 mm.

(Myodocopida) live mainly in mesopelagic and abyssal zones but are also found in shallower waters. Of the Halocyprida, all of the Halocypridoidea and a few Thaumatocypridoidea, also predators and scavengers, are truly pelagic. Of the non-pelagic Myodocopa, the nektobenthonic halocyprid Thaumatocypridoidea inhabit not only abyssal and bathyal depths but also shallow marine caves which, like deep water environments, are dark and oligotrophic, with little temperature variation. Another halocyprid group, the Cladocopina, are benthonic/nektobenthonic; they live on fine sediment bottoms in deep water, but at least some shallow water species live interstitially in coarse sediments. The Cylindroleberidoidea and Sarsielloidea (both Myodocopida) are essentially benthonic/nektobenthonic, sometimes burrowing in fine sediments; they have a wide bathymetric range, being more prevalent in shallow coastal (even intertidal) waters. Cylindroleberidoids are filter-feeders, while many sarsielloids are voracious predators. Only a few myodocopans tolerate reduced salinities and are found in estuaries (e.g., some Sarsielloidea).

Of the Podocopa, the Platycopida (Figure 5A) are all benthonic filter-feeders found predominantly in marine environments (although a few are known from brackish waters), being most diverse in warm, shallow, carbonate environments, but also living in deep waters. The Palaeocopida, highly diverse in the Palaeozoic, are today extremely rare (the Puncioidea), living interstitially in shallow marine, high-energy biogenic sands off New Zealand.

The predominantly benthonic Podocopida are the most diverse marine ostracods (Figure 5B–M), comprising representatives of six extant superfamilies. The Bairdioidea are most abundant and diverse in warm, shallow carbonate environments; benthonic detritus-feeders, unable to swim, they are mostly epifaunal crawlers but some live interstitially. Benthonic Macrocypridoidea, benthonic/nektobenthonic Pontocypridoidea and a few benthonic/nektobenthonic Cypridoidea (the candonid Paracypridinae) are found in marine and brackish waters. These three cypridocopine groups include herbivores, detritivores and carnivores; some pontocypridoids are commensal on other invertebrates such as starfish. Cytheroidea are benthonic crawlers, climbers on algae, burrowers in fine sediment, or live interstitially in coarse sand, and some are commensal, such as the Entocytheridae on freshwater and marine crustaceans. Cytheroideans include herbivores, detritivores, and carnivores. Extant Sigillioidea are rare, living in warm, shallow marine environments (probably interstitially), submarine caves, and (in at least one case) the deep sea. At least one species of the predominantly non-marine Darwinuloidea lives in brackish as well as freshwater conditions.

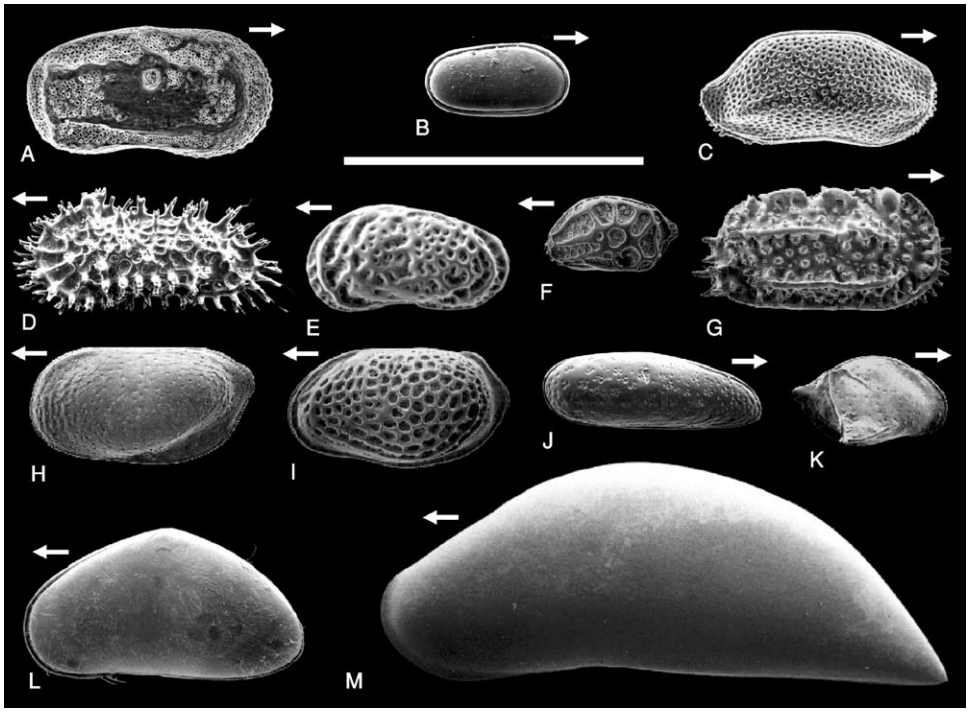


Figure 5 Shells of some Quaternary – Recent marine podocopan ostracods: A: Platycopida; B–M: Podocopida (B: Sigillioidea; C: Bairdioidea; D–K: Cytheroidea; L: Pontocypridoidea; M: Macrocypridoidea). Scale bar = 1 mm.

No myodocopans, and only three podocopan superfamilies, are found today in non-marine waters (which can include saline as well as fresh, flowing (lotic) and standing (lentic), temporary and permanent waterbodies) (Figure 6). The Cypridoidea are the most successful, with desiccation-resistant eggs allowing many taxa to inhabit temporary ponds, survive periods of drought, and achieve wide dispersal (eggs being wind-blown or carried by birds and other animals); they are benthonic/nektobenthonic, and herbivores, detritivores, predators and scavengers. Several families of the Cytheroidea have non-marine representatives, most notably the Limnocytheridae, of which many have female brood care and live in permanent lakes (but a few have desiccation-resistant eggs and live in temporary ponds). Most Darwinuloidea live in freshwater lakes, rivers, and springs; benthonic crawlers, they may filter-feed as well as grazing on algae and detritus, and the exclusively parthenogenetic females brood eggs and early instars. Some non-marine ostracods have highly specialized habitats, such as the limnocytherid genus *Elpidium* which lives in the water held by bromeliad plants in South America. Some freshwater ostracods are hypogean (predominantly cypridoidean Candonidae), living in deep, well-oxygenated ground water; others are found where ground water reaches the surface in springs and seepages, including hot springs up to 54°C.

Some representatives of essentially freshwater ostracod groups inhabit merely damp environments such as fen soils, mosses, and leaf-litter, including species of Cypridoidea, Cytheroidea, and Darwinuloidea. The Terrestricytheroidea are thought to have entered such habitats from marine littoral environments, however, and are exclusively associated with high intertidal and supratidal vegetation.

Geological History and Evolution

Thanks to an excellent fossil record (the best of any arthropod group) the geological history of the Ostracoda is well-documented and understood in most marine and non-marine environments (Figure 7). Some ostracod habitats have poor preservation potential, however: commensal ostracods (marine and non-marine) and semi-terrestrial taxa have little or no fossil record. The fossil record of Ostracoda begins in the Ordovician and their origins and relationships to other crustacean groups are obscure.

Ordovician ostracods were marine, probably meiobenthonic (possibly nektobenthonic) animals; faunas dominated by Palaeocopida (beyrichicopines and binodicopines) but including Podocopida (e.g., sigillioideans), Platycopida and Leperditicopida, were essentially confined to shelf areas and show lithofacies-related assemblages that can be related to depth; marginal marine tidal flat assemblages dominated by leperditicopids and palaeocopids are recognized, but there is no evidence for deep marine (bathyal–abyssal) faunas. The earliest myodocopans, marine nektobenthonic forms, appear in the Late Ordovician (Figure 7).

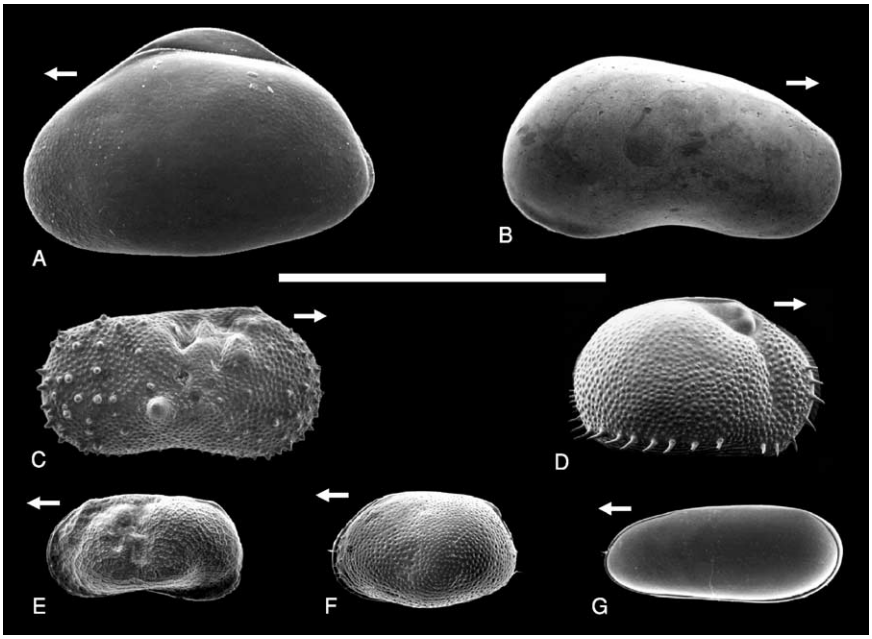


Figure 6 Shells of some Recent freshwater podocopid ostracods. A–D: Cypridoidea; E–F: Cytheroidea (Limnocytheridae); G: Darwinuloidea. Scale bar = 1 mm.

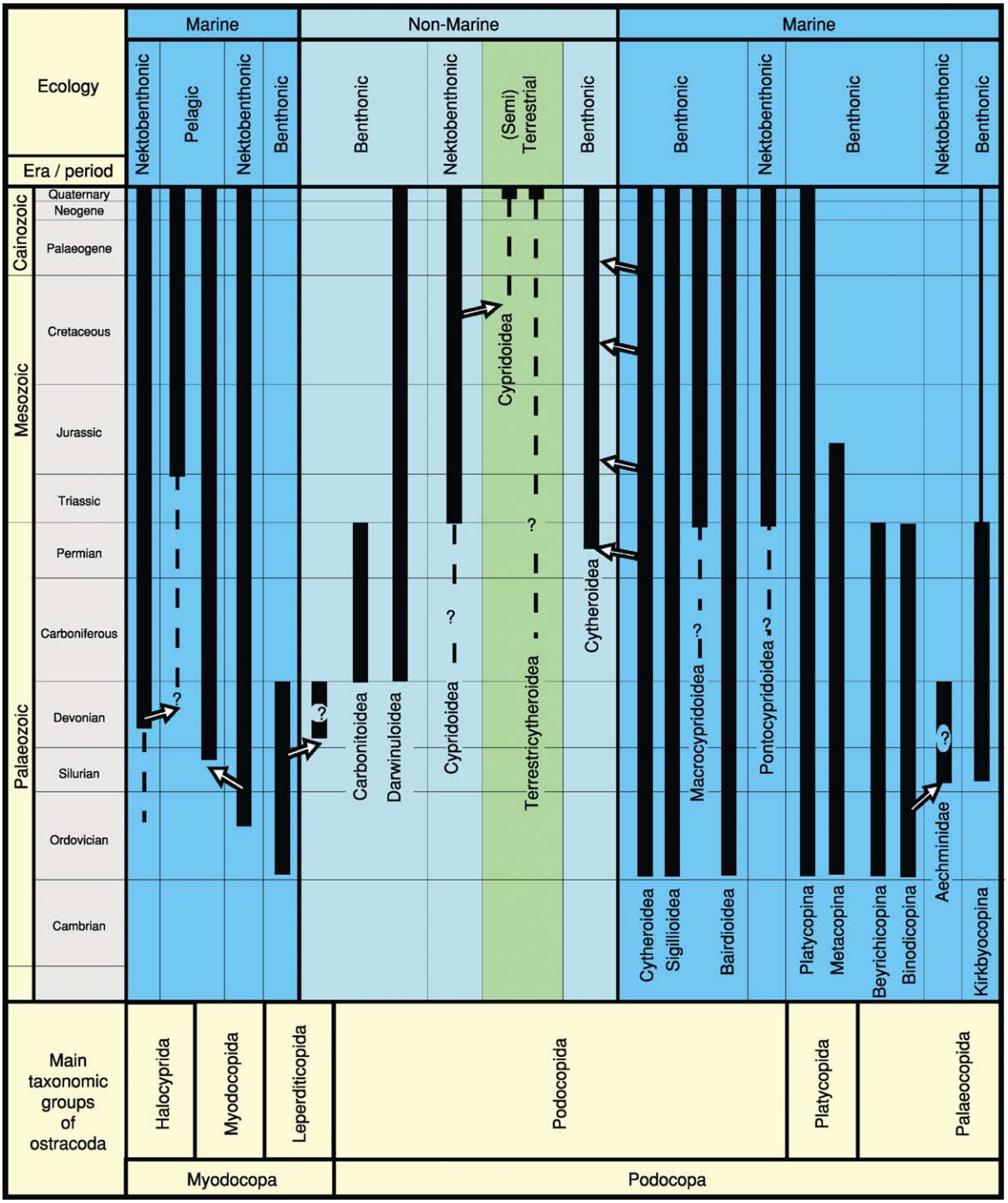


Figure 7 Stratigraphical time-lines of the major ecological radiations of the Ostracoda.

In the Silurian, some palaeocopids (Figure 8), platycopids, and leperditicopids may have begun to adapt to marginal marine, brackish water (and perhaps hypersaline) conditions; diverse palaeocopids dominated shallow marine shelf waters, giving way to 'non-palaeocope' faunas (including metacopines) in deeper water. The Ordovician–Devonian leperditicopids include the largest known ostracods (up to 50 mm long) and inhabited marginal marine environments, including tidal flats and estuaries, possibly entering freshwater environments in the Devonian; it has been speculated that they occupy an ancestral

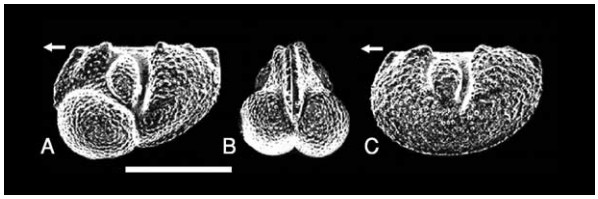


Figure 8 A Silurian palaeocopid ostracod (Beyrichicopina). A–B: female, lateral and anterior views (note large anteroventral brood pouches); C: male. Scale bar = 1 mm.

position in myodocopan evolution, but their affinities are controversial and there is some doubt that they are true ostracods.

The exclusively marine bairdioideans, one of the oldest extant podocopid lineages, are known from the mid-Ordovician onwards; they show a considerable variety of both deep- and shallow-water taxa by the Devonian, although post-Palaeozoic forms are most diverse in shallow carbonate environments. The podocopid cytheroidean family Bythocytheridae originated in the Ordovician, possibly sharing a common ancestor with the Bairdioidea, and underwent major radiations in the Devonian and again in the Jurassic–Cretaceous; while many cytheroidean families include at least some taxa that have adapted to brackish waters, bythocytherids never seem to have achieved this and are found only in fully marine salinities.

Distinctive Palaeozoic marine assemblages from the Devonian onwards have been related to shallow benthonic (high energy; dominated by thick-shelled ornamented podocopans), deep benthonic (low energy; thin-shelled spinose podocopans), and pelagic environments (myodocopans). The low-energy benthonic assemblage may represent a 'palaeopsychrospheric' fauna inhabiting deep cold waters below the thermocline in a stratified ocean, but this interpretation remains controversial.

The pioneer colonization of pelagic marine environments by myodocopan ostracods took place in mid-Silurian (Figure 7), after an ecological shift from a nektobenthonic mode of life, probably in response to bottom water oxygen deficiency and the opportunities offered by the plankton-rich, well-oxygenated upper waters. Another myodocopan group, the halocypridoidean halocyprids, may have undergone a similar shift, but they have an extremely poor fossil record and evidence for the existence of pelagic halocypridans since the Devonian, as shown on Figure 7, is limited and controversial; in the Silurian – Carboniferous they are represented by entomozooideans, a group of questionable affinity. A Silurian–Devonian palaeocopid binodocopine family, the aechminids, with a large, hollow dorsolateral spine on each valve, are also thought to have been pelagic, since it is difficult to imagine such a carapace morphology as functionally suitable for a benthonic organism; since the end of the Palaeozoic, however, there have been no pelagic podocopans.

The first undoubted non-marine ostracods, the podocopid superfamilies Carbonitoidea and Darwinuloidea, entered freshwater environments in the Early Carboniferous and radiated to high diversity in the Late Carboniferous–Permian; at the end of the Permian the former became extinct and the latter reduced in diversity (Figure 7). Post-Palaeozoic non-marine faunas are dominated by Cypridoidea and limnocytherid Cytheroidea; both lineages may have Late Palaeozoic origins but their early history is controversial, as is the possibility of a Carboniferous brackish-freshwater radiation of some platycopids, which today are an exclusively marine group. Ostracods, which may represent the first limnocytherids, proliferated in Late Carboniferous and Permian lakes and coal swamps. It is notable that darwinuloideans, which in Palaeozoic assemblages show distinctive sexual dimorphism (the females having expanded posterior brood chambers), have apparently been exclusively parthenogenetic since the Early Mesozoic.

Almost all of the Palaeocopida, so characteristic of Palaeozoic marine faunas, became extinct at the end of the Permian; only the Puncioidea survived and are represented today by a single living species. Post-Palaeozoic marine faunas are dominated by cytheroidean Podocopida. Major radiations of cytheroidean families took place in the Mesozoic, for example the Cytheruridae (Triassic onwards), the Progonocytheridae and Schulerideidae (mid-Jurassic–Early Cretaceous), the Trachyleberididae and Brachycytheridae (Late Cretaceous onwards), and the Cytherettidae, Hemicytheridae, Loxoconchidae, Leptocytheridae, and Xestoleberididae (Tertiary). The origins of the diverse modern deep-sea (psychrospheric) benthonic ostracod fauna (predominantly cytheroideans) were in Mesozoic faunas which evolved in a thermospheric ocean and were forced to adapt to cooling conditions in the Tertiary. Metacopine platycopids radiated in the Triassic and Early Jurassic and then became extinct. Platycopina were often the dominant group (in terms of abundance rather than diversity) in the chalk seas of the Late Cretaceous; their highest diversity today is in sub-tropical or tropical shallow carbonate environments.

In contrast to podocopans, the fossil record of myodocopans is poor, since in many the valves are weakly or not at all mineralized. The nektobenthonic cladocopine halocyprids are usually better calcified, however; they originated in the Palaeozoic and became relatively diverse in the Mesozoic, achieving their greatest diversity in the deep (bathyal) waters of

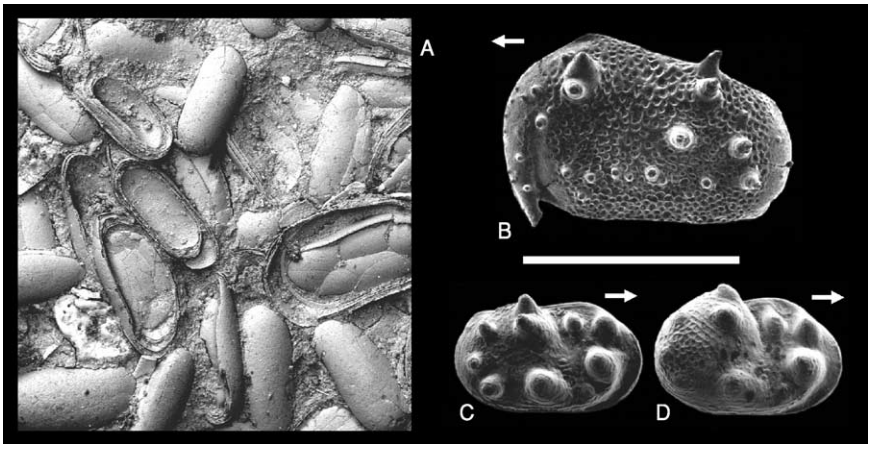


Figure 9 Some Early Cretaceous nonmarine ostracods. A: bedding-plane assemblage, mainly Darwinuloidea; B: *Cypridea* (Cypridoidea); C–D: Cytheroidea, Limnocytheridae (C: male; D: female, with expanded posterior brood chamber). Scale bar = 1 mm.

the modern Arctic Ocean, where competing pandemic deep-sea cytheroideans are excluded by physical barriers such as the shallow Bering Strait.

Mesozoic to modern non-marine faunas are dominated by Cypridoidea, but include significant cytheroidean (particularly the Limnocytheridae) and darwinuloidean components (Figure 9). The marine cytheroidean family Cytherideidae also gave rise to significant post-Palaeozoic non-marine radiations in brackish, hypersaline, and freshwaters. A major Mesozoic radiation of cypridoidean podocopids took place in non-marine environments from the latest Jurassic onwards. A key factor may have been their possession of desiccation-resistant eggs, enabling them to inhabit temporary water bodies and achieve wide dispersal. However, much of the Early Cretaceous cypridoidean diversity is attributable to one family, the Cyprideidae (*Cypridea* and related genera), which subsequently became extinct in the Palaeogene, although they co-existed with representatives of extant families; the reasons for their differential success and subsequent demise are obscure.

Applications

Benthonic marine ostracods, being facies-controlled, are of limited value for wide-scale biostratigraphy, but their small size, morphological diversity, and excellent fossil record have nevertheless rendered them extremely useful on regional or intrabasinal scales from the Ordovician to the Quaternary. Ostracods have also been used with great success in establishing zonation schemes and correlations in non-marine depositional settings, such as the Mesozoic 'Wealden' basins of England.

Quaternary deep-sea benthonic ostracod assemblages characteristic of different water masses have considerable potential for palaeoceanographic and palaeoclimatic studies, such as the identification of mode switches in thermohaline circulation. Ostracods were used to great effect in documenting the development of a stratified ocean (with a shallow, warm thermosphere overlying a deep, cold psychrosphere) during the Cenozoic, as well as other palaeoceanographic events, such as the opening of the Drake Passage and the closure of the Panama Isthmus.

Shallow and marginal-marine ostracods have proved valuable as indicators of sea-level change, both by the recognition of depth-related shelf assemblages and by the identification of intertidal taxa as sea-level proxies.

Two groups of marine ostracods have been used as fossil proxies for dissolved oxygen levels. Platycopids, as filter-feeders and brooders, may be better able than podocopids to survive episodes of low oxygen availability in the benthic marine environment, by virtue of more efficient mechanisms (branchial plates) for circulating water (for respiration) through the domiciliar and brood spaces in the carapace. The 'platycopid signal' indicative of low-oxygen conditions, first recognized in differential success of platyocopids compared to podocopids during the mid-Cretaceous Cenomanian–Turonian Oceanic Anoxic Event, has been widely applied, not only in the Mesozoic but also in the Palaeozoic; in the latter case palaeocopids, metacopines, and platycopines have all been counted as filter-feeders. Variation in the size and shape of the vestibulum in the podocopid cytheroidean genus *Krithe* and its close relatives has also been used (controversially) as an indicator of mid-Cretaceous to Quaternary marine oxygen levels. However, it must be noted that neither of these attractive applications is so far adequately supported by evidence of the biology, ecology and distribution of modern taxa.

Ostracods are particularly effective for palaeoenvironmental analysis in marginal marine settings, bearing in mind the need to distinguish between autochthonous and allochthonous components of assemblages (the former representing the local enviroment, the latter being mixed by post-mortem transport, perhaps by tidal currents, from a wider range of environments). The transition from marine to brackish conditions is usually reflected in changes in species composition and a reduction in the diversity of assemblages. Intraspecific variation, for example in carapace ornament, can often be related to different environmental conditions, although caution is advised since different taxa undoubtedly respond in different ways. In the living euryhaline cytherideid *Cyprideis torosa* (a cytheroidean podocopid) variations in the development of nodes and the shape of sieve pores are influenced by salinity, and this knowledge has been effectively applied to the interpretation of Neogene and Quaternary fossil examples; the relationships are complex, but noded forms with round sieve pores may generally be taken to indicate low-brackish salinities, while smooth forms with irregular pores record hypersaline conditions.

Differences in non-marine ostracod assemblages reflect different water chemistries and habitat types as well as overall salinity. That Mesozoic Purbeck-Wealden faunas of England are essentially non-marine is as much evident from the absence of platycopids and bairdioideans as from the abundance of cypridoideans (especially *Cypridea* and its allies), cytheroideans (limnocytherids and cytherideids), and darwinuloideans (**Figure 9**); some cytheroidean taxa with marine affinities may indicate periodic marine incursions in coastal settings, or alternatively the existence of saline inland lakes. Darwinuloideans and some limnocytherids had brood care and must have lived in permanent water bodies, while many *Cypridea* species may represent temporary ponds, perhaps in ephemeral river beds. Limnocytherids have also been related to high alkalinity waters, *Cypridea* to low alkalinities.

Quaternary non-marine ostracods can be interpreted in similar ways and it has been shown that different species inhabit different lakes with the same salinity (ionic concentration) but different water chemistry (ionic composition). They are increasingly being used as sources of biogenic calcite for the analysis of trace metals, stable isotopes, and even amino acids, with applications in palaeoclimatology (e.g., as palaeotemperature and palaeosalinity proxies) and age determination. Temperature is an important influence on the geographical distribution of both marine and non-marine ostracods, and many taxa are effective indicators of cold (glacial) and warm (interglacial) conditions. Ostracods are frequently included in multi-proxy studies of Quaternary archaeological sites.

See Also

Fossil Invertebrates: Arthropods.

Further Reading

Athersuch J, Horne DJ, and Whittaker JE (1989) *Marine and brackish water ostracods*. Synopses of the British fauna (New Series), No. 43. EJ Brill: Leiden.

Benson RH (1990) Ostracoda and the discovery of global Cainozoic palaeoceanographical events. In: Whatley R and Maybury C (eds.) *Ostracoda and global events*, pp. 41–58. London: Chapman and Hall.

Carbonel P, Colin J-P, Danielopol DL, Löffler H, and Neustrueva I (1988) Palaeoecology of limnic ostracodes: a review of some major topics. *Palaeogeography, Palaeoclimatology, Palaeoecology* 62: 413–461.

De Deckker P, Colin J-P, and Peypouquet J-P (eds.) (1988) *Ostracoda in the Earth Sciences*. Amsterdam: Elsevier.

Holmes JA and Chivas AR (eds.) (2002) *The Ostracoda: applications in Quaternary research*. AGU Geophysical Monograph **131**. Washington DC: American Geophysical Union.

Griffiths HI and Holmes JA (2000) *Non-marine ostracods and Quaternary palaeoenvironments*. Quaternary Research Association Technical Guide No. 8. London: Quaternary Research Association.

Horne DJ (2002) Ostracod biostratigraphy and palaeoecology of the Purbeck Limestone Group in southern England. *Special Papers in Palaeontology* 68: 53–70.

Martens K (ed.) (1998) *Sex and parthenogenesis. Evolutionary ecology of reproductive modes in non-marine ostracods*. Leiden: Backhuys.

Meisch C (2000) *Freshwater Ostracoda of western and central Europe*. Susswasserfauna von Mitteleuropa **8/3**. Heidelberg: Spektrum Akadamischer Verlag.

Park LE and Smith AJ (eds.) (2003) *Bridging the gap: trends in the ostracode biological and geological sciences*. Palaeontological Society Papers **9**.

Siveter DJ (1984) Ecology of Silurian ostracods. *Special Papers in Palaeontology* 32: 71–85.

Siveter DJ, Vannier JMC, and Palmer D (1991) Silurian Myodocopes: pioneer pelagic ostracods and the chronology of an ecological shift. *Journal of Micropalaeontology* 10(2): 151–173.

Siveter DJ, Sutton MD, Briggs DEG, and Siveter DJ (2003) An ostracode crustacean with soft parts from the Lower Silurian. *Science* 302: 1749–1751.

Whatley R (1995) Ostracoda and oceanic palaeoxygen levels. *Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut, Hamburg* 92: 337–353.

Williams M, Floyd JD, Miller CG, and Siveter DJ (2001) Scottish Ordovician ostracodes: a review of their palaeoenvironmental, biostratigraphical and palaeobiogeographical significance. *Transactions of the Royal Society of Edinburgh: Earth Sciences* 91: 499–508.

Palynology

P Coxon and G Clayton, Trinity College, Dublin, Ireland

Introduction

'Palynology' is a term that was introduced in 1944 to cover a broader range of studies than was covered by the earlier term, 'pollen analysis', which had distinct Quaternary connotations. Initially, palynological research dealt mainly with spores and pollen, but many other types of organic-walled microfossils were subsequently investigated. The useful term 'palynomorph', thus includes a broad spectrum of microfossils (such as the miospore shown in Figure 1) that are resistant to the mineral acids and other reagents commonly used for fossil extraction from rocks and sediments. As a consequence of the inclusive definition, based on chemical composition of the microfossil wall, palynology studies encompass microfossils of diverse biological affinities, both plant and animal. Although many different techniques for the study of palynomorphs have been employed, by far the most commonly used are transmitted light microscopy for routine examination and scanning electron microscopy for the elucidation of surface structure and ornamentation, at much higher magnifications.

Major differences separate Quaternary and pre-Tertiary palynology. In the former, almost all taxa recorded are from extant plants, whereas in the latter, the taxa are almost invariably extinct. This is strongly reflected in the palynomorph classifications utilized, with pre-Tertiary palynomorphs typically being assigned to 'form' (i.e. morphology-based) rather than natural genera and species. Equally striking differences are apparent in terms of biostratigraphy, with Quaternary dating and correlation strongly tied to the recognition of changing climate but with pre-Tertiary biostratigraphy being normally based on the first appearances and extinctions of taxa in the stratigraphic record. The Tertiary represents a complex 'overlap' interval, with uncertainties concerning both the appropriate scheme for the classification of palynomorphs and the biostratigraphic approach adopted.

Figure 1 *Knoxisporites stephanephorus* Love 1960. A transmitted light photomicrograph of a typical Mississippian miospore.

Quaternary Palynology

Palynologists working on Quaternary material have, for the most part, concentrated primarily on the pollen of gymnosperms and angiosperms as well as on the spores of bryophytes and pteridophytes (this work being colloquially referred to as 'pollen analysis'). However, a range of Quaternary fossil material is considered as palynological and could be referred to here (including fungal spores). The problems of reconstructing complex Quaternary environments from a wealth of detailed evidence have led palynologists to use a wide range of fossil groups in order to provide multiproxy evidence that can allow more accurate models of past changes. There is also a wide literature on other microscopic fossil material, e.g., charcoal, cuticle fragments, phytoliths, and testate amoebae.

Quaternary palynology has the distinct advantage that taxa are, by and large, still extant, allowing the use of modern reference material collected from living plants in addition to keys, pollen and spore atlases, and World Wide Web-based identification pages offering digital images of pollen and spores. Such a wealth of information allows identification of Quaternary (and some Tertiary) fossil palynological material to the level of family, often to genus, and sometimes to species, giving the Quaternary palynologist a wide range of options with regard to palaeoenvironmental reconstruction and other applications relevant to palynological data.

Early palynological research in the Quaternary concentrated on vegetation history and produced (albeit initially unrefined) palaeoecological reconstructions. This research quickly became more sophisticated,

leading to the gathering of much detailed information on former plant distributions and to a tradition of palaeobotanical studies that continues, with considerable enhancement of the technique, to this day. The palynological elucidation of detailed vegetation histories of north-western Europe was followed by attempts to produce biostratigraphic correlations, which, in some instances, merged into climatostratigraphies. Holocene palynological research was inevitably linked to the human influence on both landscape and vegetation and early work began to show the importance of the anthropogenic impact on natural ecosystems. Palynological research into such influence on the environment is still of primary importance today, and a detailed knowledge of past environment has allowed present and future management and protection of sensitive ecosystems to be better planned.

Problems are inherent in the time-transgressive nature of temporally and spatially detailed Holocene palynological records; this has caused obvious problems with correlation and has led to a movement away from relying on biostratigraphy deduced from Holocene sequences. For Late Pleistocene and Holocene work, radiocarbon dating has permitted detailed correlation and synthesis of palynological data from sequences younger than, in theory, *ca.* 40 000 years, although the most certain dates come from within the limits of reliable radiocarbon calibration methodologies (<13 000 years). Even within these limits, the complexities of atmospheric radiocarbon production severely hamper attempts to date important environmental/geological transitions (e.g., the radiocarbon 'plateau' of the critical last glacial–interglacial transition some 13 000–10 000 radiocarbon years ago).

Despite calibration problems with radiocarbon dating, the ability to date palynological assemblages from the Holocene has enabled palynologists to map the timing of migration and vegetational succession in great detail, allowing isopoll and isochron maps to be constructed and giving palynologists a great deal of evidence regarding plant migration, vegetational succession, and the timing of climatic (and other environmental) impacts on vegetation. Figure 2 shows reconstructions of European vegetation zones at four different time periods. The use of radiocarbon-dated palynological data allows such detailed palaeoecological maps to be presented. However, biostratigraphy remains important in many older (Pleistocene) palynological records, for which 'pollen analysis' has long been used to characterize the oscillating nature of the climate at a coarse level (i.e., warm stage/cold stage) and at a high resolution (e.g., the analysis of vegetation history and biostratigraphy of interglacial or temperate-stage sequences). Figure 3 shows the palynological record from four successive interglacials from southern France. Pollen diagrams such as these from interglacial sequences have an important role in Pleistocene biostratigraphy, palaeoecological studies, and biogeographical analyses, and in understanding regional and continental vegetational histories.

The distinctive nature of successive temperate stages is emphasized in Europe by the geographical 'trap' that is formed during cold stages, when climatic conditions deteriorate and thermophilous taxa become locally extinct. During cold stages, plant taxa inhabit refugia to the south and east, but there are severe barriers to migration (e.g., the Pyrenees, the Alps, and the continentality of eastern Europe and the Mediterranean). Each successive cold stage saw extinction of taxa from Europe, and so throughout the Pleistocene, the number of taxa recorded in each interglacial generally declines. This decline in represented taxa provides unique assemblages in temperate stages and the differing rates of migration of taxa, and their relative success provides distinctive vegetational successions in each interglacial (see Figure 3).

More recently, problems with using biostratigraphy on short or incomplete sequences have been highlighted, and the use of advanced dating methodologies and mammalian biostratigraphy (e.g., in the UK and north-western Europe) has suggested that there has been a compression of the number of temperate stages that actually exist into a smaller number of named stages, because of a lack of palynological resolution. The 'missing stages' are glaringly apparent from the far more complete marine oxygen isotope records. However, the apparent resolution in the biostratigraphy of the interglacial sequences in south central France (Figure 3) shows the value of palynological work on complete temperate-stage deposits, and palynology still holds great value for analysing temperate-stage palaeoenvironments.

Inferences regarding climate change are of paramount importance in research stemming from modern palynology, and the need to understand climate variability has led to palynological data becoming particularly important as a proxy data source for climatic parameters. In this regard, the use of taxa to act as a proxy for climate variables was recognized early on; this approach has continued to be used as models of the relationships between individual taxa, and even whole fossil assemblages, and climate parameters are clarified. This research has proved particularly important in large-scale modelling of global climate change using fossil data from widespread regional sources.

The most recent geological period, the Quaternary, is characterized by dramatic shifts in climate forced

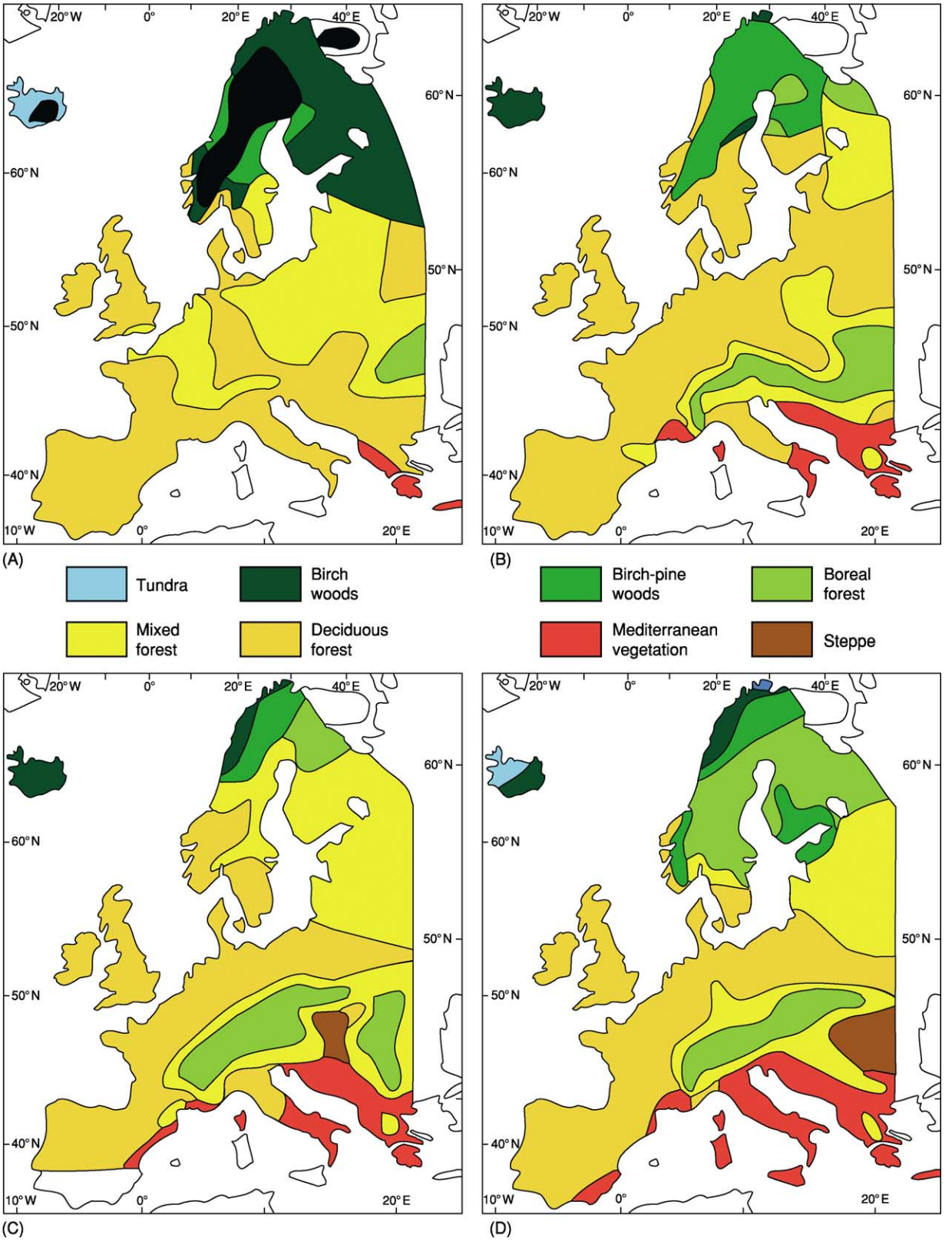


Figure 2 Vegetational regions of Europe reconstructed from pollen data for (A) 9000, (B) 6000, and (C) 3000 years BP and (D) the present. After Huntley B and Prentice IP (1993) Holocene Vegetation and Climates of Europe. In: Wright HE Jr, *et al.* (eds.) *Global Climates Since the Last Glacial Maximum*. University of Minnesota Press.

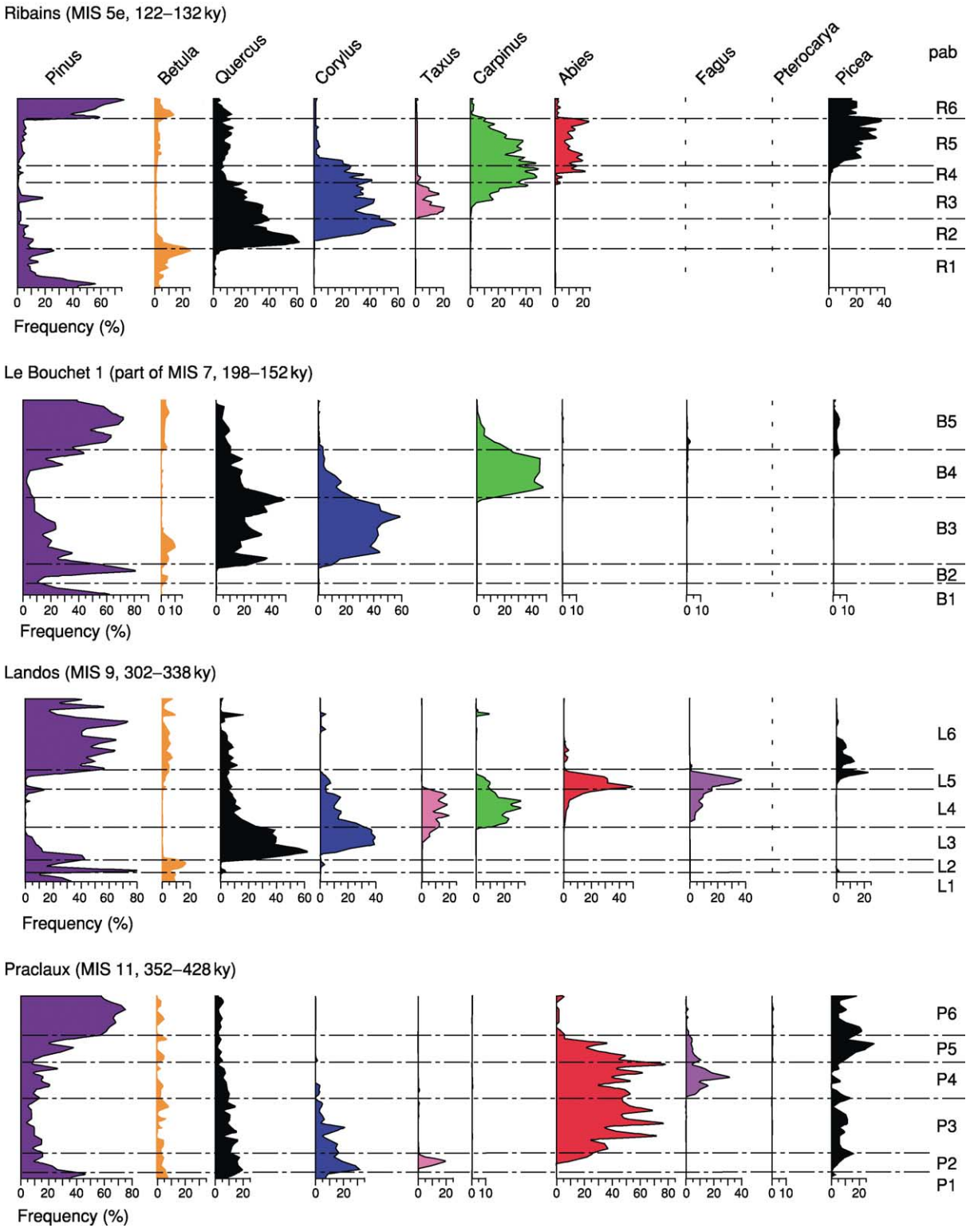


Figure 3 Relative percentage pollen diagrams (selected taxa) of four successive Pleistocene temperate stages (interglacials) from the Velay region of France. The correlation to marine isotope stages (MIS) and tentative ages of the deposits are shown. The behaviour of some of the individual taxa (coloured curves) during each temperate stage and the unique succession of pollen assemblages are apparent. Possible local pollen assemblage biozones (pab) are indicated; these can be subdivided as necessary to produce more detailed descriptions of assemblages. The regional disappearance of taxa – e.g., *Pterocarya*, after the Middle Pleistocene (MIS 11), and *Fagus*, by the last interglacial (MIS 5e) – is also apparent. After De Beaulieu JL and Reille M (1995) Pollen records from the Velay craters: A review and correlation of the Holsteinian Interglacial with isotopic stage 11. *Mededelingen van de Geologische Dienst* 52: 59–70.

by the astronomical position of Earth relative to the sun. The problems of reconstructing complex Quaternary environments have led palynologists to use a wide range of fossil groups in order to provide multiproxy evidence that can allow more accurate models of past changes.

Pre-Quaternary Palynology

Hydrocarbon exploration provided the main impetus for palynological biostratigraphy (or 'palynostratigraphy') in the second half of the twentieth century, due largely to the realization that small fragmentary rock samples, such as 'ditch cuttings' from drilling, could yield thousands of intact palynomorphs. The erection of zonal schemes for non-marine successions based on spores and pollen in the Carboniferous (Figure 4) was followed rapidly by the publication of similar schemes for the Devonian and later systems. The stratigraphic resolution of these palynological zonations compares very favourably with many schemes based on invertebrates; the average duration of Upper Palaeozoic spore and pollen biozones/sub-biozones is less than 3 Ma.

Detailed zonations of marine Palaeozoic rocks have been erected, based mainly on acritarchs (*see* **Microfossils:** Acritarchs) and chitinozoa (*see* **Microfossils:** Chitinozoa), whereas palynostratigraphic work on marine Mesozoic and Tertiary successions has relied principally on dinoflagellate studies. In relatively recent times, palynological investigations of Precambrian rocks have produced spectacular results in terms of both enabling much needed biostratigraphic correlation and shedding light on the nature of early life in Earth's oceans.

The limited palaeogeographical distribution of many pre-Quaternary spore and pollen species restricts the applicability of most palynological zonations to specific regions, and in most cases rules out any possibility of 'global' correlation. However, the converse of this obvious disadvantage is that floral provinces can be recognized, which can complement palaeomagnetic and other evidence in making large-scale palaeogeographical reconstructions.

Other Geological Applications

Not all Phanerozoic palynological research has been biostratigraphic. Inspired by palaeoecological studies of Quaternary peats, pioneering work on spore and pollen assemblages from Palaeozoic coal seams led to convincing interpretations of the gradually changing environmental conditions responsible for the accumulation of the peat. More comprehensive

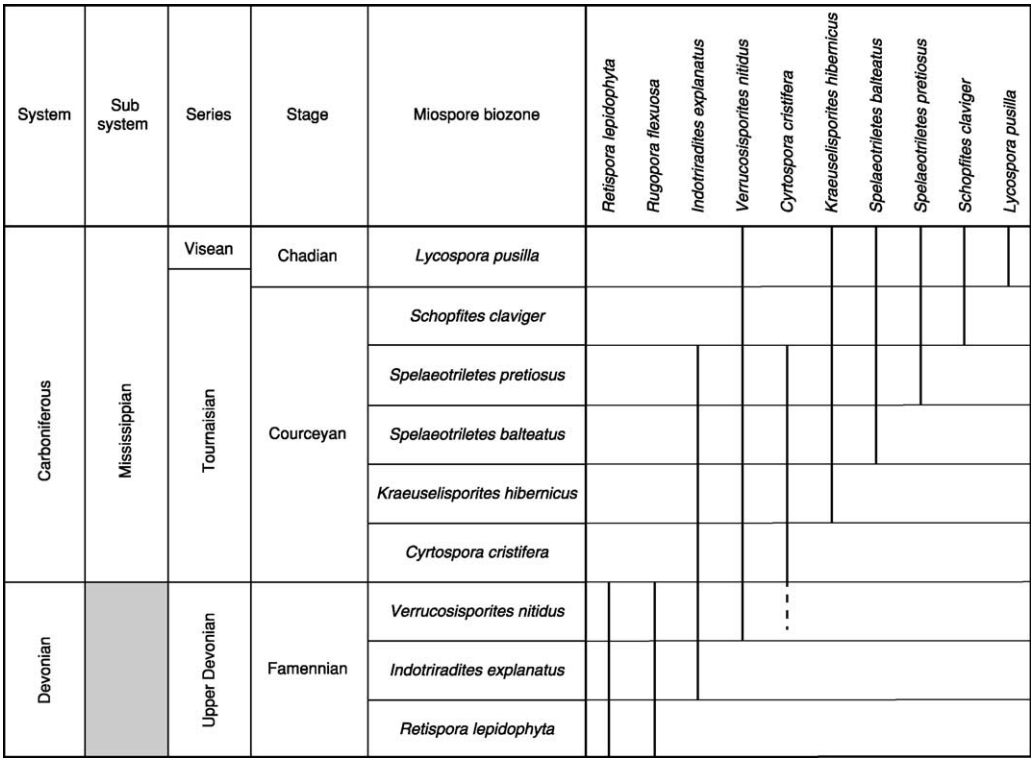


Figure 4 Part of the miospore zonation for the Late Devonian and Early Carboniferous of western Europe, showing the stratigraphic ranges of selected miospore taxa.

Figure 5 Specimens of the acritarch *Veryhachium* from four Devonian samples of different maturity. With increasing temperature, the colour changes from almost colourless, through yellow, to brown, and eventually to black.

palaeoenvironmental interpretations based on palynological investigations of lithologies other than coal are often referred to as 'palynofacies analyses'. These studies extend beyond the identification of just the palynomorphs present, to consideration of all of the 'palynodebris' – the microscopic organic matter present, including particles such as woody debris, charcoal, and cuticle. Interpretations based on palynofacies analysis typically enable distinction to be made between sediments deposited in marine and non-marine environments, and, in the former case, permit some estimate of proximity to shoreline and energy level of the depositional environment.

A further example of the non-stratigraphic applications of palynology concerns the gradual carbonization of organic matter resulting from increased temperature. The 'thermal maturity' of sediments is critical in terms of hydrocarbon generation but is often difficult to assess accurately. The irreversible darkening of palynomorphs that results from carbonization (Figure 5) can be measured and may be used in place of traditional methods such as vitrinite reflectance determination to estimate maturity in rocks deficient in vitrinite.

See Also

Biozones. **Microfossils:** Acritarchs; Chitinozoa. **Palaeoclimates**. **Palaeoecology**. **Petroleum Geology:** The Petroleum System. **Tertiary To Present:** Pleistocene and The Ice Age.

Further Reading

Batten DJ (1999) Palynofacies analysis. In: Jones TP and Rowe NP (eds.) *Fossil Plants and Spores: Modern Techniques*, pp. 194–198. London: Geological Society.

Clayton G and Coxon P (1999) Spore and pollen biostratigraphy. In: Jones TP and Rowe NP (eds.) *Fossil Plants and Spores: Modern Techniques*, pp. 225–229. London: Geological Society.

Donald G (1996) Non-Aquatic Quaternary. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications*, pp. 879–910. Salt Lake City: American Association of Stratigraphic Palynologists Foundation.

Fensome RA, Riding JB, and Taylor FJR (1996) Dinoflagellates. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications*, pp. 107–169. Salt Lake City: American Association of Stratigraphic Palynologists Foundation.

Jarzen DM and Nichols DJ (1996) Pollen. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications*, pp. 261–291. Salt Lake City: American Association of Stratigraphic Palynologists Foundation.

MacDonald GM (1996) Non-Aquatic Quaternary. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications*, pp. 879–910. Salt Lake City: American Association of Stratigraphic Palynologists Foundation.

Miller MA (1996) Chitinozoa. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications*, pp. 307–336. Salt Lake City: American Association of Stratigraphic Palynologists Foundation.

Moore PD, Webb JA, and Collinson ME (1991) *Pollen Analysis*, 2nd edn. London: Blackwell Scientific Publications.

Playford G and Dettmann ME (1996) Spores. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications*, pp. 227–260. Salt Lake City: American Association of Stratigraphic Palynologists Foundation.

Reille M (1992) *Pollen et Spores d 'Europe et d' Afrique du Nord*. Marseille: Laboratoire de Botanique Historique et Palynologie.

Strother PK (1996) Acritarchs. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications*, pp. 81–106. Salt Lake City: American Association of Stratigraphic Palynologists Foundation.

Wright HE Jr, Kutzbach JE, Webb T III, Ruddiman WF, Street-Perrott FA, and Bartlein PJ (eds.) (1993) *Global Climates Since the Last Glacial Maximum*. Minneapolis: University of Minnesota Press.

MICROPALAEONTOLOGICAL TECHNIQUES

I J Slipper, University of Greenwich, Chatham Maritime, UK

Introduction

The techniques used in the study of micropalaeontology tend not to be discussed in any great detail in most published accounts; many reports dismiss the subject of methods used with 'standard techniques were employed' or some such statement. This is often a mistake, because each group of microfossils will benefit from different treatments and each rock type will present challenges to the technician. To judge results of any scientific work, it is necessary to know which methods were used (e.g., during sample preparation). The focus of this article is the many ways in which microfossils may be extracted from their matrix and then prepared for study. Topics covered include field sampling/procedures and laboratory techniques, including extraction of microfossils of different compositions, such as calcareous, organic walled, phosphatic, and siliceous, from rocks of varying lithology. General methods of washing and sieving and specialist methods of separation, such as flotation and magnetic separation, are also covered.

Field Procedures

The approach in the field varies according to the type of work undertaken and the fossil group being studied. The sampling interval should be commensurate with the objectives: a high-resolution study may require taking samples every 10 cm whereas a reconnaissance survey can be carried out with samples taken every metre. The amount taken in the field is also dependent on the fossil group, on the fossil sizes, and the estimated abundance of specimens. Sample quantities taken in the field are usually larger than is needed for laboratory study. This allows some of the raw sample to be held in reserve for use either as reference material or as replacement in case of mishaps, or, more likely, when the results prove that further study of a different nature is required. The most important aspect of field procedure is awareness and avoidance of possible contamination from sources other than that of interest. The route to avoiding contamination is routine cleaning of the sampling equipment between taking successive samples. Of equal importance is good record keeping and labelling. Accurate logs should be drawn up in the field and the horizons sampled must be measured in relation to key marker horizons. Microfossils may also be found in abundance on bedding surfaces, such as ostracods in the Cretaceous Wealden Clay. In these cases, laboratory processing will destroy many specimens, and it is far better to recover large oriented blocks and to examine the surfaces directly under a binocular microscope. Where spot sampling is undertaken, the procedure is to remove all weathered material from the surface of the outcrop and to excavate a clean, unweathered block of the sediment, which is then placed in a clean labelled bag for transport and storage.

Extraction Methods

The techniques that have been devised to extract microfossils from their matrix are as varied as the number of different fossil groups and the types of rock in which they are found. There is no single overall method for extracting microfossils from rock, though some general-purpose methods may enhance achieving this. To be successful, a preparator selecting a method needs to be aware of the composition not only of the fossils concerned, which may be calcareous, siliceous, phosphatic, organic, or, in some cases, combinations of these, but also of the composition of the matrix, which could be any of the manifold types of sedimentary rocks. In cases in which the compositions of the microfossil and matrix differ, such as ostracods in a clay, simple methods may suffice, but in cases in which the two are similar, such as foraminifera in chalk or radiolaria in chert, the processes involved become complex. In these cases, other factors, such as differences in grain size, crystallinity, structure, and porosity of the matrix, must become the focus of attention for the method.

The methods surveyed here have been devised by many workers in laboratories around the world looking for efficient ways to separate the precious microfossils from the rock, while retaining a representative sample of the original assemblage in terms of abundance and diversity, in addition to minimizing any damage to the fragile specimens; which are easily damaged during processing. Some of the processes are dangerous and should be undertaken only in a properly equipped laboratory with fume cupboards and protective clothing, whereas others can easily be carried out with minimal equipment, such as an oven, a sieve, and some water. Though the techniques may be associated with certain groups, many

methods are suitable for use across different groups. The selection of the method, therefore, will depend as much on the facilities and methods of the laboratory and the experience of the technician as on the 'correct' laboratory manual method.

Calcareous Microfossils

Foraminifera (*see* **Microfossils:** Foraminifera) and ostracods (*see* **Microfossils:** Ostracoda) both secrete calcitic material that is preserved in the fossil record, and are generally of a similar size and are often found together in the same environments. There are exceptions to this, but techniques suitable for one group may be applied to the other. Calcareous nannofossils, however, by virtue of their much smaller size, require slightly more specialized treatment.

Foraminifera and ostracods The sample size required is generally between 250 and 500 g; once cleaned, coarsely broken up, and oven dried, the sample is weighed prior to the start of processing. Weighing enables the specimen abundance to be estimated on completion of the laboratory work. During the drying stage, there is a possibility that any clay in the sample could be baked hard, rendering any further processing more difficult, so drying is carried out at low temperatures for many hours. The extraction of foraminifera and ostracods from argillaceous sediments is relatively straightforward and comes close to being a standard technique; this is because both the composition and the grain size differ between fossil and matrix.

If the sample is poorly consolidated, a simple boiling in distilled water may be sufficient to break down the rock; this can be done in a microwave oven or on a hotplate. Stirring the sample is avoided because this can easily damage specimens. Various agents can be added to the water if the initial boiling is ineffective. A detergent such as Teepol may help, but this has been shown to cause some dissolution of delicate calcitic material. One of the problems with clay-rich sediments is that they tend to be sticky and difficult to disperse. In these circumstances, a deflocculating agent will help, such as sodium carbonate (washing soda), sodium hydroxide, or sodium hexametaphosphate (Calgon). The ideal result is to turn the rock into a fine soup that may then be sieved to concentrate the residue.

Better lithified sediments, such as shales, require treatment harsher than simple boiling in a water softener. In these cases, a weak solution of hydrogen peroxide (15%) is used. On contact with the rock, the effect of the generation of small bubbles of oxygen within the pore spaces breaks the rock apart. However, hydrogen peroxide can cause dissolution of the microfossils, so the time spent immersed has to be kept to minimum. This is particularly important if the rock contains any pyrite, because reaction with hydrogen peroxide will create sulphuric acid, which further dissolves the microfossils. Other similar methods employ sodium hypochlorite (domestic bleach) or sodium hydroxide; these are less destructive but take a longer time to be effective.

For marly sediments, a very effective method is to use an organic solvent, such as petroleum spirit, paraffin, or white spirit (turpentine substitute). After the rock has been dried and all pore spaces opened, it is immersed in spirit for several hours. The soaking time is dependent on the porosity of the rock; for very fine-grained sediments, a longer time is required for the spirit to enter the rock. The spirit is then decanted off, filtered, and saved for reuse. The vessel containing the rock is then topped up with water, which will cause the sample to disintegrate and form a thick slurry. The slurry is then boiled in excess water and washed with a detergent before sieving.

If the rock type is carbonate rich, the difference between the composition of the fossil and the rock is minimal and chemical methods may not be effective. In these cases, physical breakdown of the rock is achieved by forces of crystallization within the pore spaces; chalks will respond well to this approach. The most commonly used material for this procedure is sodium sulphate decahydrate (glauber salt) in a supersaturated solution. After the rock chips have been dried in the oven, the still-warm sample is immersed in the solution and returned to the oven and allowed to soak thoroughly. When saturated, the liquid is decanted and the sample is placed in a freezer overnight. Rapid freezing has the effect of causing many nucleation sites, allowing fast growth of very small crystals in the pore spaces. This is sufficient to break the rock apart and has less destructive power than does slow crystal growth, and therefore gives better fossil recovery.

Calcareous nannofossils Calcareous nannofossils (*see* **Fossil Plants:** Calcareous Algae) as a group are composed of several forms, such as coccoliths, nannoconids, and discoasters. They range in size between 0.25 and 30 μm, and, due to this very small size, the number of specimens in fine-grained pelagic sediments can often exceed 1 million cm^{-3}. Therefore, when collecting in the field, only a very small sample size (approximately 1 cm^3) is required. Contamination must be very carefully avoided in all stages of nannofossil preparation: the surface of the rock must be scraped clean, cleaned tools must be used to extract the sample, and the sample must be placed in a clean, labelled bag. In the laboratory, any glassware

must be cleaned with a 10% solution of hydrochloric acid and rinsed with distilled water.

Several methods, each of increasing complexity, can be used for extracting nannofossils from the surrounding matrix. The simplest may be done rapidly by making a smear slide. This is a small scraping of sediment onto a glass slide; this specimen is then diluted and mixed with a little distilled water. The slide is dried on a hotplate and finished by mounting a cover slip with a medium that has a refractive index between the two extremes of calcite, 1.484 and 1.658. Caedax or Cellosize and Elvacite have been used successfully. These smear slides are useful for reconnaissance work, but for more detailed analysis, other methods are available. The most commonly used technique is separation by centrifuge, which has replaced methods such as wet sieving, gravity settling, and elutriation. The object is to prepare slides that contain a high proportion of nannofossil to sediment while retaining the diversity of the original sample. To achieve this, the sample is broken down by crushing or shaking in distilled water buffered with sodium carbonate to a pH of around 8 to 8.5. This should then pass through a 45μm sieve. Often a water softener such as sodium hexametaphosphate is added to deflocculate the clays, but some authorities avoid this because it has a slightly corrosive effect on the fossils. Similarly, ultrasonic vibration is used by some to disaggregate the sample further, but this has been reported to damage some specimens. Many variations on the centrifuge speeds have been suggested, but a simple two-stage method involves initially centrifuging with a low speed to settle out particles coarser than 30 μm, retaining the finer material in suspension, then a second stage of rapidly centrifuging the liquid to concentrate the particles down to 3 μm. Speeds and times will vary according to the equipment, so the instrument manufacturer's charts should be consulted for calculating the correct variables.

Nannofossils may be examined either by light or electron microscopy. For light microscopy, a drop of the suspension is placed onto a slide and prepared as for the smear slide. A high-powered research-grade optical microscope with magnifications of 800× or 1200× (oil immersion lenses) is required. Observations are made either by using polarizing attachments and viewing in cross-polarized light or by using phase-contrast methods. Though the results obtained in this way are rapid, and often used in industrial applications, more detailed images may be obtained by employing electron microscopy. The transmission electron microscope (TEM) has been used to see the structural relationships of the elements of the nannofossils, but the preparation for the TEM requires making carbon replicas of the specimens, which is a painstaking and skilled operation and is rarely used now. A faster and more convenient method is to coat the specimens in a thin layer of gold and observe using scanning electron microscopy (SEM). Figure 1 shows examples of specimens using various observation methods.

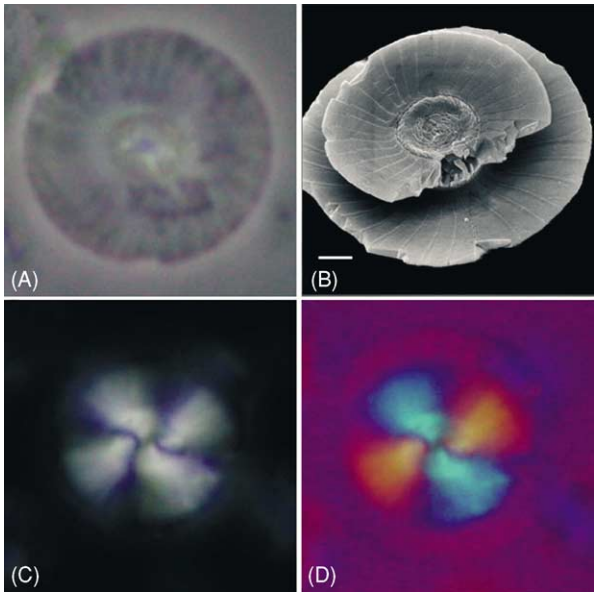


Figure 1 Calcareous nannofossils (*Calcidiscus*) visualized using different methods. (A) Phase contrast; (B) scanning electron microscopy; scale bar = 1 micron; (C) crossed polarized light; (D) gypsum tint plate. Photographs by Dr J Young.

Acid-Insoluble Microfossils

Conodonts Conodonts (*see* **Microfossils:** Conodonts) are found in marine rock types as diverse as limestones, shales, marls, ironstones, and cherts, and the methods of extracting them from their matrix are just as varied. However, unlike the calcareous microfossil groups, the phosphatic nature of conodonts renders them suitable for some acid preparations, and a variety of acids have been used for this purpose. Although conodonts may be visible on bedding surfaces in the field, it is far more usual to employ spot sampling at intervals that have been accurately logged. A bulk sample size of 1–2 kg should be sufficient to yield many conodonts. Because conodonts are durable and quite resistant to weathering and erosion, care needs to be taken to avoid contamination of samples from material washed down slopes. Again, only fresh rock should be sampled. For all lithologies, samples are pretreated by cleaning the exterior of the bulk sample, oven drying, and then coarsely crushing to centimetre-sized pieces.

Carbonate-cemented rocks will respond to diluted acids and many methods have been tried. The use

of monochloroacetic acid, although giving a rapid reaction, has been shown to damage conodont elements. Similarly, the use of unbuffered acids, such as formic and acetic acids, can also cause dissolution. However, when 10% formic acid (HCOOH) is buffered with calcium carbonate and tricalcium carbonate to a pH of less than 3.6, it becomes a suitable digesting medium. Similarly, 10% glacial acetic acid (CH_3COOH), when buffered with sodium acetate or calcium acetate to a pH of 3.5, can be used quite effectively. The technique is similar with both acids and involves immersing the crushed rock into the acid in a fume cupboard and waiting for the effervescence to cease. Formic acid may require up to 1 day before the residue is washed with water and sieved, whereas acetic acid will require many regular changes of acid to dissolve the matrix, and this process can continue for several weeks. For other rock types, such as shales, methods similar to those used for calcareous microfossils are generally used. Once the residue has been sieved, conodonts may be further concentrated by either electromagnetic or heavy liquid separation.

Siliceous microfossils Several groups of microfossils have siliceous skeletons, including radiolaria, diatoms, silicoflagellates, and ebridians. These may occur in both clastic and carbonate rock types, and by virtue of their chemistry are resistant to some acids that may be used successfully to extract them. However, this is not always necessary, because diatoms may be found in unconsolidated sediments and can even form the majority of the rock, which is then termed a 'diatomite'. Radiolarians may also be numerous enough to form radiolarian earth. Various methods are used for loosely consolidated sediments, involving several stages: (1) removal of the carbonate component by using dilute acids, such as hydrochloric, formic, nitric, or acetic acids, (2) removal of organic matter using oxidizing agents, such as hydrogen peroxide or stronger nitric acid, and (3) dispersal of the clay component by using water softeners. With more indurated material, the strength of the acid and the times of digestion are increased. For extraction of radiolarians from mudstones, shales, and marls, a combination of 15% hydrogen peroxide and a degreasing agent such as MP 10 is effective. Very often radiolarians occur in chert beds, siliceous shales, siltstones, or mudstones, which contain silica of the same composition as the fossils. In these cases, 10% hydrofluoric acid (HF) has to be used with extreme caution to etch the radiolarians out of the rock.

The observation of siliceous fossils such as radiolaria is made more difficult because the refractive index of the shells is similar to that of the glass on which the specimens are mounted, rendering them invisible. Techniques to overcome this involve selecting a mounting medium with a high refractive index to contrast with that of the glass, and observing in transmitted light with dark-field illumination, or by coating the specimens with silver nitrate, thus allowing oblique incident light to be used. Larger specimens may be observed directly using a stereozoom microscope. The electron microscope is also used for studies requiring observation of the finest detail and is invaluable in taxonomic work.

Organic microfossils As a group, organic microfossils include pollen and spores, dinoflagellates, acritarchs (*see* **Microfossils:** Acritarchs), tasmanitids, Chitinozoa (*see* **Microfossils:** Chitinozoa), and scolecodonts. With any palynological methods, a clean working environment is essential to reduce the risks of sample contamination, because the specimens are only micrometres in size and are easily transferred between tools and vessels. Although methods vary widely according to the fossil group and the lithology being treated, a broad working plan would follow several stages, including cleaning, disaggregation and dispersal, chemical extraction, and concentration. Palynological preparations differ from other techniques in the chemical extraction stage, because the fossils are composed of sporopollenin or pseudochitin, which allows the use of strong acids; these methods should only be carried out in properly equipped laboratories. Initial disaggregation is achieved by cleaning, crushing, and soaking or boiling in strong detergents such as 'Quaternary O'. Other dispersal agents are used depending on the lithology, such as KOH, NaOH, or Na_3PO for peat; KOH for coals; or NaOH for shales.

Chemical extraction takes several stages to complete and may consist of demineralization and maceration, depending on the rock type. Demineralization is needed to remove the inorganic component of the rock, and different acids are used to achieve this. Carbonates are removed by treatment with HCl; silicates are removed using HF. If sulphates or sulphides are present, they are removed by treatment with HNO_3. Other phases, such as heavy minerals, are removed using heavy liquids. The second stage involves the removal of unwanted organic material; plant disintegration products in the residue may consist of cellulose, hemicellulose, and lignin. Cellulose is broken down by the process of acetylation using a reagent that is a mixture of concentrated sulphuric acid (H_2SO_4) and glacial acetic acid. The broken down cellulose is removed by centrifuging and washing in glacial acetic acid, and finally washing in distilled water. Acetylation also has the beneficial effect of darkening spores and dinoflagellate cysts, which

enhances their appearance for microscopical study. The lignin found in peats, though resistant to acetylation, is susceptible to oxidation and becomes soluble in a hot alkali solution. However, pollen and spores are also easily destroyed by the process of oxidation, so the effect on these grains has to be monitored and the process stopped before damage occurs. Many methods of oxidation may be used, depending on the lithology of the specimen. For example, silicates and argillaceous rocks will respond to 10% nitric acid to which is added 10% sodium hydroxide or Schulze's solution (a mixture of potassium chlorate and concentrated nitric acid); carbonates require fuming nitric acid (95%); peats can be treated with glacial acetic acid, sodium chlorate, and sulphuric acid; and oxidation of coals can be achieved by using Schulze's solution, fuming nitric acid, or hydrogen peroxide.

After these various chemical techniques have been completed, the final stages involve washing by centrifuge. The resulting palynomorphs in the residue are finally washed in 95% alcohol before staining with compounds such as Safranine-O, Malachite green, or Bismarck brown in 0.1% solutions, and are then transferred to a storage phial in distilled water with a mould inhibitor. To prepare slides for study, a small amount of the residue is pipetted onto a cover slip, and after the water has evaporated, a mounting medium, such as Elvacite 2044, Canada Balsam, or epoxy resin, is applied to the glass slide and the cover slip is placed on the slide and allowed to cure.

The usual method of observation is with a conventional biological microscope. However, modern methods of illustration and study involve the use of confocal laser scanning microscopy (CLSM), which gives high-resolution, blur-free images with a greater depth of focus than is possible with conventional microscopy. These images may also be reconstructed to give three-dimensional information and can be viewed as rotating computer animations.

Physical Separation and Concentration Methods

Additional techniques are sometimes required after the initial processing has been completed and the sample is thoroughly disaggregated. The purpose is to concentrate the sample to increase the ratio of microfossil specimens to rock particles such that time spent examining the residue is kept to a minimum.

Sieving

The principal aim of sieving the sample is to remove particles of finer or larger sizes, compared to the fossils of interest. Usually the finer particles are the deflocculated and dispersed clays; fragments larger than the microfossils may be stones, shelly material, or rock that has not broken down completely. The choice of sizes of sieves is often dependent on the fossil group being studied. Usual practice for the larger calcareous microfossils is to use a nest of sieves, of sizes ranging from (bottom to top) finest (75 μm), to medium (250 μm), to coarse (1 mm). If studying planktonic foraminifera, it is essential that the finest sieve is either 75 or 63 μm, because heterohelicids and similar-sized foraminifera will easily fall through coarser sizes. It is worth noting that sieves are designed to separate sedimentary particles that are equant in shape; however, many microfossils are elongated in one direction and may pass through a sieve lengthways. Therefore, the smallest microfossil dimension should be considered when choosing sieves to retain a certain element of the fauna. Occasionally microfossils can become stuck in the mesh, and are then liable to be transferred to another sample as a contaminant. In order to identify such errant specimens, the sieve is stained with a solution of methylene blue; once the surplus stain is washed off, any contaminating material takes the blue colour and is readily spotted in the residue.

Magnetic Separation

Rock particles in residues often contain a proportion of iron, and calcareous or phosphatic microfossils do not, thus it is possible to use differences in magnetic properties to separate the two fractions. For example, this technique is particularly successful in concentrating calcareous microfossils in a residue containing glauconite. This is achieved by using a Frantz Isodynamic Magnetic Separator, which consists of a very large electromagnet that can be tilted in two planes, a vessel into which the sample is placed, and a vibrating motor that moves the particles down a chute. The chute divides into two sections so that separate collecting buckets can receive the magnetic and non-magnetic grains.

Flotation

A common method for concentrating specimens in a residue is to use differences in the specific gravities of the fossil and the matrix. A suitable heavy liquid can then be used to separate the two components, such that one floats and one sinks. Each fraction can then be collected separately. Toxic liquids such as carbon tetrachloride, bromoform, and tetrabromoethane, which have been used widely in the past, have now been superseded by safer compounds, such as sodium

polytungstate, the density of which is altered by the addition of water. To separate conodonts, for example, a liquid with a specific gravity of about 2.75 is used; this allows calcite to float and the conodonts to sink. The specific gravity of pollen ranges from just over 1 to about 1.5, whereas the minerals encountered in pollen analyses have specific gravities ranging from 2.0 to over 3.0 (gypsum, 2.3; feldspar, 2.55; quartz, 2.6; calcite, 2.7; dolomite, 2.8). Therefore, a solution specific gravity between 1.56 and 2.0 should be used for pollen separation.

Picking

The final stage in any processing of the groups of large microfossil is that of picking the specimens. This is usually carried out by thinly spreading out the residue on a flat plate such that each grain can be easily seen, and then using a low-power binocular microscope at 10–20× magnification, scanning each grain to identify the microfossils. When a specimen is spotted, it is removed by using a fine, sable-hair brush of gauge 000 that has been wetted in distilled water. The specimen is transferred onto a numbered gridded slide that has been prepared with some adhesive. Gum tragacanth with a mould inhibitor has been widely used, but water-soluble glues such as Pritt Stick work just as well. Alignment and careful arrangement of the specimens on the slide aids in identification and counting.

See Also

Conservation of Geological Specimens. **Fossil Plants:** Calcareous Algae. **Microfossils:** Acritarchs; Chitinozoa; Conodonts; Foraminifera; Ostracoda; Palynology. **Sedimentary Environments:** Depositional Systems and Facies.

Further Reading

Aldridge RJ (1990) Extraction of microfossils. In: Briggs DEG and Crowther PR (eds.) *Palaeobiology – A Synthesis*, pp. 502–504. Oxford: Blackwell Scientific Publications.

Allman M and Lawrence DF (1972) *Geological Laboratory Techniques*. London: Blandford.

Austin RL (1987) *Conodonts: Investigative Techniques and Applications*. Chichester: Ellis Horwood Limited.

Brasier MD (1980) *Microfossils*. London: Allen and Unwin.

Feldman RM, Chapman RE, and Hannibal JT (1989) *Paleotechniques. The Paleontological Society Special Publication 4*. New Haven, CT: The Paleontological Society.

Green OR (2001) *A Manual of Practical Laboratory and Field Techniques in Palaeobiology*. Dordrecht: Kluwer Academic Publishers.

Hodgkinson RL (1991) Microfossil processing: a damage report. *Micropaleontology* 37(no. 3): 320–326.

Kummel B and Raup D (1965) *Handbook of Palaeontological Techniques*. San Francisco: WH Freeman and Co.

Taylor RJ and Hamilton GB (1982) Techniques. In: Lord AR (ed.) *A Stratigraphical Index of Calcareous Nannofossils*, pp. 11–15. Chichester: Ellis Horwood Limited.

MILANKOVITCH CYCLES

See **EARTH: Orbital Variation (Including Milankovitch Cycles)**

MILITARY GEOLOGY

E P F Rose, Royal Holloway, University of London, Egham, UK

Introduction

Throughout the nineteenth century, geology was widely perceived as a science with potential military applications. However, geologists were not operationally deployed upon the battlefield until the two twentieth-century world wars. Then, German forces made far greater use of military geologists than their British and American opponents, although the roles on both sides were similar. Military geologists guided the quarrying of aggregates and other mineral resources for tactical or strategic use; the development of secure water supplies adequate to support troop concentrations; aspects of military engineering, particularly with regard to fortification and the construction of underground facilities; and terrain analysis, notably

to facilitate or impede amphibious assault and cross-country military 'going'. By the dawn of the twenty-first century, small numbers of military geologists were established for routine use by armed forces, both European (notably British and German) and American.

Historical Background

Military applications of geology first became apparent in Napoleonic times. Indeed, the first general to take geologists as such on campaign was Bonaparte himself. The French army he led into Egypt in 1798 was accompanied by a civilian Commission of Sciences and Arts that included Déodat de Dolomieu (after whom the mineral dolomite was later named) (Figure 1) and several of Dolomieu's former geology students, recent graduates of the School of Mines in Paris – notably Louis Cordier (later immortalized by the mineral name cordierite), François-Michel de Rozière, and Victor Dupuy. However, their role was that of 'mineralogists' to support the army by exploring the geological resources of the country, rather than by contributing tactical or strategic advice. Karl von Raumer (1783–1865), from 1811 Professor of 'Mineralogy' at the University of Breslau (now Wroclaw in Poland), served as a staff officer in Prussia's 1813–14 war of independence from Napoleonic domination, but was used to communicate dispatches rather than appraise geology. The distinguished Swiss mining geologist, Johann Samuel Gruner (1766–1824) (also known as von Grouner), having moved to Bavaria in 1803, became a captain in a volunteer rifle battalion when Bavaria joined the alliance against Napoleon late in 1813. Postwar in 1820, seemingly at the instigation of the Bavarian Bureau of Military Topography, he combined his military and geological experience to write a memorandum on the relationship between geology and military science. Published posthumously in 1826, this is the earliest known work in its field.

Figure 1 Déodat de Dolomieu (1750–1801), in 1798, by André Dutertre. A former cavalry officer and Knight of Malta, Dolomieu taught geology at the School of Mines, Paris, before serving as a senior scientist with Napoleon's expeditionary forces in Egypt. From portrait supplement to Saintine, X-B, Marcel JJ and Reybaud L (eds.) (1830–1836) *Histoire Scientifique et Militaire de l'Expédition Française en Égypte*. Paris: Dénain & Delamare. By permission of the British Library, London (Shelfmark1311.h.2). Also from Rose EPF (2004) Napoleon Bonaparte's Egyptian campaign of 1798: the first military operation assisted by geologists? *Geology Today* 20: 24–29, by permission of Blackwell Publishing Ltd.

In the UK, geological mapping was initially perceived as a military skill. From 1809 to 1814, military (Board of Ordnance) objectives and funding generated geological fieldwork by J. MacCulloch, and from 1814 to 1826 his geological mapping in Scotland. From 1826 to 1846, Royal Engineer officers (successively captains J.W. Pringle, J.E. Portlock, and H. James) pioneered government geological surveys in Ireland. The 'British' Geological Survey was founded in 1835 and sustained until 1845 under military (Board of Ordnance) auspices, and its earliest directors (H.T. De la Beche until 1855, R.I. Murchison until 1871 (*see* **Famous Geologists:** Murchison)) were both men who had received a military rather than a university education. The world's oldest geological society, the Geological Society of London founded in 1807, included amongst its earliest influential members those also active in the reserve army (Lieutenant G.B. Greenough) or the regular army (T.F. Colby, J.W. Pringle, and J.E. Portlock), and a veteran of the Napoleonic wars in receipt of military half-pay throughout his geological career (Lieutenant W. Lonsdale).

Because of its evident practical applications, geology was soon introduced into the curriculum at many military colleges. It was taught at Addiscombe in Surrey to officer cadets of the East India Company's army, by J. MacCulloch from 1819 to 1835 and by D.T. Ansted from 1845 until college closure in 1861. For the British army, J. Tennant lectured on geology to Royal Engineer and Royal Artillery cadets at the Royal Military Academy, Woolwich, from 1848 to 1868; T. Rupert Jones to cadets of the infantry and cavalry at the Royal Military College, Sandhurst, from 1858 to 1870; Jones also to young officers of all arms at the Staff College, Camberley,

until 1882; and A.H. Green at the School of Military Engineering, Chatham, from about 1888 to 1896. Jones (Figure 2) was employed full time as a military professor of geology, but the others were essentially visiting professors who concurrently held better known appointments elsewhere – Ansted and Tennant for a while at King's College London, Green at Oxford. The United States Military Academy at West Point began some geological teaching in about 1823. Publications by R.B. Smith in 1849, F.W. Hutton in 1862, and articles within a massive three-volume, *Aide-Mémoire to the Military Sciences*, by G.G. Lewis and others (first edition, 1846–52; second edition, 1853–62), published primarily for use by the Royal and the East India Company's Engineers, demonstrated quite specifically the importance of geology to the British military profession; however, military teaching of the subject declined sharply towards the end of the nineteenth century as perception of its practical value waned.

Professional geologists were not used operationally as such until World War I, and then primarily as a response to the near-static battlefield conditions on the Western Front in Belgium and northern France. The German army deployed the engineering geologist W. Kranz in 1914, and H. Philipp plus many more geologists in 1915. A German military geological service was constituted in 1916, and made use of some 250 geologists by the end of the war in 1918 when it was disbanded. In the British army, far fewer geologists were used: Lieutenant (later Captain) W.B.R. King from 1915 to guide the development of water supplies; Major (later Lieutenant-Colonel) T.W. Edgeworth David from 1916 to guide siting of mine tunnels and dugouts; and a few others, primarily in the Tunnelling Companies of the Engineer Corps. All returned to their former civilian life at the end of the war: King to the 'British' Geological Survey and then the academic staff of the University of Cambridge and later London, and David as Professor of Geology and Physical Geography to the University of Sydney in Australia. Led by Lieutenant-Colonel A.H. Brooks, nine (of potentially 18) geologists were assigned for service with the American Expeditionary Force in 1918, primarily from the United States Geological Survey, but the war ended before their full operational deployment.

The German army was quick to reinstate a military geological organization prior to World War II, from 1937. Military geology textbooks were generated by J. Wilser in 1921, E. Wasmund in 1937, K. von Bülow and others, C. Mordziol, and W. Kranz, all in

Figure 2 Professors of the Staff College, Camberley, in 1874. The geologist, T. Rupert Jones, is seated at the left rear. Reproduced with permission from Rose EPF, Häusler H, and Willig D (2000). In: Rose EPF and Nathanail CP (eds.) Comparison of British and German applications of geology in world war, pp. 107–140. *Geology and Warfare: Examples of the Influence of Terrain and Geologists on Military Operations*. London: Geological Society. Courtesy of the Joint Services Command and Staff College, Swindon, and the Geological Society, London.

1938. Military geologists were deployed with troops invading Poland in 1939, and France and the Low Countries in 1940. By late 1941, there were 32 teams of military geologists providing support to the German army throughout its area of occupation, the number of terms increasing to 40 by November 1943. Making use eventually of some 400 geologists, largely recruited or conscripted from university staffs, the German army thus developed the largest organization ever to provide military applications of geoscience in wartime. Additionally, smaller but still significant numbers of geologists served with the German air force, navy, Waffen-SS, or the paramilitary construction agency 'Organisation Todt'.

In contrast, the British army again used very few geologists as such: Major (later Lieutenant-Colonel) W.B.R. King from 1939 to 1943 and his Cambridge protégé Captain (later Major) F.W. Shotton from 1940 to 1945; this group was increased to perhaps a dozen in various roles by 1944, largely in preparation for the Allied invasion of Normandy, and many of them formerly academic geologists from British universities. America entered the war in December 1941, its forces soon supported by a Military Geology Unit at the United States Geological Survey. With a wartime roster of 88 geologists, 11 soil scientists, 15 other specialists, and 43 support staff, it produced 313 studies – including 140 major terrain folios, 42 other major reports, and 131 minor investigations – in total containing about 5000 maps, 4000 photographs and figures, 2500 large tables, and 140 terrain diagrams.

After the war, the British army maintained continuity in geological expertise through a small group of officers in the reserve army (the Territorial Army or, from 1953 to 1967, the Army Emergency Reserve) for peace-time engineering projects and operational planning, and active service in times of crisis, whether occasioned by military conflict or humanitarian need. German capability in military geology was abolished at the end of the war, but later re-established within the Bundeswehr. By the end of the Cold War in 1990, the German army employed more than 20 full-time geologists – as civilians but with commitment as reserve army officers. In the USA, geological support for the armed forces continued to be based on the United States Geological Survey which, between 1945 and 1972, briefly used about 150 scientists to compile terrain intelligence on a global scale.

Quarrying, Aggregates, and Mineral Resources

Quarrying became important to the German army in World War I. From 1915 especially, in regions such as Alsace, German military geologists produced maps showing sites for quarrying of construction materials – building stone, sand, and aggregates – maps that were used to guide the construction of fortification systems by the Engineer Corps. Other maps plotted reserves of economic significance, such as coal and metalliferous ores. A significant proportion of German military geologist tasks (Figure 3) were related to the winning of raw materials in occupied land. The Allies, fighting to defend their own terrain, made use of local civilian rather than military geological expertise for such tasks.

Raw materials also formed a significant proportion (Figure 3) of German military geologist tasks in World War II. Fortification systems, such as the West Wall bordering Germany and France, and the Atlantic Wall that marked the western limit of German-occupied Europe, consumed vast quantities of aggregates for concrete in 1938–40 and 1940–44, respectively. The Atlantic Wall alone included some 15 000 bunker and casemate complexes with outer walls of reinforced concrete at least 2 m thick, eventually absorbing about 17.6 million tonnes of concrete.

Knowledge of potential sources of raw materials was also required for operational planning. For Operation Sealion, the German invasion of England scheduled for September 1940, planning included the preparation of maps by military geologists at scales of 1 : 100 000 and 1 : 50 000 to show the positions of all working and disused quarries in south-east England, the nature of material produced (e.g., limestone, sand, and clay), and its potential engineering use. Such maps, their data derived mainly from published topographical maps plus geological maps and memoirs, were widely used by the German army in preparation for attack.

In defence, the Allies relied largely on civilian agencies, such as the Geological Survey of Great Britain, for geological advice. When counter-attacking, however, military geologists were used. To support the Allied liberation of Normandy in June 1944, the (British) Geographical Section of the General Staff reprinted all the French 1 : 80 000-scale geological maps of the region. These, plus military geologist advice, guided the operation of units, such as the Quarry Group, Royal Engineers. During the 11 month campaign in north-west Europe, this generated in total some 1 376 000 m^3 of crushed rock from 49 different quarry sites (Figure 4), successively in France (Figure 5), Belgium, and Germany – aggregates used for road repair, widening, and construction; hard standings for ammunition and stores depots; vehicle and gun parks; and airfield runways. The mobile, mechanized conflict of World War II gave the maintenance and construction of roads and airfields a high priority.

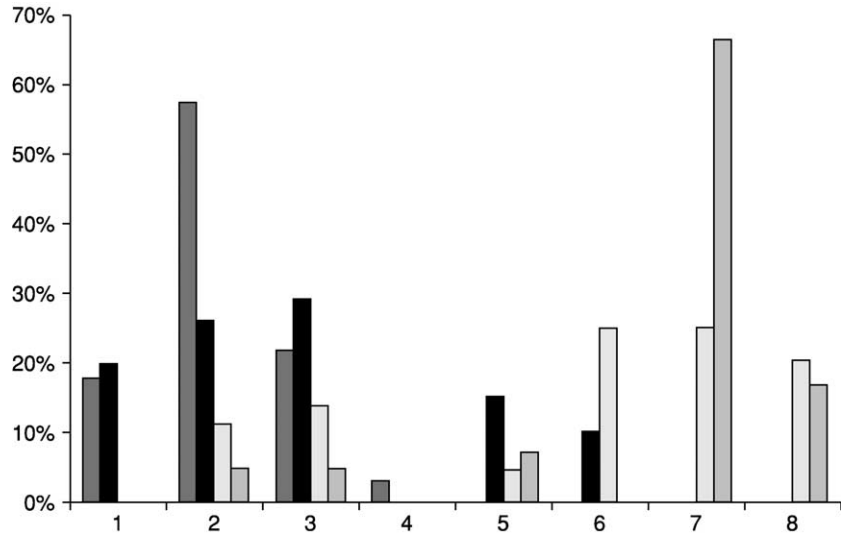


Figure 3 Categories of major tasks undertaken by German military geologists, showing percentage of tasks calculated for each of four time periods to demonstrate a change in focus from World War I to the end of the twentieth century. ▩, World War I (1914–18), based on 215 tasks undertaken by the military geology group directed by C. Mordziol between February and November 1917; ■, World War II (1939–45), based on 640 tasks; □, Cold War (1947–89), based on tasks for III (German) Corps between April 1971 and March 1972; ▩, late twentieth century, based on 42 main tasks undertaken by III (German) Corps in 1990. 1, Raw materials; 2, water supply/drainage; 3, infrastructure; 4, general and regional geology; 5, trafficability/exercises; 6, obstacles/barriers; 7, environmental protection; 8, other tasks. Environmental protection features more highly in post-1970 work for III Corps than for other German corps. Data courtesy of D. Willig, derived from Rose EPF, Häusler H, and Willig D (2000). Comparison of British and German applications of geology in world war, pp. 107–140. In: Rose EPF and Nathanail CP (eds.) *Geology and Warfare: Examples of the Influence of Terrain and Geologists on Military Operations*. London: Geological Society.

In the postwar years, military quarrying has waned considerably in importance (Figure 3), but the expertise has been maintained for occasional operational use. Thus, following the Falkland Islands' conflict between the UK and Argentina in 1982, aggregate for repair and extension of the Islands' principal airport was provided from quarries operated by Royal Engineers, supported by advice from a military geologist.

Water Supply

Potable water has long been recognized as a vital military resource. Without a secure water supply, no defended locality, however well fortified and garrisoned, can withstand siege for long. Moreover, troops massed for campaign need, in total, copious supplies of drinking water.

Problems of water supply and drainage related to the army's field positions were, from late 1914, amongst the earliest German military geologist tasks assigned on the Western Front in World War I. Overall, they represented by far the greatest number of tasks (Figure 3). From 1915, German military geologists were preparing a variety of water supply maps to provide information in a manner readily understood by combat troops, especially to guide the siting of new boreholes and dams for ponding small reservoirs.

Also in 1915, the British army deployed its first operational military geologist, 'Bill' King, as a hydrogeologist. Most of his early work consisted of compiling water supply maps. Later work involved the selection of sites for boreholes. Over 470 borings for water were emplaced behind the Western Front by the British army during the war, and King supervised and interpreted many of these. Improved techniques for recording and mapping hydrogeological data were developed as a military expedient at this time.

In World War II, geological appraisals relating to water supply were still important to the German army, if less so than in World War I (Figure 3) because conflict was relatively more mobile. Military water supply maps ranged from the very detailed, large-scale maps prepared for some areas under long-term occupation, to the simpler, smaller scale maps required for operational use. Thus, in the 2 month planning period for Operation Sealion, German military geologists gave a high priority to preparing water supply maps at scales of 1 : 50 000, 1 : 100 000, and 1 : 250 000. These depicted the major lithologies which cropped out in south-east England, typically grouped into three categories according to

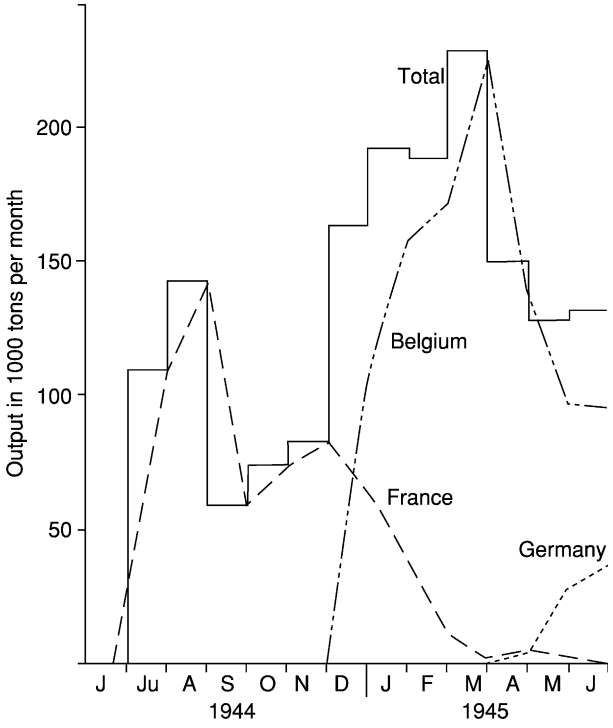


Figure 4 Monthly 1000 ton (1020 tonne) quantities of stone produced or distributed by the Quarry Group, Royal Engineers, from June 1944 to June 1945, in total (full line) and progressively from sources in France, Belgium, and finally Germany. Redrawn, with permission from Blackwell Publishing Ltd., from Rose EPF and Pareyn C (1995) Geology and the liberation of Normandy. *Geology Today* 11: 58–63.

suitability for the emplacement of shallow wells: suitable (overprinted with crossed diagonal lines), unsuitable (no overprint), and only partly suitable (with unidirectional diagonal line overprint) (Figure 6).

In the British army, 'Bill' King was deployed once more to France as a hydrogeologist to serve with the British Expeditionary Force, from 1939 until its evacuation in 1940. Dowsers rather than geologists were initially used to guide well drilling in North Africa and the Middle East, but with little success. There was a marked improvement in 1941 when 'Fred' Shotton was deployed as a military hydrogeologist, with support from 42nd Geological Section of the South African Engineer Corps, whose members were experienced in the use of electrical resistivity and other geophysical techniques employed in the exploration for groundwater in arid zones. From 1943, first King and then Shotton were tasked with geological preparations for the liberation of Normandy, which included the complication of water supply maps, amongst others (principally for beach landings, and airfield construction). Following the D-Day landings, well drilling in Normandy to develop groundwater, largely by borings through Middle Jurassic limestones, was carried out primarily under military geological supervision.

Postwar, the British army has retained a well drilling capability on a near-permanent basis. Part-time reserve army geologists have been used to guide its operation

Figure 5 Military quarrying of Middle Jurassic limestone near Caen, Normandy, August 1944. © The Imperial War Museum, London: photograph CL811.

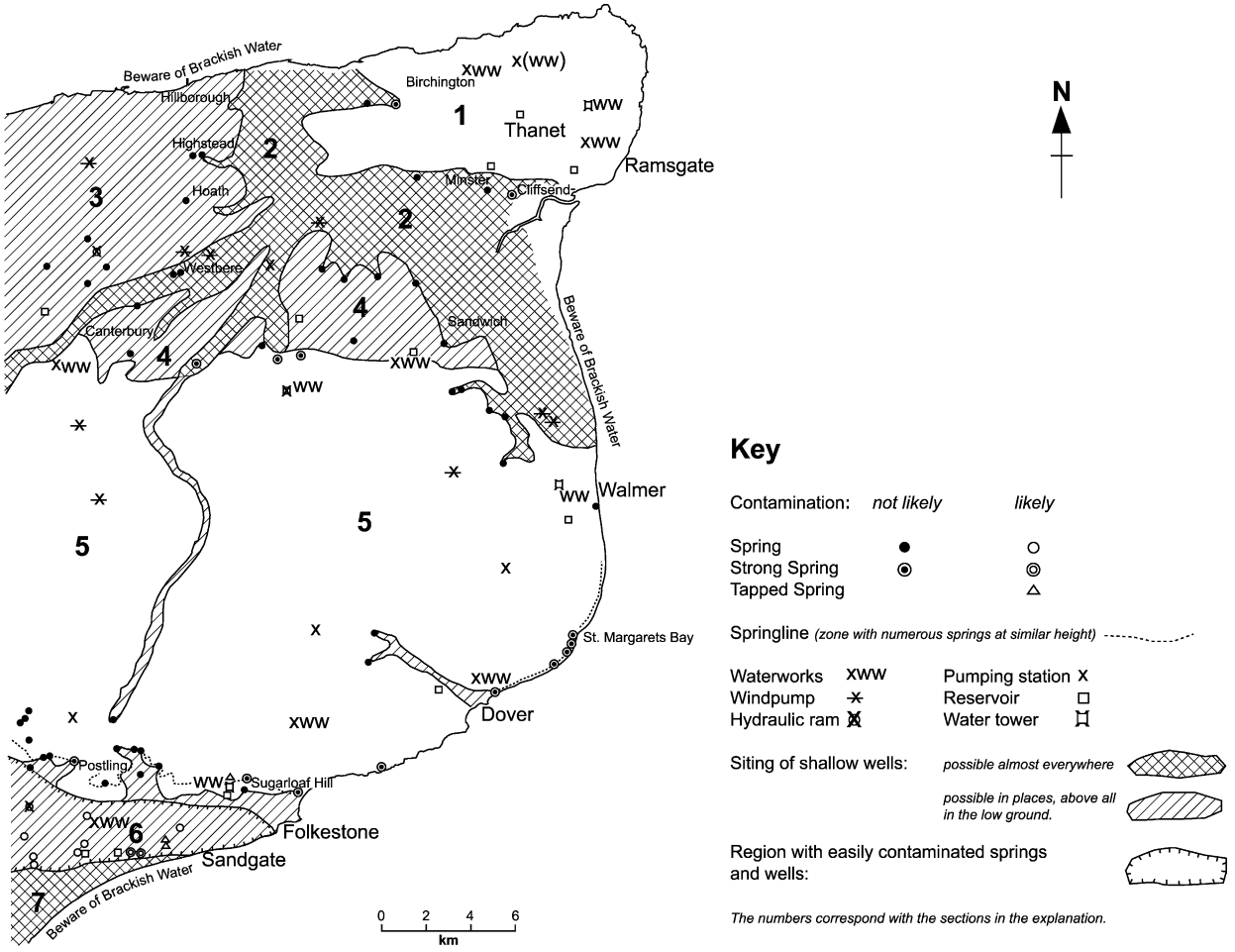


Figure 6 Map of east Kent, showing the seven water supply regions depicted on the 1 : 50 000-scale water supply map sheet 117 prepared by German military geologists to facilitate planning for the invasion of England in September 1940. Key translated from German, but retaining the symbols adopted for the original map. Redrawn with permission from Rose EPF, Mather JD, and Willig D (2002) German hydrogeological maps prepared for Operation 'Sealion': the proposed invasion of England in 1940. *Proceedings of the Geologists' Association* 113: 363–379, Fig. 4. © 2002 The Geologists' Association, London.

both in the UK and overseas (Figure 7). The German army has also used its military geologists in a water supply role (Figure 3), notably under United Nations or North Atlantic Treaty Organization auspices, in countries such as Somalia, Bosnia-Herzegovina, Kosovo, and Afghanistan.

Engineering Geology

Although tunnelling to emplace mines is one of the oldest applications of engineering to the science of war, during World War I it was employed on a vast and unprecedented scale on the Western Front. Beginning in 1915, it reached a peak in 1916 with some 25 000 Allied troops actively engaged in tunnelling, faced by a comparable number of Germans, and nearly 1500 mines fired on both sides in the process of more or less continuous underground warfare. Through test boring and collation of geological data, including the compilation of some of the first water table maps to be prepared in Europe, Edgeworth David provided detailed information on the extent and depth of strata most suitable for mining by the British Expeditionary Force (Figure 8).

Later, as from 1917 the front line was held by firepower rather than manpower, troops were deployed in depth and protected from artillery and aerial bombardment by dugouts (underground shelters), subways (connecting tunnels) and other subsurface excavations. To guide their siting, David prepared specialist geotechnical maps (Figure 9), illustrating the relative suitability of ground for dugout construction – credited to be amongst the first environmental/engineering geology maps ever published. German military geologists were used on similar 'infrastructure' tasks (Figure 3) related to the construction of field positions,

Figure 7 Royal Engineers boring for water in the South Arabian Federation in 1964. Military geologists from the British reserve army were used to guide such deployments during a time of intermittent conflict. © Crown Copyright/MOD. Reproduced with the permission of the Controller of Her Majesty's Stationery Office, and courtesy of the Royal Engineers Library, Chatham.

especially trenches, mines, and dugouts (to minimize time, labour, and materials, and the risk of collapse or water inflow), and site selection for standing camps, munition dumps, heavy gun positions, and airfields.

On demobilization, there were many more German than Allied geologists who returned to civilian life with practical experience in engineering geology. From them developed the expertise more fully deployed in World War II (Figure 3). The German army and navy were noted for using geological guidance when planning and constructing major fieldworks, whether fortifications, underground installations and factories, airfields, military bases, or command HQ facilities. In addition to the massive fortifications of the West Wall and Atlantic Wall, examples include the successful construction of bomb-proof submarine pens on the coast of Norway at sites with considerable geotechnical problems, such as the instability of beach sands and weak 'quick' clays, and active residual stresses and associated rebound-relief structures in the rock formed as a consequence of the region's glacial and tectonic history. Moreover, the Germans successfully located many military installations and critical manufacturing plants underground throughout Europe during the war.

The British seldom used military geologists for such tasks, except on Gibraltar, where a military geologist was briefly available during 1943 to guide some of the extensive tunnelling work undertaken to enhance the rocky peninsula as a fortress.

When America entered the war, its military engineers had to operate in terrain which varied from the

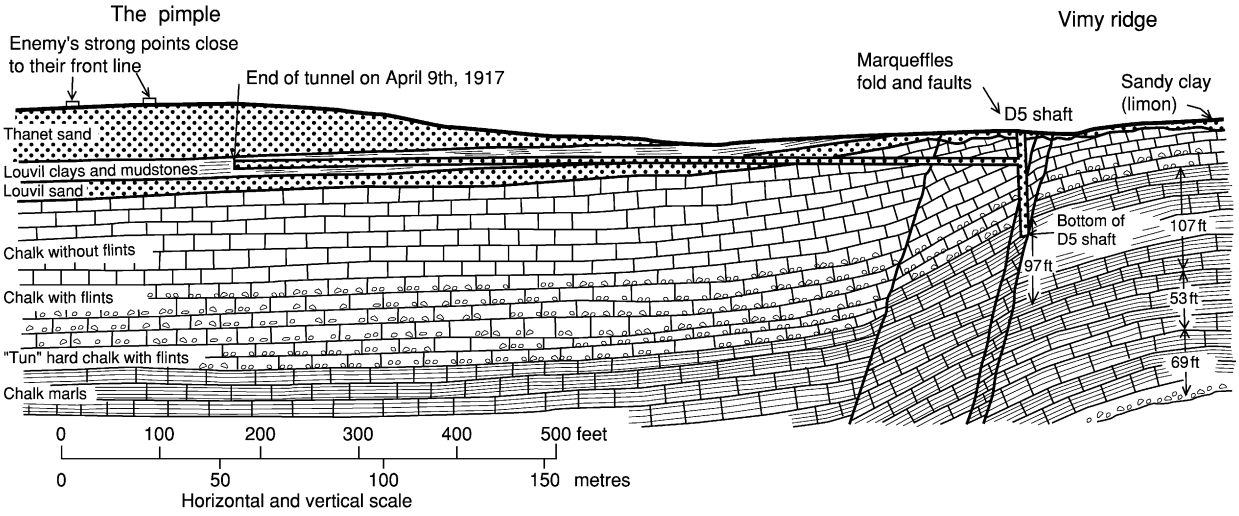


Figure 8 Geological cross-section east–west through Vimy Ridge, France, showing a mine tunnel excavated forward from the British lines, mostly in Louvil clays, which provided the quiet, dry, and stable conditions ideal for tunnelling. Redrawn, with permission from Blackwell Publishing Ltd., from Rosenbaum MS and Rose EPF (1992) Geology and military tunnels. *Geology Today* 8: 92–98.

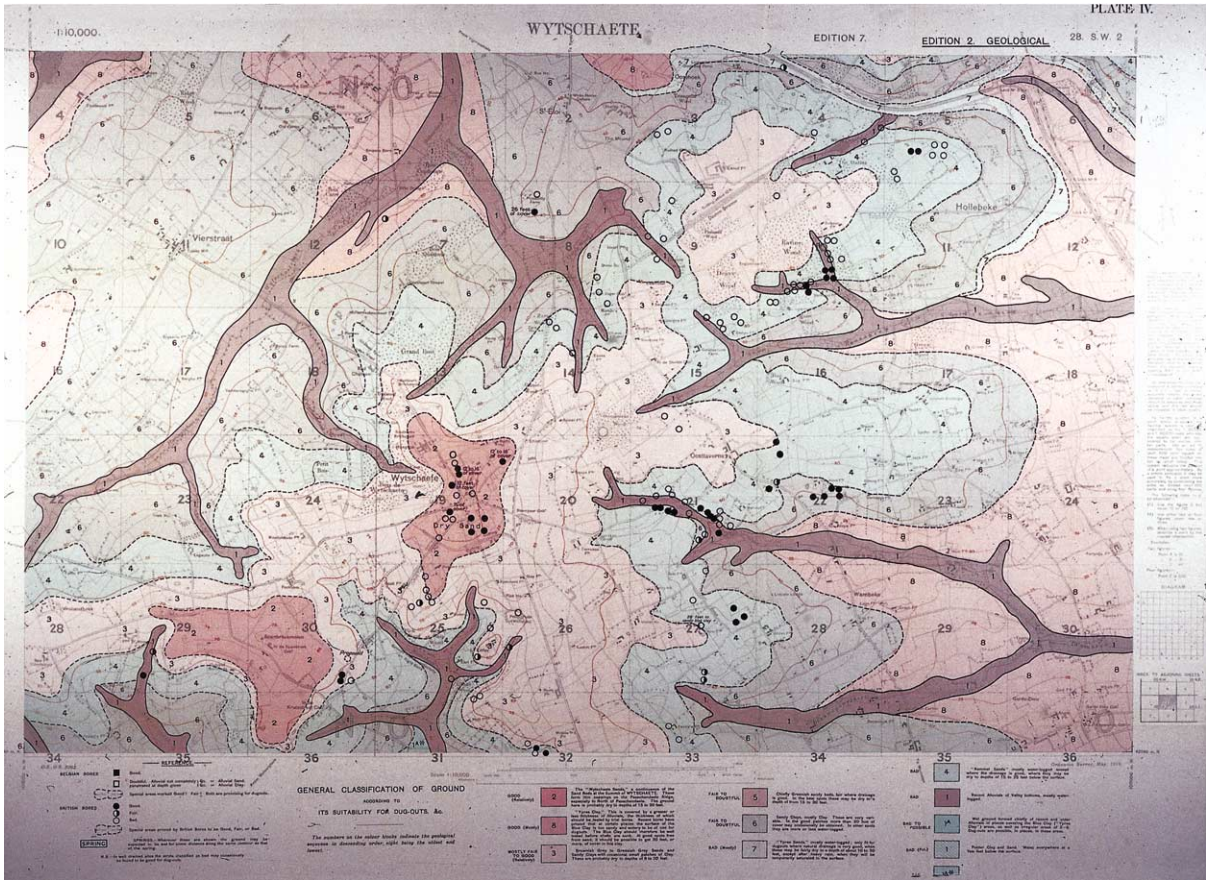


Figure 9 'Geological' map of Wytschaete, Belgium, originally at a scale of 1 : 10 000, showing the relative suitability of the ground for the construction of dugouts. The redder the formation shown on the map, the drier and more suitable for dugout construction; the bluer the formation, the wetter and less suitable. Reproduced with permission from the Ordnance Survey map of May 1918.

frozen Arctic to the jungles of the tropics. In northern regions, permafrost was a problem. In the island region of the South Pacific, terraced coral reefs and associated soft to weathered clay-mud and sand provided a very different challenge, because of physical variability and rapid variations in rock strength. Postwar, military engineering geology continued to be important to America as its role as the global superpower developed; however, its importance in European priorities declined sharply (Figure 3) as during peacetime major engineering works for military clients were largely undertaken by civilian agencies or contractors rather than by uniformed military engineers.

Terrain Analysis

In the nineteenth century, German language authors showed early appreciation of the relationships between geology, terrain, and warfare, but terrain analysis for military operations proved to be of relatively minor importance in World War I, at least in the near-static battlefield conditions of the Western Front. Studies of regional geology formed a very minor proportion of German military geological tasks (Figure 3), and the same was true of tasks carried out by the much smaller number of Allied military geologists.

In the more mobile conflict of World War II, the situation was different. A significant proportion of tasks undertaken by German military geologists related to the effect of terrain on the off-road trafficability of ground for military vehicles of different types, tracked or wheeled, light or heavy, and on the avoidance in attack or enhancement in defence of natural obstacles to troop movements (Figure 3). For the invasion of south-east England scheduled for September 1940, German military geologists prepared coastal maps at a scale of 1 : 50 000, which categorized beaches and cliffs in terms of suitability for amphibious assault, based on features related to both geomorphology and geology. A start was made on construction/engineering maps of inland areas, showing the distribution of major rock types, with assessment of the workability (by hand tools

to explosives), stability of excavations (with or without gabions), and permeability to groundwater (Figure 10). Another series was started to show variations in off-road trafficability, i.e., maps from which the rate of military 'going' could be predicted.

Later in the war, the German army developed map-making roles for all major combat areas, and the German navy developed specialist geotechnical maps for coastal regions to facilitate defence from Allied assault. Around 1942, a marine geographical unit was established by the Naval High Command with the primary objective of preparing specialist maps of selected coastal areas to facilitate defensive planning and fortification. The unit included a nucleus of professional geographers and at least one geologist. Also in 1942, a research detachment for special duties was created to provide terrain mapping and evaluation for the Armed Forces High Command. Initially, this was tasked with investigating the passability of the central Libyan Desert for troops moving from the south. Later, the scope of its work widened, and specialist maps were generated for many areas, at scales ranging from 1 : 50 000 to 1 : 500 000, by teams which included one or more specialists in geography, geology, plant ecology, meteorology, soil science, cartography, and photogrammetry, working in cooperation. This multidisciplinary work generated some 36 terrain evaluation maps between September 1943 and February 1945, covering areas from northern Finland to the mouth of the River Dnieper, and from north-west Germany to Greece. These were complemented by specialist trafficability maps of the Caucasus, the Pyrenees, and the eastern Alps, generated largely in 1942–43 by the German army's military geographical service. 'Military geology' maps were prepared by teams of German army geologists for some areas, e.g., to guide fortification of the Channel Islands (Figure 11).

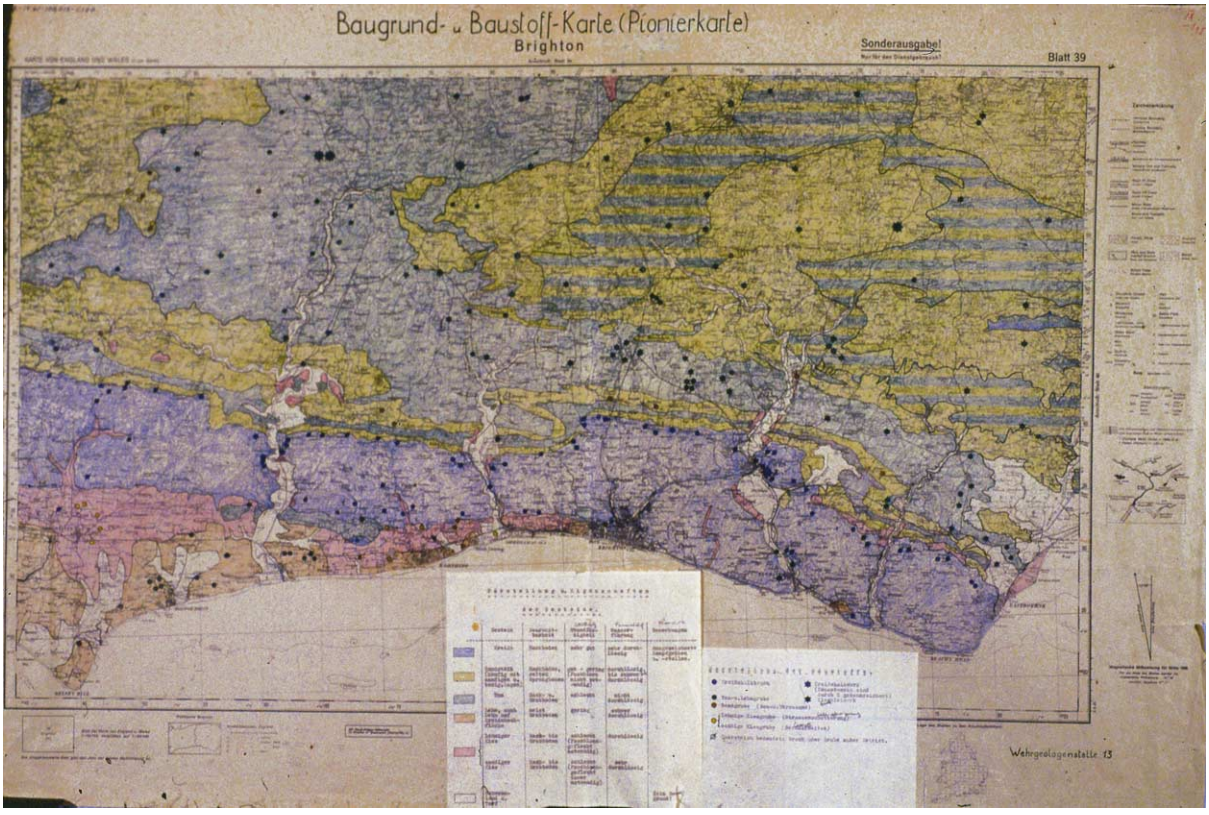


Figure 10 Building terrain and construction materials (engineering) map of the Brighton area of southern England, prepared at a scale of 1 : 100 000 by German military geologists to assist planning for the invasion scheduled for September 1940. Quarry sites are plotted on to a simplified geological map, and the key classifies the major rock types in terms of engineering properties. Reproduced with permission from Baugrund-und Baustoff-Karte, Brighton [Cartographic Record]; #23-2; German Language Geologic Maps, Record of the Geological Survey, Record Group 57; National Archives at College Park, College Park, MD, USA. Also by permission from Rose EPF and Willig D (2004). Specialist maps prepared by German military geologists for Operation Sealion: the invasion of England scheduled for September 1940. *The Cartographic Journal* 41: 13–35.

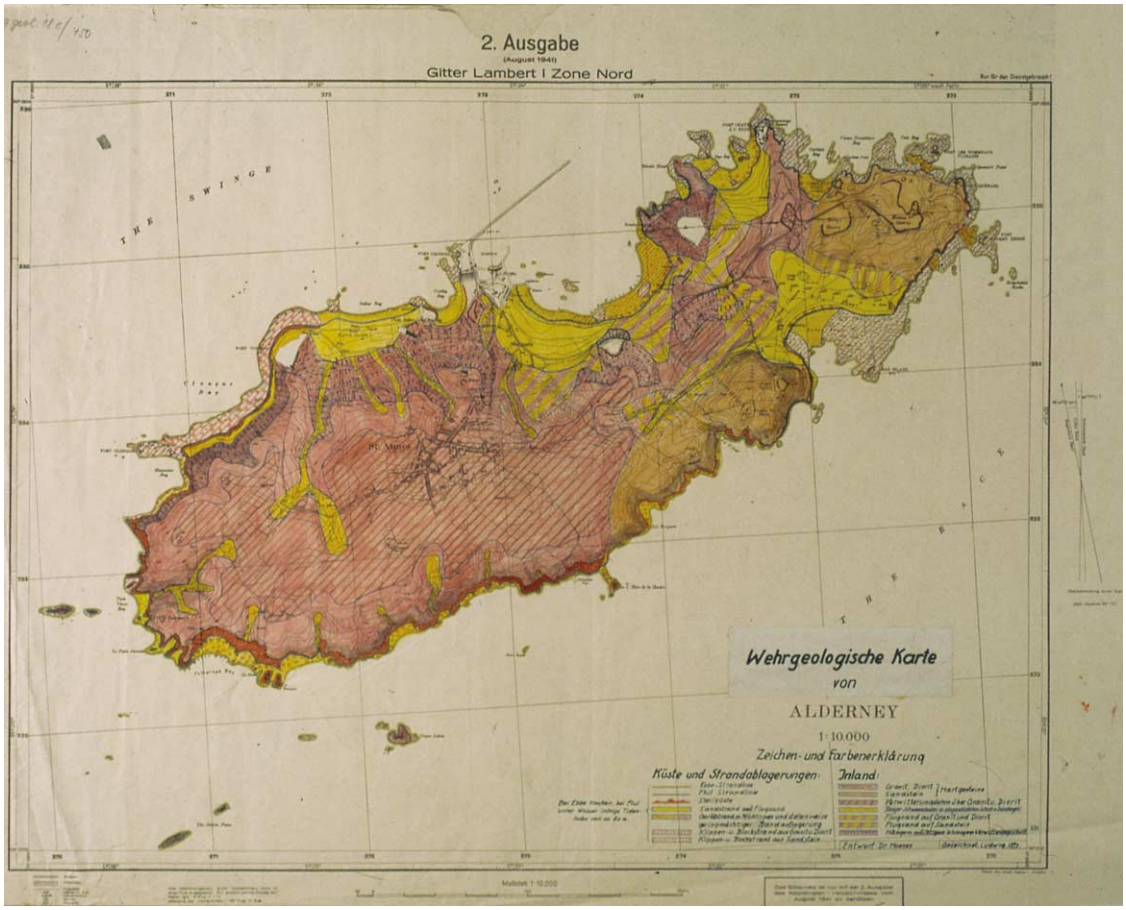


Figure 11 Military geology map of Alderney, one of the British Channel Islands occupied by German troops during World War II, compiled by German military geologists in 1942. The key provides an appraisal of rock type and geomorphology for both coastal and inland areas in terms that would facilitate fortification and troop deployments against British counterattack. Reproduced from Wehrgeologische Karte von Alderney [Cartographic Record]; #20-Great Britain-Channel Is.; German Language Geologic Maps, Records of the Geological Survey, Record Group 57; National Archives at College Park, College Park, MD, USA.

Allied geologists also helped to generate specialist maps as the war progressed, the British especially to facilitate operations in Normandy in 1944 (Figure 12). Additionally, late in the war, cross-country movement (CCM) maps were developed by the Allies, especially the Americans, by multidisciplinary effort.

The preparation of such specialist maps, including those for areas without ground access, became a focus of interdisciplinary military activity during the Cold War of 1947–89. During this time, the Military Geology Branch of the United States Geological Survey focused research on strategic studies, areas of the Pacific region formerly occupied by the Japanese, Alaska and other areas of permafrost, military engineering geology and CCM maps for Germany, suitability of arid lands both inside and outside the USA for airfield construction, special intelligence, and studies to assist interpretation of global seismic signals for the Nuclear-Test Detection Program.

Conclusion

During the nineteenth century, there was a closer relationship between the military and some of the founding fathers of the science of geology than is commonly realized. In Britain, in particular, military objectives and funding generated initiatives both in the teaching of geology and in government-sponsored geological mapping. However, such initiatives waned through the second half of the century.

During the twentieth century, rather than teach the basic elements of geology to army or at least engineer officers in general, Britain, Germany, and the USA all incorporated professional geologists into their armed forces. During World Wars I and II, the role of geologists was partly to gather, collate, and interpret geological information for particular military use, and partly to guide site investigation and well drilling. Much military geology was merely

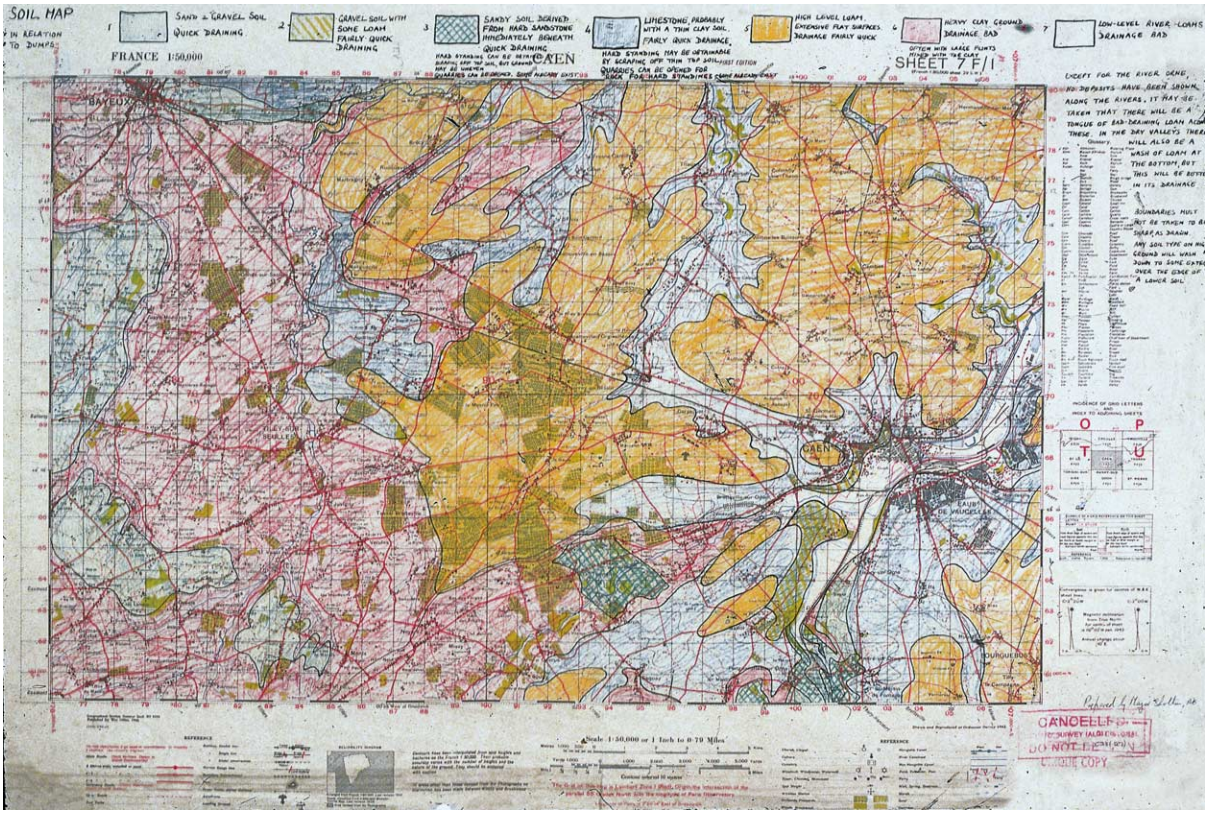


Figure 12 Soil map in relation to dumps, Caen: one of the specialist maps prepared by British military geologists in support of the 1944 campaign in Normandy. This 1 : 50 000-scale topographical map was hand coloured by Major F. W. Shotton to distinguish areas according to seven categories, given (from left to right) in the key at the top: 1, sand and gravel soil, quick draining; 2, gravel soil with some loam, fairly quick draining; 3, sandy soil derived from hard sandstone immediately beneath, quick draining; hard standing can be obtained by scraping off top soil but ground may be uneven; quarries can be opened, some already exist; 4, limestone, probably with a thin clay soil, fairly quick drainage; hard standing may be obtained by scraping off thin top soil; quarries can be opened for rock for hard standings, some already exist; 5, high-level loam, extensive flat surfaces, drainage fairly quick; 6, heavy clay ground, drainage bad; often with large flints mixed with the clay; 7, low-level river-loams, drainage bad; peat-marsh along coast at Meuvaines. Donated by the Ministry of Defence to British Geological Survey. © NERC. See also Rose EPF and Pareyn C (2003) *Geology of the D-Day Landings in Normandy, 1944, Geologists' Association Guide 64*. London: Geologists' Association.

hydrogeology or engineering geology constrained by military objectives and urgency. Additionally, the need to communicate complex data to non-geologists under operational conditions led to the development of a wide range of specialist engineering geological maps to depict ground conditions in order to guide specific military activities. Examples include: water supply maps to guide the siting of boreholes and the abstraction of potable groundwater; geotechnical maps to guide the excavation of mines and dugouts; resource maps to guide the quarrying of materials for tactical or strategic use; 'diggability' maps to guide the siting of trench systems and anti-tank ditches; and CCM (trafficability or 'going') maps to indicate the potential effect of terrain and varying climatic conditions on the off-road movement of troops and vehicles. Some but not all of these map categories were derived from, or subsequently gave rise to, similar non-military specialist maps – for use in planning environmental or civil engineering works. Some depicted very complex data, but the maps most popular with military commanders, irrespective of nationality, were commonly those that depicted ground in terms of only three categories for a specific military activity: 'go' (suitable); 'no go' (unsuitable); and 'slow go' (intermediate or needing further qualification) (e.g., Figures 6 and 9).

By the start of the twenty-first century, in addition to these traditional roles, there was a trend for military geologists in peacetime to be used to protect the environment rather than the state, particularly in Germany (Figure 3) and the USA. In the Western world, large areas of land are used for military

training, and there is a continuing need to monitor and control the use of this land for a number of reasons: to protect soil and aquifers from significant contamination by explosive residues, propellant combustion products, and fuel spillage; to minimize scarring or erosion of soil pavements by off-road use of vehicles and foot traffic; and to assess the effects of vehicular compaction of soils in these areas, with its implications for permeability, rainfall runoff, and consequent patterns of erosion. Remote sensing, coupled with geological and geographical expertise, has facilitated the preparation of a wide range of specialist maps. It has also been used to predict sites of clandestine tunnels and underground facilities; cave systems potentially adapted for terrorist use; and areas in which bombing might artificially induce landslides and so deny road communication to an enemy. Geophysical techniques have been used to supplement geological appraisals in the detection and location of underground facilities potentially of use for espionage or as a means to avoid arms control inspections. Geologists and the military thus continue to share an interest in a variety of ground properties.

See Also

Engineering Geology: Aspects of Earthquakes; Geological Maps; Geophysics. **Environmental Geology**. **Famous Geologists:** Murchison. **Geological Engineering**. **Geological Field Mapping**. **Geological Surveys**. **Geology, The Profession**. **History of Geology From 1780 To 1835**. **History of Geology From 1835 To 1900**. **History of Geology From 1900 To 1962**. **Remote Sensing:** Active Sensors; GIS; Passive Sensors. **Sedimentary Rocks:** Dolomites.

Further Reading

Caldwell DR, Ehlen J, and Harmon RS (eds.) (2004) *Studies in Military Geology and Geography*. Dordrecht: Kluwer Academic Publishers.

Doyle P (1998) *Geology of the Western Front, 1914–1918, Geologists' Association Guide 61*. London: Geologists' Association.

Doyle P and Bennett MR (eds.) (2002) *Fields of Battle: Terrain in Military History*. Dordrecht: Kluwer Academic Publishers.

Ehlen J and Harmon RS (eds.) (2001) *The Environmental Legacy of Military Operations, Reviews in Engineering Geology 14*. Boulder, CO: Geological Society of America.

Häusler H (1995) Die Wehrgeologie im Rahmen der Deutschen Wehrmacht und Kriegswirtschaft. *Informationen des Militärischen Geo-Dienstes, Vienna* 47: 1–155; 48:1–119.

Rose EPF (1996) Geologists and the army in nineteenth century Britain: a scientific and educational symbiosis? *Proceedings of the Geologists' Association* 107: 129–141.

Rose EPF (2004) The contribution of geologists to the development of emergency groundwater supplies by the British army. pp. 159–182. In: Mather JD (ed.) *200 Years of British Hydrogeology, Geological Society Special Publication 225*. London: Geological Society.

Rose EPF and Hughes NF (1993) Sapper geology. *Royal Engineers Journal* 107: 27–33, 173–181, 306–316.

Rose EPF and Nathanail CP (eds.) (2000) *Geology and Warfare: Examples of the Influence of Terrain and Geologists on Military Operations*. London: Geological Society.

Rose EPF and Pareyn C (2003) *Geology of the D-Day Landings in Normandy, 1944, Geologists' Association Guide 64*. London: Geologists' Association.

Rose EPF and Rosenbaum MS (1993) British military geologists. *Proceedings of the Geologists' Association* 104: 41–49, 95–108.

Rose EPF, Mather JD, and Willig D (2002) German hydrogeological maps prepared for Operation 'Sealion': the proposed invasion of England in 1940. *Proceedings of the Geologists' Association* 113: 363–379.

Underwood JR Jr and Guth PL (eds.) (1998) *Military Geology in War and Peace, Reviews in Engineering Geology 13*. Boulder, CO: Geological Society of America.

von Bülow K, Kranz W, and Sonne E (1938) *Wehrgeologie*. Leipzig: Quelle and Meyer.

Willig D (1999–2003) Entwicklung der Wehrgeologie: Aufgabenspektrum und Beispiele. *Fachliche Mitteilungen des Amtes für Wehrgeophysik, Traben-Trarbach* 225: 1–116; 226: 1–82; 227: 1–77.

MINERAL DEPOSITS AND THEIR GENESIS

G R Davis, Imperial College London, London, UK

The Nature of Mineral Deposits

The modern industrialized world makes use of an astonishing range of rocks and minerals, either for direct use or for processing to other useful forms such as metals. Mineral deposits are formed in places where natural processes have concentrated certain minerals, and some of these deposits are rich enough to be exploited as economic deposits. New rocks (including mineral deposits) are generated through weathering and sedimentation, and through volcanic activity and igneous intrusion, and existing rocks (including mineral deposits) are metamorphosed by tectonic activity and changes in their pressure/temperature environment within the Earth's crust. The formation of mineral deposits should be expected to be part of this general scheme of petrogenesis, bodies of rock where valuable minerals have become concentrated by processes active in their crustal environment over long periods of time. Many mineral deposits are simply expressions of peak geological conditions within more widely dispersed and less intense mineralization. For instance, the thin black Kupferschiefer shale of the large central European Permian basin is typically mineralized with copper sulphides. In only a few places, however, does mineralization reach sufficient concentration and extent to be called a deposit, and in fewer still, such as Lubin, is it an economic mineral deposit. Once seen as geological accidents, mineral deposits are now widely regarded as a normal and revealing part of the overall geological picture. This transformation was illustrated in 1972 by the appearance of a major textbook entitled Ore Petrology (Stanton) alongside classical texts on Igneous Petrology and Sedimentary Petrology. The processes by which anomalously high concentrations are achieved in mineral deposits, and particularly metalliferous deposits, are far less well understood than those operating to form anomalously high, but more abundant and familiar concentrations of common minerals such as SiO_2 as quartzite and $CaCO_3$ as limestone or chalk. Fortunately, our understanding of many ore-forming processes is constantly being broadened through detailed observation of both known deposits in existing mines, and new kinds of deposits revealed by mineral exploration. Field observations are supplemented by the wide range of laboratory studies including fluid inclusions, stable isotopes of sulphur and oxygen, and age dating, in addition to petrological methods of microscopy on transparent and opaque minerals. Knowledge about mineral deposits, the geological controls that influence their characteristics, the environments in which they develop and the genetic processes that produce them, is the basic science that is applied by economic geologists in their professional functions as mining geologists and exploration geologists (*see* **Economic Geology**).

The following introduction is a brief and necessarily incomplete outline of some aspects of the great variety of mineral deposits. The scheme of arrangement is based on the association of mineral deposits with the major geological rock groups – igneous, sedimentary and metamorphic. The order of presentation is pragmatic, moving from deposits and their genetic processes that are readily understood, to those that are enigmatic and controversial. We start, therefore, with those deposits and processes that we can observe at the present time in our own familiar surface environments, and progress through associations with basic igneous rocks, felsic igneous rocks, ancient sedimentary rocks, and, finally, metamorphism.

Present Day Associations and Processes

The easiest processes to understand are those we can actually observe in operation at or near the Earth's surface, namely the weathering of rocks, the transportation of the products to form secondary deposits, and the precipitation of minerals from solution. Prolonged weathering of suitable bedrock, by no means a simple process in detail, produces some major resources that can be easily mined (Figure 1). The world's dominant Al resource is bauxite (139 million tonnes per annum) derived either from limestones as 'terra rossa', mainly in Jamaica, or from laterite type profiles, with the largest deposits in Australia, Brazil and Guinea. Even gold may be concentrated in, and extracted from, some lateritic and clay profiles developed over auriferous bedrock, for example in Western Australia. New Caledonia in the late nineteenth century supplied most of the world's nickel and continues to supply about 10% of annual production, mainly as green garnierite (hydrated Ni-Mg silicate) concentrated from ultrabasic host rocks into weathered bedrock and the overlying laterite.

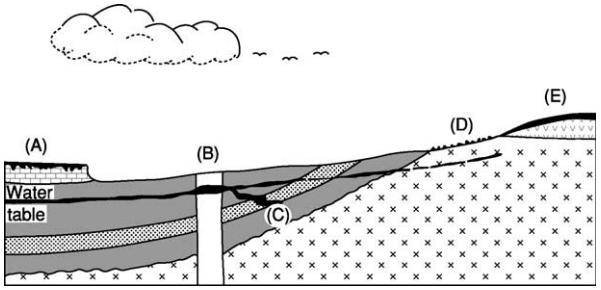


Figure 1 Diagrammatic crustal profile showing the habitats of economic deposits related to weathering processes. (A) Bauxite formed from the weathered residue of limestone. (B) Supergene enrichment of sulphide ore. (C) Uranium 'roll-front' ore body. (D) Residual placer formed on weathered basement. (E) Lateritic iron, nickel, and manganese deposits formed on weathered ferromagnesian-rich volcanics. Reproduced from Selley RC (2000) *Applied Sedimentology*, 2nd edn. p. 36. San Diego: Academic Press.

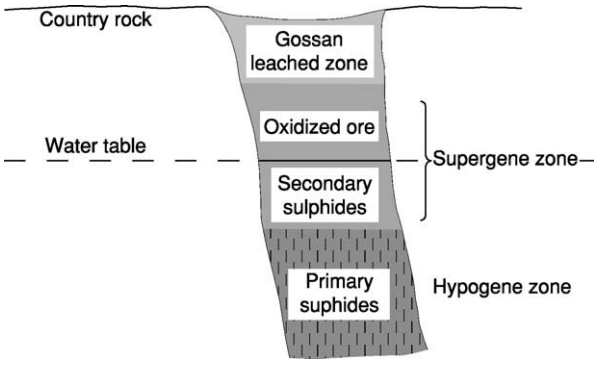


Figure 2 A simplified sketch section showing typical near-surface zonation produced by weathering of a metalliferous sulphide deposit. Reproduced from Selley RC (2000) *Applied Sedimentology*, 2nd edn. p. 37. San Diego: Academic Press.

Nickeliferous laterite with associated Co and Mn is also being mined or tested elsewhere as appropriate extractive metallurgy is developed. It is probable that residual deposits of laterite type constitute a major resource bank for the future.

Metalliferous sulphide minerals are reactive and consequently metal sulphide-rich mineral deposits weather readily in most climates to a secondary 'oxide' mineral assemblage capped by a leached iron-rich' gossan' and underlain around water table depth by an enriched supergene sulphide mineralization (Figure 2). These weathered profiles, especially for copper ores, have been economically important since early man found that the brightly coloured minerals, such as malachite, were easy to discover and to smelt to metal. In large ore bodies, such as porphyry copper sulphide deposits, the oxidized ores and supergene enriched chalcocite blankets may form an important

Figure 3 Colluvial boulders of kyanite being mined at Lapsa Buru, India. Kyanite ($A1_2\ SiO_5$) is an industrial mineral used in mullite refractories such as sparking plugs. Photo: GR Davis.

separate resource. These require different techniques of mining and treatment, but in many cases they provide an early return on capital because of the shallow working depth.

A most striking and economically important effect of prolonged weathering cycles is the conversion of banded iron formation (BIF) into high-grade iron ore by leaching of silica, thus raising the iron content and breaking down hard rock to softer fabrics that are cheaper to mine.

Minerals resistant to weathering may be liberated from their host rocks and concentrated nearby as eluvial and colluvial deposits, such as kyanite (Figure 3).

Resistate minerals may also be carried away in water by erosion, to be deposited elsewhere as placer deposits of heavy minerals such as gold, wolfram and cassiterite. Gold, the eternal metal, has provided wealth and romance through the ages. Jason and his Argonauts sought the Golden Fleece, the sheep's coat in which fine gold was trapped by washing alluvium. Two millennia later came the famous rush to the placer deposits of the Yukon, and today's prospecting by metal detector in Australia for eluvial nuggets grown in the weathering environment. Modern placers no longer provide a major source of the noble metals, but platinum placer mining is important in Russia.

Mechanically concentrated by the action of gravity in moving water, placer deposits, both fluvial and marine, are of major economic importance (Figure 4).

Cassiterite placers in Indonesia, Malaysia and Nigeria are fluviatile, and the coastal environment includes the titanium sands of Florida, and a growing number of black beach sand deposits. Mining of ilmenite, rutile and zircon sands dominates the market for these industrial minerals, which occur as very

low grade (3–6% total heavy minerals), but cheaply mined deposits at places along the wave-swept sandy shores of southern Africa, Australia, and India. Special interest attaches to placer deposits of the 'forever' mineral diamond, in South Africa. Average gem quality and value is progressively enhanced as stones are swept from their parent kimberlite pipes down the Orange River via river terrace alluvial deposits to the even more vigorous environment of the marine gravel terraces and wave cut platforms of the south-western coast of Africa. Only the best survive the journey (Figure 5).

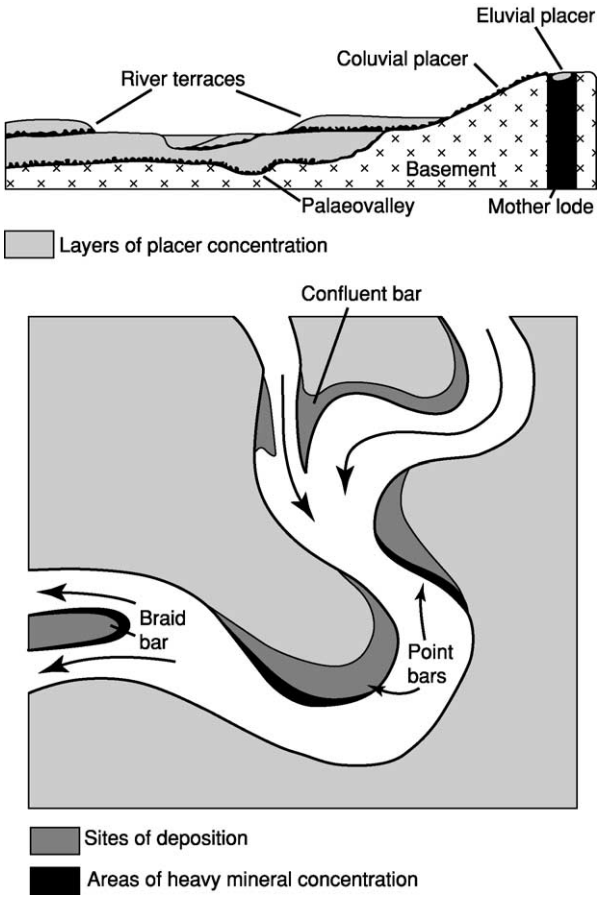


Figure 4 Cross-section and plan showing the loci of alluvial and related placer mineral deposits. Reproduced from Selley RC (2000) *Applied Sedimentology*, 2nd edn. p. 205. San Diego: Academic Press.

Precipitation of salts from solution has been known to mankind since time immemorial. Salinas, or salt pans, are widely used to evaporate sea water and recover a succession of salts, notably halite (common salt), and so mimic the natural process of salt accumulation is closed or restricted marine basins. A somewhat different process concentration of gypsum and other salts can be observed in the sabkha flats fringing hot desert seas. Both processes contribute to the thick salt strata found in many sedimentary basins. Such salt deposits not only form the basis of large scale salt mining, but play an important role in petroleum geology as source rocks, traps and seals.

One observable, but little understood process that deserves much more attention concerns the metal-fixing powers of bacteria, as seen, for example, in some iron and manganese deposits. Suggestions concerning the possible role of bacteria have been made for the Late Precambrian disseminated Cu deposits of the Zambian Copperbelt, where stromatolites are developed. Pyritic Au-U ores of the Late Archaean (*ca.* 2.8 Ga) Witwatersrand conglomerates in South Africa also contain carbonaceous material. This has been attributed to primitive organic matter or remnants of algal colonies that grew at the time of sedimentation, but another theory holds that it was introduced as oily material much later during fracturing of the lithified sediments. Progress in this field is likely to come via research in mineral processing

Figure 5 Diamond deposits on the south-western coast of Africa. (A) Diamonds are concentrated in deeply gullied wave-cut rock platforms exposed after stripping off later wind-blown sand cover. (B) Well-rounded hard pebbly gravel exposed on a raised beach. The heavy mineral content includes diamonds in hydraulic balance with larger pebbles. (Fountain pen lower left shows scale). Photo: GR Davis.

technology, where increasing use is being made of bacterially assisted leaching of refractory metallic ores. Increased understanding of the role of bacteria in removing and precipitating metals from dilute solutions may have a profound influence upon theories of ore genesis.

The clearly important role of meteoric water in residual and depositional deposits in the present-day superficial environment reflects 'the fundamental and powerful part played by fluids' in nearly every kind of mineral deposit. Subsurface fluids may be molten magmas, circulating meteoric waters, connate waters squeezed out of sedimentary basins, magmatic waters resulting from the solidification of igneous magmas, and waters generated by metamorphism. All may contain dissolved solids and/or gases, and if heated they are known as hydrothermal solutions. Because most processes of mineral concentration cannot be directly observed, various genetic theories have been advanced on the basis of available field observations and laboratory evidence. Such theories seek to explain the origin of both the mineralizing fluid and its dissolved ore-forming constituents, and to relate the time of mineral concentration to the time of formation of the host rocks. Mineralization considered to have been formed at the same time as the host rock is termed syngenetic, whereas epigenetic mineralization is considered to be introduced into the host rock at a later time.

Conceptual theories are vital to scientific understanding and advancement, but invite debate and controversy. They tend also to trap dogmatic adherents into forcing good, but inconvenient new evidence into existing pigeonholes, or brushing it aside, unless a coherent new genetic mechanism is revealed. There is no better example of this than the denial by many geologists of the evidence for continental drift, until the discovery of sea-floor spreading provided a credible mechanism for embracing that same evidence. Development of the new theory of plate tectonics during the 1960s transformed the scientific framework of geology, including views on the nature and origin of mineral deposits. During that same period some other discoveries of observable natural mineral concentration systems, currently active at or near the surface of the Earth, had a major impact on theories of ore genesis. Modern sea-floor exploration revealed the vast extent of the manganese nodules described by the Challenger expedition of 1873. This demonstrated not only an enormous potential resource of Cu, Co, Ni and other associated metals, but also the potential of cold sea water as a dilute mineralizing fluid. Also at the Earth–Ocean interface, the hot brine pools and underlying soft ferruginous muds rich in Zn, Cu and Ag discovered in 1965 in the Red Sea deeps demonstrated an exhalative deposit still forming at the present day in a continental rift system. On the mid-Atlantic ridge the discovery of active 'black smoker' hydrothermal vents and massive sulphide deposits made a dramatic impact. A 1600 m deep geothermal well in the Salton Sea in California (1962) tapped hot metalliferous brines that precipitated dark siliceous deposits containing about 20% Cu, 1% Ag and other metals in the discharge pipe at surface temperatures and pressures. Contemporaneously with these revelations, fluid inclusion and chemical studies were highlighting the potency of brines and chloride complexes as solvents and transporters of metals.

Despite current knowledge of present-day processes, the interpretation of many features of mineral deposits in ancient sediments remains controversial, and will be discussed later.

Associations with Basic and Ultrabasic Rocks

These deep-seated rock types with low silica, high Fe-Mg content are the source of leading world supplies of minerals as diverse in use as chromite, platinum group minerals, vanadium, nickel, and diamonds. In all these rocks, the problems of ore genesis are very much part of the problems of rock genesis. The valuable elements or minerals are an integral part of the geochemistry of the parent magma, which is itself the ore forming fluid.

Layered basic intrusions display magmatic segregation during undisturbed cooling and differential crystallization, a process readily understood through clear field relationships supported by laboratory studies on ore textures. Heavy oxide minerals such as chromite crystallize and settle into discrete layers as syngenetic magmatic sediments. Sulphide minerals in contrast are deposited late (in many cases at the base of the intrusion or injecting wall rocks) from an immiscible sulphide melt that persists to the last stages of magma crystallization. The oxide assemblage in strongly layered basic rocks is exemplified by the repetitive and regionally extensive chromite and vanadiferous magnetite layering of the giant early Proterozoic Bushveld Igneous Complex in South Africa. The closely associated and famous platiniferous Merensky reef is coarse grained (pegmatoidal), with minor sulphides of Ni and Cu. The sulphide assemblage is best known from the nickel mines of Sudbury, Canada, a differentiated, but not layered basic complex, now considered to have been triggered by meteoric impact, and the Norilsk deposit in flood basalts in Siberia. Ultrabasic rocks of Archaean age (which include komatiite lavas) are also host to important nickel-copper-iron sulphide segregations in

the Yilgarn block of Western Australia (another discovery of the 1960s) and to very large masses of metallurgical grade chromite at Selukwe (now named Shurugwi) in Zimbabwe. The Selukwe occurrences contrast with the genetically similar, but much smaller podiform chromite segregations in ophiolites of the Tertiary Alpine belt in eastern Europe.

A source of ilmenite alternative to beach sands is found in alkaline igneous complexes such as the anorthosites at Egersund in Norway, Allard Lake in Canada and Stillwater in the USA. The distinctive pipe-like intrusive carbonatites are closely associated with peralkaline intrusive complexes and are mined for an unusual range of minerals. The highly mineralized Palabora complex, South Africa (Figure 6), produces copper (average ore grade 0, 54% Cu), apatite, vermiculite, magnetite, baddeleyite (a zirconium oxide), uranium and by-product sulphuric acid. Kimberlite pipes, diatremes and some dykes are famous as the primary source of diamonds. Economic grades are typically only a few carats to the tonne (1 carat = 0.2 g), but most kimberlites do not contain diamonds in sufficient quantity and/or quality to meet economic criteria. These deep-seated intrusives bring xenoliths of mantle material to surface, and diamond mining has spurred geological research. This has resulted in genetic understanding that has in turn lent impetus to the current exploration boom and new discoveries – a sequence of events typical of progress in economic geology.

Figure 6 Palabora mine open pit, averages 1750 M across the rim and has reached its economic working depth of 822 M in 38 years. Beneath the open pit underground mass mining by block caving methods is now producing 30 000 tonnes per day of copper ore from the core of this very productive carbonatite pipe. Photograph courtesy of Rio Tinto plc.

Associations with Felsic Igneous Activity

Granitic magmas consolidate in a dynamic and high-energy environment dominated by superheated fluids at high temperatures and pressures. As rising magma slowly consolidates the portion still melted becomes progressively more enriched in volatiles. Release of pressure, by volcanic eruption, for example, may result in the mechanical release of gases capable of extensive fracture damage to both the intrusive and also the intruded country rocks. At high crustal levels the heat around the magma may create convection cells of surrounding groundwaters and augment the fluid content, so that episodic igneous and hydrothermal activity may persist over tens of millions of years. The fluids incorporated into the magmatic system may thus include deep connate waters and meteoric waters, both of which could be mineralized through leaching of the rocks that they traverse. These mineralized fluids may then transport and precipitate their dissolved load in rocks at higher levels, at lower temperatures and pressure, or escape to the surface. The terms hypothermal, mesothermal and epithermal are used to describe mineral assemblages attributed to hydrothermal fluids at relatively high, moderate and low temperatures and pressures (Figure 7).

The vast range of mineral deposits associated with felsic intrusions include pyrometasomatic replacement such as skarns (tactites) forming scheelite at contacts with limestones, and corundum at contacts with basic rocks. Complex coarse grained pegmatites are a product of late-stage fluids, and are mined for a variety of minerals including cassiterite, wolframite, beryl, mica and lithium minerals. Extensive vein systems may form in and around intrusive granite cupolas, well displayed by the Cornubian batholith of southwest England, mined since pre-Roman times, but no longer economic. Rich metallic mineralization of Sn-Cu-Zn-Pb displays a marked zonal pattern, and hydrothermal alteration of some granite areas (supplemented by later surface weathering) formed extensive kaolin deposits. Zonal patterns of wall-rock alteration and of metallic-sulphide mineralization also characterize the immensely productive, but low-grade porphyry copper and molybdenum deposits that contribute so much to current world production of base metals and by-product gold. These huge ore bodies are formed as impregnations at depth within and around hydrofractured calc-alkaline intrusives above subduction zones at destructive plate margins. Many gold- and silver-rich epithermal ore deposits formed nearer the surface are attributed to late-stage magmatic or mixed magmatic-meteoric fluid systems linked to the porphyry bodies. One interesting

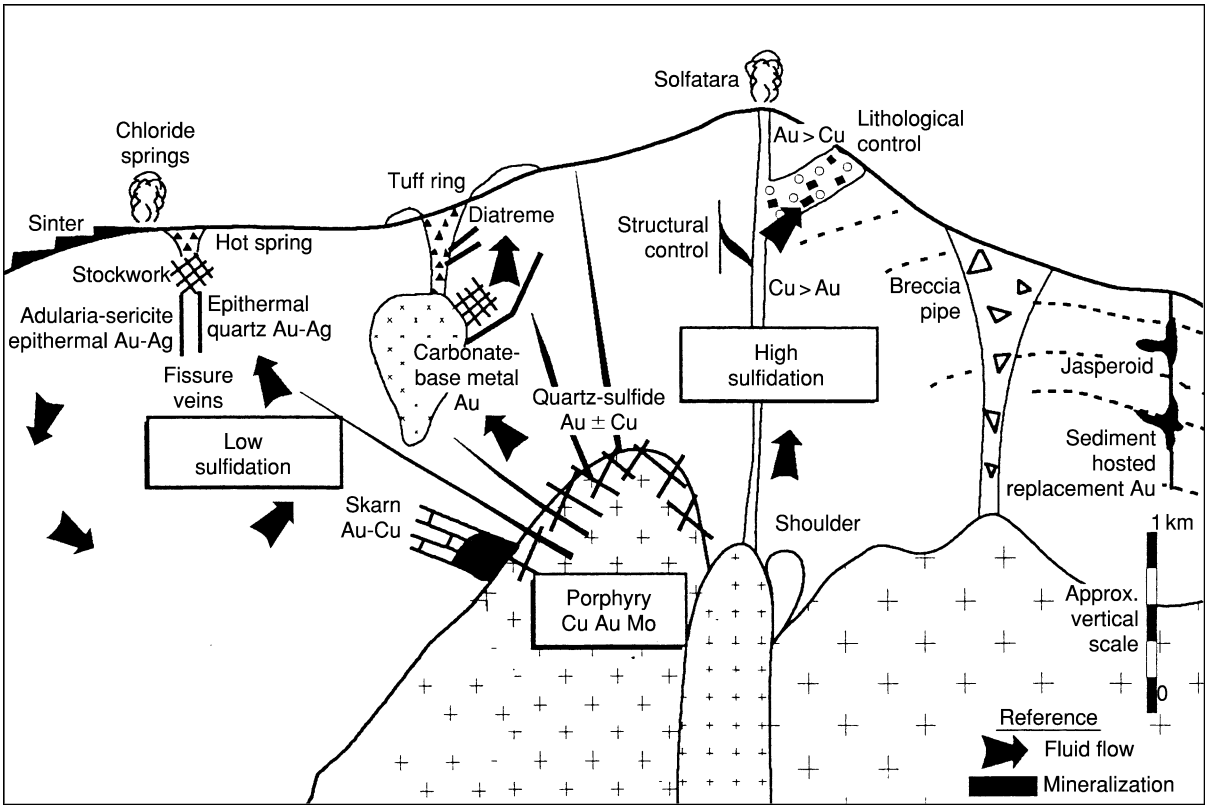


Figure 7 Pacific Rim gold-copper mineralization models. This cross-section sketch is an example of the types of mineralization associated with felsic intrusive rocks. Reproduced with permission from the Society of Economic Geologists Inc., *Southwest Pacific Rim Gold-Copper Systems: Structures, Alteration, and Mineralization*, p. 6, Fig. 1.1, Corbett GJ and Leach TT (1998).

example is at Lihir island within the Pacific 'ring of fire', where large-scale mining recently commenced on very fine-grained gold in pyritic mineralization within an inactive volcanic caldera. The Carlin-type deposits of Nevada, on the other hand, are situated along regional structural trends. The gold mineralization is commonly so finely disseminated within the dark impure calcareous host rocks that mining limits are determinable only by assay, and the discovery in the 1970s was hailed as 'virus gold'.

Rhyolites and dacites, the extrusive forms of granites and granodiorites, are host rocks to economically and geologically important types of base-metal sulphide mineralization that are transitional between sub-surface and surface. The Iberian pyrite belt of southern Spain and Portugal, and the Kuroko deposits of Japan are leading examples. Subsurface mineralization of transgressive stockworks in explosive rhyolite domes underlies massive stratiform sulphide bodies that are interpreted as deposited at surface in a submarine environment (Figure 8). These two closely associated deposits, of distinctly different style and appearance, are attributed to the same mineralizing ore fluids at work in the subsurface and, after exhalation, above the rock-water interface.

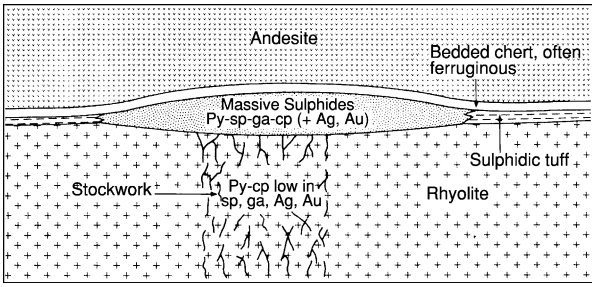


Figure 8 Schematic section through an idealized volcanic-associated massive sulphide deposit showing the underlying feeder stockwork and typical mineralogy. Py = pyrite, sp + spha- sphalerite, ga = galena, cp = chalcopyrite. Reproduced from Evans AM (ed.) (1995) *Introduction to Mineral Exploration*. Oxford: Blackwell Science.

Associations with Ancient Sedimentary Rocks

Knowledge of present-day surface processes assists in the understanding of ancient sedimentary deposits such as coal, and salt deposits formed by evaporation of brines from closed basins or sabkhas. Geological details vary, but the genetic picture is clear enough to guide exploration for further resources. However,

controversy and lack of understanding surrounds many deposits found in sediments, especially metalliferous ores. Dispute tends to centre on the nature and origin of the ore forming fluid, its movement through the crust, the origin of the metals, and the mechanism of their precipitation.

The controversy is well illustrated by the continuing debate concerning the origin of the giant Witwatersrand gold deposits of South Africa, the world's greatest gold producer for over a century. A well-preserved Late Archaean sedimentary basin, about 6 km thick and 280 km across, contains pyrite-uraninite-gold mineralization in conglomerates developed at several unconformities from the bottom to the top of the sequence (Figure 9). Economic interest is focused on a series of conglomerates towards the top of the basin and situated near its edge. Despite compelling evidence for a paleoplacer origin, undisputed signs of some hydrothermal activity in the sediments are still being used to argue the minority view that epigenetic hydrothermal processes played a major role in the introduction of the gold. Another syn-sedimentary hypothesis holds that gold was precipitated from solution, on the evidence of carbon ('thucholite') that may be derived from primitive organisms. The uncertainties continue, as some others hold that the carbonaceous matter was introduced from an external source long after consolidation of the sediments.

Ores in sediments may be discordant irregular masses of clearly epigenetic origin; or concordant and possibly, therefore, either syngenetic with the enclosing sediment, or diagenetic (that is, due to processes of diagenesis). The term 'stratabound' is used to indicate confinement of the mineral deposit between or within sedimentary strata, and 'stratiform' to indicate that the deposit shows internal layering or stratification.

Most Phanerozoic ironstones are sedimentary, and the intriguing Banded Iron Formation, widespread in Proterozoic and Archaean basins, is well documented as the major sedimentary protore of high-grade iron ore production. Many stratiform massive base-metal sulphide deposits, once considered by many to be epigenetic replacements due to hydrothermal fluids from hidden igneous bodies, have been shown to be the products of hydrothermal exhalations and a volcano-sedimentary environment, as for the Kuroko and Iberian deposits mentioned above. At the other pole are the Late Proterozoic stratabound Cu-Co disseminated sulphide ores of the Copperbelt in northern Zambia and the adjacent Congo. Mineralization of just a few percent sulphides is confined to strata only tens of metres thick, but extending (in large regional folds) over thousands of square kilometres. The only igneous rocks known nearby are basement granites that form the hilly topography upon which the cupriferous sediments were deposited, and much later basic dykes. The detailed evidence, including a marked facies zoning upwards and shoreward of pyrite-chalcopyrite-bornite-chalcocite (Figure 10), convinced most geologists that the ores are not simple hydrothermal replacements as favoured after mine development in the 1920s. An origin due to syngenetic precipitation of pyrite and copper sulphides by bacterial action has long been advocated, together with effects due to diagenesis and metamorphism, but not convincingly demonstrated. Current research based on increased regional knowledge (over 1000 million

Figure 9 A typical specimen of Witwatersrand gold ore from South Africa. Pyrite and minor gold in the groundmass supporting well-rounded, unmineralized quartz pebbles, 25 cm across. Photo: GR Davis.

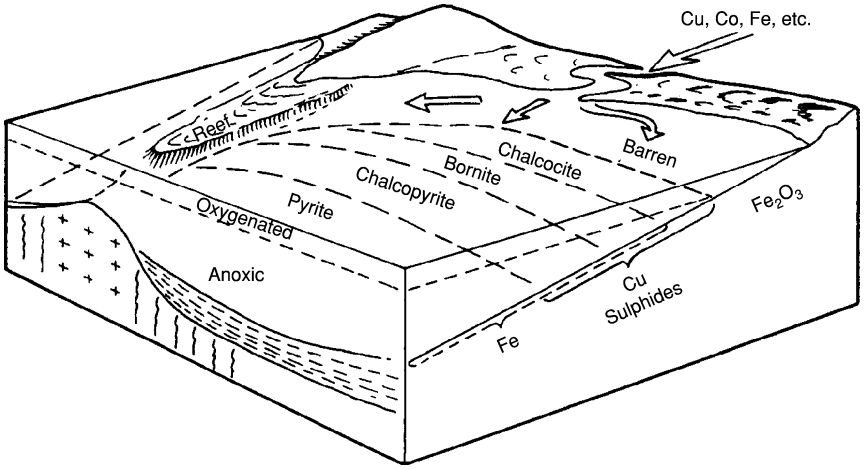


Figure 10 A diagram to illustrate the environment of sedimentation of Zambian Copperbelt host rocks. A marked zonation of copper and iron sulphides exists near the shoreline formed by the hilly basement rocks. The diagram (1976) illustrates a syngenetic theory of origin for mineralization, which is not universally accepted. Reproduced from Fleischer VD, Garlick WD, and Haldane R (1976) Geology of the Zambian copperbelt. In: Wolf KH (ed.) *Handbook of Stratabound and Stratiform Ore Deposits,* vol. 6, pp. 223–350. New York: Elsevier.

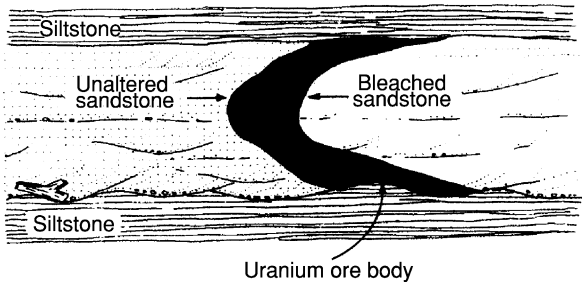


Figure 11 Sketch illustrating the occurrence of carnotite roll-front ore bodies in fluvial channel sandstones. The direction of front migration in this case was from right to left. Reproduced from Selley RC (2000) *Applied Sedimentology*, 2nd edn. p. 38. San Diego: Academic Press.

tonnes averaging about 2.7% Cu has been mined) and laboratory studies suggests that metal-bearing sulphate brines were introduced through tectonic thrust faulting and that copper sulphides were epigenetically precipitated by organic matter in the host rocks. Neither theory has demonstrated the source of the metals in the seawater (for syngenesis) or in the deep-seated fluids (for epigenesis), or fully explained all the observed field relationships and laboratory data.

Knowledge of diagenetic processes is advancing and is illustrated by roll-front uranium deposits in Colorado, where carnotite has been precipitated from migrating pore waters in confined aquifers (Figure 11). The mottled lead sulphide sandstone ore bodies of the extensive Laisvall and related deposits of the Swedish Caledonides may also be due to migrating mineralized fluids.

The discordant masses of zinc and lead sulphide mineralization that are widespread in sedimentary carbonate sequences, such as the Mississippi Valley deposits of the USA, have clearly been deposited later than their host rocks. The ore forming fluids were for many years attributed to igneous sources, for lack of a better option. It has now been shown in some well-documented cases that oilfield brines migrating up-dip from deep sources are the ore forming fluids from which sulphides were precipitated. From years of meticulous field observations backed up by laboratory studies (such as fluid inclusions, stable isotope and geochemical analysis) there has emerged an elegant and logical understanding that both oil and ore deposits are linked through such common controls as source, migration path, trap and seal. The effect of this advance in knowledge about mineral deposits has been dramatic. The philosophy of prospecting for new mineral fields of this type has switched, to put the case over-simply, from searching carbonate sequences near intrusive granites to studying fluid migration paths in sedimentary basins containing hydrocarbon source rocks and evaporites. An understanding is developing of the thermochemical reactions (at elevated pressures and temperatures) between hydrocarbons and sulphate brines that can release aggressive rock-altering fluids and hydrogen sulphide, a powerful precipitant of base metals. This example, and that of the Zambian Copperbelt, demonstrates that ores in ancient sediments are best studied, like the hydrocarbon deposits to which they may be linked, in the context of the entire geological history of the containing sedimentary basin.

Metamorphic Associations

Metamorphic processes may either generate new mineral deposits, or modify the characteristics of existing deposits. A number of industrial minerals are formed by metamorphism, including kyanite for refractories, and garnet for abrasives. Talc forms in metamorphosed ultrabasic rocks, which are also the major source of chrysotile asbestos. When olivine rich rocks such as dunite are serpentinized, the fibrous form of serpentine, chrysotile, is formed in cross-fibre veins that seem to develop under conditions of tensile stress and hydrofracturing. A dense enough network of veins (a few per cent by volume) of good-quality fibre converts large masses and lenses of serpentinized rock into a mineral deposit of economic potential, usually suitable for mass-mining methods (Figure 12). Despite the health hazards associated with asbestos, world production is still about 2 million tonnes per annum. There are large deposits in the Urals, the Palaeozoic of the eastern Townships in Quebec, and in the Archaean at Zvishavani (Shabani) in Zimbabwe. Tectonic deformation is also an important precursor to epigenetic mineralization ('ground preparation') as it forms shear zones or fractures that provide channelways for mineralizing fluids. Dynamic metamorphism and the repetitive action of a process known as seismic pumping may drive mineralizing fluids from depth into higher level vein systems. Activated by earthquake movements, fluids accumulating at depth in large fault systems are expelled upwards through channelways where dissolved substances may be precipitated as successive generations of vein fillings. Large numbers of vein systems, commonly containing gold, are associated with the shear and fracture zones typical of greenstone schist belts in Archaean cratons and Precambrian Shield areas. Geological and mining conditions may thus be very similar in such gold fields areas as far apart as Australia, Canada and West Africa.

A second and geologically important aspect of metamorphism relates to the fact that pre-existing mineral deposits, like all other rocks, must suffer the effects of regional metamorphism. Deposits not only become faulted, folded and sheared into new geometric shapes and attitudes that affect the cost of mining, but may respond by recrystallization in ways that materially affect their amenability to mineral extraction and beneficiation. Sulphide minerals are particularly prone to recrystallize and anneal, or become mobilized into new open spaces such as rock cleavage. Veins of gangue and ore minerals may also be formed by lateral secretion driven by metamorphic temperatures and stresses. The overall effect is to overprint and mask the original ore fabrics, especially at the micro and meso scales. Such difficulties lead to additional debate about controls on ore formation and ore genesis, which in their turn affect decisions on how and where to explore for similar mineralization. Some examples include the famous massive sulphide Pb-Zn ores of Broken Hill (Australia) in strongly folded sillimanite grade gneisses, the folded disseminated Cu deposits of the Zambian copperbelt, the great conglomerate hosted Au-U deposits of the Witwatersrand (South Africa) and the twice-folded and metamorphosed stratabound Cu-Co deposits at Kilembe (Uganda).

Figure 12 A network of cross-fibre chrysotile asbestos veins in serpentinite, Havelock mine, Swaziland. This length of fibre and intensity of mineralization constitutes superior grade ore. Photo: GR Davis.

Discovery of Further Mineral Deposits

The AGI definition of economic geology includes the 'application of geological knowledge and theory to the search for.... mineral deposits', recognizing the fact that the extractive industries must meet future demands for new materials and replace mined out deposits with new discoveries. Over the centuries the world's need for mineral supplies has always been met by some combination of chance finds at outcrop, and indirect indications from prospecting methods such as panning, geophysical and geochemical surveys, aerial photography and earth satellite imagery. Whether the search is for an extension to ore at a working mine or a new grassroots discovery, the choice of where to look for success is based on someone's idea that a given part of the Earth's crust will actually contain economic mineralization. The best ideas are based upon two factors – superior knowledge of the observed and detailed characteristics and field associations of the type of deposit targeted, and a valid theory of its genesis. These two factors have long been used empirically in prospecting, but over the past 20 years the available information has been systematically gathered and codified into a large number of 'deposit models'. These models are continually being tested, improved, and added to as discoveries of new and unexpected types of mineralization are made. For example, the huge Olympic Dam copper-uranium deposit, discovered under 350 m of barren cover rocks in South Australia in 1975, now represents a broad and complex group referred to as iron oxide-copper-gold deposits.

The search becomes more difficult as the number of undiscovered deposits decreases and the emphasis moves further towards deposits deeply buried and hidden under overlying rock cover. This growing challenge is being met with an integrated approach of new and broader ideas backed by an ever-improving tool-kit of geophysical and geochemical field exploration technologies and elegant laboratory techniques. The tool-kit also includes readily available space satellite imagery that now provides the broad perspective on which to study major crustal features and lineaments that may link the surface to the depths below the crust. Economic mineral concentration is seldom an isolated event, but rather one result from a dynamic system linking tectonics, magmatism, volcanicity, and sedimentation in complex ways. To guide exploration to best advantage the entire system should be investigated and understood as far as possible. Twenty-first century economic geologists, striving to understand the ore forming systems in our complex Earth's crust, may be poised to take the biggest step forward in the subject since 1556, when Georgius Agricola (Bauer) famously classified ore deposits and recorded the state of the art of mining in his classic work, *De Re Metallica*.

See Also

Economic Geology. **Mining Geology:** Exploration; Mineral Reserves; Hydrothermal Ores; Magmatic Ores. **Sedimentary Rocks:** Banded Iron Formations; Ironstones.

Further Reading

Cooke DR and Pongratz J (eds.) (2002) *Giant Ore Deposits: Characteristics, Genesis and Exploration. CODES Special Publication 4*. Hobart: University of Tasmania.

Craig JR, Vaughan DJ, and Skinner BJ (2001) *Resources of the Earth: Origin, Use and Environmental Impact,* 3rd edn. New York: Prentice-Hall.

Davis GR (1988) Is Metallogeny a Practical Exploration Tool? *Episodes* 11(2): 105–110.

Derry DR (1980) *World Atlas of Geology and Mineral Deposits*. London: Mining Journal Books.

Evans AM (1993) *Ore Geology and Industrial Minerals, an introduction,* 3rd edn. Oxford: Blackwell.

Kirkham RV, Sinclair WD, Thorpe RI, and Duke JM (eds.) (1993) *Mineral Deposit Modeling. Geological Association of Canada Special Paper 40*.

Selley RC (2000) *Applied Sedimentology,* 2nd edn. San Diego: Academic Press.

Stanton RL (1972) *Ore Petrology*. New York: McGraw-Hill.

MINERALS

Contents

Definition and Classification

E H Nickel, CSIRO Exploration and Mining, Wembley, WA, Australia

Introduction

Minerals, the individual components comprising rocks, are generally defined in terms of chemical composition and crystal structure, and most classification systems are based on these properties. Minerals are formed by geological processes and include both terrestrial and extraterrestrial materials. The total number of minerals generally accepted as valid by the mineralogical community is about 4000. Mineraloids are mineral-like substances such as synthetic materials, human-influenced substances, and some biological materials that do not satisfy all the criteria for the definition of a mineral species.

Definition of a Mineral Species

Minerals are substances formed by geological processes that occur on Earth or in extraterrestrial bodies. A mineral is defined on the basis of its chemical composition and crystal structure, and to qualify as a mineral species, a substance must have a unique combination of these properties. Some minerals have a well-defined composition (quartz, for example, is SiO_2), but many have a variable composition whereby some chemical constituents are replaced by others. For example, the mineral olivine can be regarded as a solid-solution series between forsterite (Mg_2SiO_4) and fayalite (Fe_2SiO_4) in which Mg^{2+} and Fe^{2+} substitute for each other in the crystal structure. In such

a case, the two end-members are regarded as species, with a compositional range extending from the end-member composition to the midpoint of the solid-solution series, namely 50 mol%. 'Olivine' is therefore regarded as a series or group name. Other solid-solution series are more complex, as in the case of ternary or higher order solid solutions, as exemplified by the mutual replacement of F, OH, and Cl in the apatite series, giving rise to the end-member species fluorapatite, hydroxylapatite, and chlorapatite, respectively. Further complications may involve coupled substitutions, such as in britholite-(Ce), in which the replacement of Ca^{2+} in the apatite structure by Ce^{3+} is balanced by the substitution of P^{5+} by Si^{4+}. In all such cases, the species is defined on the basis of the predominant ion in a specified structural site.

Some minerals with identical compositions have different crystal structures; examples include quartz, tridymite, and cristobalite, all of which have the chemical formula SiO_2 but which crystallize in the hexagonal, orthorhombic, and tetragonal crystal systems, respectively. Such substances are called polymorphs and qualify as separate mineral species because they have different structures. On the other hand, there are minerals with the same crystal structure but with different compositions, such as galena (PbS), periclase (MgO), and halite (NaCl). These, too, are regarded as separate species.

A structural variant that is sometimes encountered, notably in micaceous minerals, is created by the rotation of one structural unit with respect to its structurally equivalent unit. When this perturbation is distributed throughout the structure as a regular stacking sequence, the structural variant is called a polytype. Polytypes are not regarded as separate species and are indicated by the addition of a hyphenated suffix to the root mineral name. The suffix takes the form of a numeric symbol that represents the periodicity with respect to the basic structural element, an alphabetic symbol that represents the crystal system, and sometimes a numeric subscript that represents the type of stacking. Some examples of mica polytypes are muscovite-1M (monoclinic, with a periodicity of 1), muscovite-$2M_2$ (monoclinic, with a periodicity of 2 and a particular stacking sequence), muscovite-3T (trigonal, with a periodicity of 3), and muscovite-2A (anorthic, i.e., triclinic, with a periodicity of 2).

Although the crystal structure of a mineral is one of the important properties defining a mineral species, some mineral substances do not have a long-range structural arrangement of the atoms comprising the mineral. Minerals lacking this long-range order are termed amorphous (if they were created in this form) or metamict (if they originally possessed long-range structural ordering, but the structural order was later destroyed, usually as a result of ionizing radiation). An amorphous mineral may be accepted as a valid species if there is evidence that it is chemically homogeneous and if spectroscopic characterization shows that it is unique. A metamict mineral may be accepted as a valid species if it can be established with reasonable certainty that the original substance was a crystalline mineral, generally by returning it to its original crystalline state by heat treatment.

Many minerals were formed under conditions of high temperature or pressure (or both) and are metastable under ambient conditions; others may tend to hydrate or dehydrate when removed from their place of origin. Such minerals may require special procedures to prevent their decomposition before an investigation is complete. The use of special procedures in the investigation does not preclude the acceptance of a metastable or unstable substance as a mineral, if it can be adequately characterized and if it meets the other criteria for a mineral.

The Validation of Mineral Species

Minerals have been given names since before the dawn of history, and some of these names have survived as valid species names to the present day. However, the vast majority have not. Following the establishment of the Commission on New Minerals and Mineral Names (CNMMN) of the International Mineralogical Association (IMA) in 1959, the introduction of new mineral names has been strictly controlled, and publication of new mineral names and descriptions requires prior approval by the national representatives of the CNMMN. Proposals for the creation and naming of a new species require a complete description of the mineral, including chemical composition; crystallographic, physical, and optical properties; crystal structure (if possible); and specification of the type locality where the mineral was discovered. If approved, the specimen providing this information becomes the type specimen for the species and is the specimen to which all subsequent descriptions of the mineral from other localities must be compared.

Minerals in the literature prior to the formation of the CNMMN have been progressively winnowed down to a generally accepted list of valid mineral species by the mineralogical community in general, and by activities of the CNMMN in particular. Subcommittees have been established by the CNMMN for the review of mineral groups, including the amphibole, mica, platinum-group alloys, pyrochlore, pyroxene, and zeolite groups, and reports of these subcommittees have been published in the open literature. Discreditation or redefinition of individual species has also been done under the auspices of the CNMMN.

Mineral Names

The author(s) of the original description of a valid new mineral have the prerogative of naming the mineral. However, the name must be approved by the CNMMN prior to publication. Mineral names have various derivations, the principal ones being the geographical locality of the discovery, a particular characteristic of the mineral, and the name of a person. Such names are sometimes called "trivial" to distinguish them from systematic ones. In an effort to reduce the proliferation of trivial names, the CNMMN has approved the use of root names with suffixes. Suffixed mineral names were introduced by AA Levinson in 1966 for rare-earth minerals, and involves the addition of a hyphenated chemical symbol in brackets after the root name, e.g., synchysite-(Ce). Such suffixes are generally referred to as Levinson modifiers, and the nomenclature of all minerals with one or more rare-earth elements predominating in a structural site must conform to this usage. The system of Levinson modifiers has subsequently been extended to some other mineral groups, notably the zeolites.

Mineral Varieties and Varietal Names

In addition to valid mineral names, which apply to mineral species or groups, other names are commonly used for particular varieties of minerals, generally those with distinctive coloring. This practice is especially common in gemology. Deeply coloured varieties of corundum, for example, have been given varietal names such as ruby (red) and sapphire (blue). Such names have no validity as species names, and their use is not controlled by the CNMMN.

Mineraloids

Mineraloids are substances that have some of the properties of minerals, but are not regarded as valid minerals, usually because they have not been formed exclusively by geological processes. One class of mineraloids is those produced by biological processes, such as mineral-like calculi in animals or organic crystals in plants. A pearl is therefore classified as a mineraloid rather than as a mineral. Some biogenic substances are subsequently found to be formed by geological processes as well, such as the urinary calculi whewellite and weddellite, and these then qualify as mineral species.

Another class of mineraloids includes synthetic substances. Even though they may have a definite chemical composition and a known crystal structure, they do not qualify as minerals because geological processes have not been involved in their creation. Such substances are called anthropogenic. Substances that are formed by a combination of anthropogenic and geological processes are also classified as mineraloids. These include substances formed as a result of mine or waste-dump fires and by the action of water on man-made substances. In the past, such substances were accepted as mineral species, as, for example, the submerged ancient Laurium slag 'minerals', but in recent years the CNMMN has enunciated a policy of not accepting occurrences of such substances as minerals.

Mineral Classification

Historical Background

The ancient classification of minerals was based mainly on their practical uses, minerals being classified as gemstones, pigments, ores, etc. Probably the earliest classification based on external characteristics and on some physical properties, such as colour, fusibility, malleability, and fracture, was that of Geber (Jabir Ibn Hayyaan, 721–803), later extended by Avicenna (Ibn Sina, 980–1037), Agricola (1494–1555), and AG Werner (1749–1817). This system was substantially refined by F Mohs (1773–1839) in his *Natural-History System of Mineralogy* (1820). With Werner, physical classification attained its maturity, and was in general use, by the end of the eighteenth century. Linnaeus (1707–78) attempted to classify minerals primarily by their external morphology, with a hierarchical system involving subdivision into genus, order, and class. A purely chemical classification was proposed by T Bergmann (1735–84), but this approach was premature, because many chemical elements had not been discovered at that time and analytical procedures were in their infancy. AF Cronstedt (1722–65) seems to have been the first to devise a classification scheme involving both chemical and physical properties, with chemistry predominating. Systematic crystallography was initiated by JBLR de l'Isle (1736–90), and this concept was applied by RJ Haüy (1743–1822) in *Traité de Minéralogie* (1801), in which he presented a mineral classification scheme based on the 'nature of metals', or, as it would be expressed now, the nature of cations.

With advances in chemistry, chemical properties became increasingly important, and a chemical classification of minerals was proposed in 1819 by JJ Berzelius (1779–1848). He recognized that minerals with the same non-metal (anion or anionic group) have similar chemical properties and resemble each other far more than do minerals with a common metal. He considered minerals as salts of anions and anionic complexes, namely, as chlorides, sulphates,

silicates, etc., rather than as minerals of Zn, Cu, etc. At this time, CS Weiss (1780–1856) introduced the seven crystal systems (1815) and Mitscherlich discovered isomorphy (1819) and polymorphy (1824). Gustav Rose (1798–1873) combined chemistry, isomorphy, and morphology to produce a chemical–morphological mineral system, and this was further developed by P von Groth (1843–1927) in his five editions of *Tabellarische Übersicht der Mineralien nach ihrer Kristallographisch-chemischen Beziehungen*. JD Dana (1813–95), in his *System of Mineralogy* (1837), based his classification system primarily on chemistry, and this emphasis has been maintained throughout subsequent editions of the *System*.

After 1912, following the discovery of the phenomenon of X-ray diffraction by crystals by M von Laue (1879–1960) and WH Bragg (1862–1942), and the elucidation of the first crystal structure (the mineral halite) in 1914 by WL Bragg (1890–1971), the crystal structures of minerals began to be taken into account in mineral classification schemes. The first classification of this type, which took into account the distribution of interatomic bonds in a structure, involved the structures of silicates, determined by Machatschki in 1928. This new field was rapidly expanded by WL Bragg in *The Crystalline State* (1933) and in the first edition of *Atomic Structures of Minerals* (1937). This combination of chemistry and structure in mineral classification was subsequently applied to many other categories of minerals, such as fluoraluminates (by Pabst; 1950), aluminosilicates (by Liebau; 1956), silicates and other minerals with tetrahedral complexes (by Zoltai; 1960), phosphates (by Liebau, in 1966, and by Corbridge, in 1971), sulphosalts (by Makovicky; 1981 and 1993), and borates (by Heller, in 1970, and by Strunz, in 1997).

The classification of silicates on the basis of polymerization of corner-sharing SiO_4 tetrahedra from insular groups to dimers, chains, rings, sheets, and frameworks proved to be a particularly useful scheme but, with the exception of the borates, this concept could not be comprehensively extended to other categories of minerals. The polymerization of cation-centred polyhedra by sharing corners, edges, and faces to form various configurations has been applied in the classification of some minerals, such as sulphates (by Sabelli and Trosti-Ferroni; 1985), copper oxysalts (by Hawthorne; 1993), and phosphates (by Hawthorne; 1998), but such schemes have only limited applicability to minerals as a whole. Other classification schemes, such as those stressing genetic aspects of mineral formation (by Kostov; 1975) or interatomic bonding (by Godovikov; 1997) have not been generally adopted by the mineralogical community.

Current Comprehensive Classification Systems

Chemical composition and crystal structure are the two properties that define a mineral species, thus it is no surprise that systems based on one or both of these properties are the most widely used. Russian mineralogists have been particularly prolific in devising crystallochemical classification schemes, e.g., Betechtin (in 1961), Povarennykh (in 1966), Lazarenko (in 1971), Godovikov (in 1983), Semenov (in 1991), and Bulakh (in 1995). Among other European mineralogists, such classification schemes appear to be less profuse, and the following few examples have been selected to illustrate the diverse approaches to mineral classification.

A classification based entirely on chemical composition is that of Hey (Table 1), whose *Chemical Index*

Table 1 Principal categories in Hey's chemical classification system

Category	*Description*
1	Elements and Alloys (including the arsenides, antimonides, and bismuthides of Cu, Ag, and Au)
2	Carbides, nitrides, silicides, and phosphides
3	Sulphides, selenides, tellurides, arsenides, and bismuthides (except the arsenides, antimonides, and bismuthides of Cu, Ag, and Au)
4	Oxysulphides
5	Sulphosalts – sulpharsenites, sulphantimonites, and sulphobismuthites
6	Sulphosalts – sulphostannates, sulphogermanates, sulpharsenates, sulphantimonates, sulphovanadates, and sulphohalides
7	Oxides and hydroxides
8	Halides
9	Borates
10	Borates with other anions
11	Carbonates
12	Carbonates with other anions
13	Nitrates
14	Silicates not containing aluminium
15	Silicates of aluminium
16	Silicates containing aluminium and other metals
17	Silicates containing other anions
18	Niobates and tantalates
19	Phosphates
20	Arsenates
21	Vanadates
22	Phosphates, arsenates, or vanadates with other anions
23	Arsenites
24	Antimonates and antimonites
25	Sulphates
26	Sulphates with halide
27	Sulphites, chromates, molybdates, and tungstates
28	Selenites, selenates, tellurites, and tellurates
29	Iodates
30	Thiocyanates
31	Oxalates, citrates, mellitates, and acetates
32	Hydrocarbons, resins, and other organic compounds

of Minerals (1950) has gone through three editions. In Hey's classification, minerals are divided into 32 main categories that are subdivided into smaller groupings on the basis of predominant cations. Most of the other classification systems combine the criteria of chemical composition and crystal structure. The system used by the International Centre for Diffraction Data (1993) categorizes the minerals into 177 groups based on crystallographic criteria; the groups are divided into subgroups on the basis of chemistry and/or crystallography. The principal groupings in *Fleischer's Glossary of Mineral Species* (1999) are similar, but lack the subgroupings.

In the system advocated in 1983 by Lima-de-Faria (Table 2), minerals are divided into aqueous and nonaqueous categories; these are then subdivided on the basis of element ratios and then on the basis of crystal-structure type. In the Dana system (Table 3), minerals are divided into 22 categories, based mainly on anion composition, except for the silicates, which comprise six categories based on polymerization of the SiO_4 tetrahedra. Further subdivision of the nonsilicates is based primarily on composition, whereas that of the silicates is based on the configuration of the SiO_4 units.

The Strunz system (Table 4), probably the most widely used classification system, divides minerals into 10 classes, based on anion composition. Further subdivisions are based on chemical and structural criteria, which are different in each of the classes. In Class 1, the primary subdivision is based on composition and the secondary one is based on structure. In Class 2, the primary subdivision is based on composition (chiefly the cation/ion ratio), the secondary one is based on further compositional criteria, and the tertiary one is based on structure. In Class 3, the primary subdivision is based on composition (principally on the compositional complexity and the presence or absence of combined H_2O), the secondary one is based on cation:anion ratio, and the tertiary one is based on structure. In Class 4, the primary subdivisions are based on cation:anion ratio and the presence or absence of H_2O, with separate subdivisions for the uranyl hydroxides, vanadates with 5- or 6-coordinated V atoms, and arsenites, antimonites, bismuthites, sulphites, selenites, tellurites, and iodates. Further subdivisions are based on cation size and structure type. In Class 5, the carbonates are subdivided first on the presence or absence of additional anions and H_2O, secondly on cation size, and ultimately on crystal structure. The nitrates are subdivided on the basis of the presence or absence of OH and H_2O. In Class 6, the primary subdivision is based on the number of borate units in the chemical formula and the secondary one is based on structure. In Class 7, the primary subdivision is focussed on the presence or absence of additional anions and H_2O,

Table 2 Principal categories in the classification system of Lima-de-Faria

Category	*Type*
Minerals without water molecules in their structure	A
	A_mB_n
	$A_pB_qC_r$
	$A_pB_qC_rD_s$
	$A_pB_qC_rD_sE_x$
	$A_pB_qC_rD_sE_xF_y$
	$A_pB_qC_rD_sE_xF_yG_z$ ...
	$A_pB_qC_rD_sE_xF_yG_z$
Minerals with water molecules in their structure[a]	A_pB_q ... E_xF_y ... *n* (aq.)

[a]Work on this category is in progress.

Table 3 Principal categories in the Dana classification system

Native elements and alloys	Phosphates, arsenates, and vanadates
Sulfides and related compounds	Antimonates, antimonites, and arsenites
Oxides	Vanadium oxysalts
Halogenides	Molybdates and tungstates
Carbonates	Organic compounds
Nitrates	Nesosilicates: insular SiO_4
Iodates	Sorosilicates: isolated tetrahedral noncyclic groups, $N > 1$
Borates	Cyclosilicates
Sulphates	Inosilicates: two-dimensionally infinite silicate units
Selenates and tellurates; selenites and tellurites	Phyllosilicates
Chromates	Tektosilicates

Table 4 Principal categories in the Strunz classification system

Class	*Description*
1	Elements (metals and intermetallic alloys; metalloids and nonmetals; carbides, silicides, nitrides, phosphides)
2	Sulphides and sulphosalts (sulphides, selenides, tellurides; arsenides, antimonides, bismuthides; sulpharsenites, sulphantimonites, sulphbismuthites, etc.)
3	Halides
4	Oxides (hydroxides, $V^{[5,6]}$ vanadates, arsenites, antimonites, bismuthites, sulphites, selenites, tellurites, iodates)
5	Carbonates (+ nitrates)
6	Borates
7	Sulphates (selenates, tellurates; chromates, molybdates, wolframates)
8	Phospates, arsenates, vanadates
9	Silicates (germanates)
10	Organic compounds

the secondary one is based on cation size, and the tertiary one is based on structure; uranyl sulphates, chromates, molybdates/wolframates, and thiosulphates are in separate subdivisions. In Class 8, the primary subdivision is based on the presence or absence of additional anions and H_2O, the secondary one is based on cation size and $OH:H_2O$ ratio, and the tertiary one is based on crystal structure. Separate subdivisions are reserved for uranyl phosphates and arsenates and for polyphosphates, polyarsenates, and polyvanadates (with V in fourfold coordination). In Class 9, the primary subdivision is based on the degree of polymerization of the SiO_4 tetrahedra in the structure; the secondary one is based on the presence or absence of additional anions, the coordination number of the cations, and, in certain cases, on the periodicity of the polymerized units. Tertiary subdivision is based on crystal structure. Zeolites are in a separate division and are further subdivided on the basis of structure. The minerals of Class 10 are divided into acetates, oxalates, benzine salts, hydrocarbons, and miscellaneous organic minerals.

See Also

Analytical Methods: Mineral Analysis.

Further Reading

Bayliss P (2000) *Glossary of Obsolete Mineral Names*. Tucson: The Mineralogical Record Inc.

Blackburn WH and Dennen WH (1997) *Encyclopedia of Mineral Names. The Canadian Mineralogist, Special Publication 1*. Ottawa: Mineralogical Association of Canada.

Clark AM (1993) *Hey's Mineral Index*, 3rd edn. London: Chapman & Hall.

Gaines RV, Skinner HCW, Foord EE, *et al.* (1997) *Dana's New Mineralogy*, 8th edn. New York: Wiley and Sons.

Johnsen O (2002) *Minerals of the World*. Princeton, NJ: Princeton University Press.

Kampf A and Gerhold G (eds.) (1998) *The Photo-Atlas of Minerals*. CD format. Los Angeles: Gem and Mineral Council, Los Angeles County Museum of Natural History.

Klein C (2002) *The Manual of Mineral Science*, 22nd edn. New York: John Wiley & Sons.

Lima-de-Faria J (2001) *Structural Classification of Minerals, vol. 1: Minerals with A, A_mB_n and $A_pB_qC_r$ General Formulas*. Dordrecht: Kluwer Academic Publishers.

Lima-de-Faria J (2003) *Structural Classification of Minerals, vol. 2: Minerals with $A_pB_qC_rD_s$ to $A_pB_qC_rD_sE_xF_yG_z$ . . . General Formulas*. Dordrecht: Kluwer Academic Publishers.

Lima-de-Faria J (2004) *Structural Classification of Minerals, vol. 3: Minerals with $A_pB_q . . .E_xF_y . . .nA_q$. General Chemical Formulas and Organic Minerals*. Dordrecht: Kluwer Academic Publishers.

Mandarino JA and Back ME (2004) *Fleischer's Glossary of Mineral Species 2004*. Tucson: Mineralogical Record Inc.

Martin RF (ed.) (1998) *The Nomenclature of Minerals: A Compilation of IMA Reports*. Ottawa: Mineralogical Association of Canada.

Nickel EH and Grice JD (1998) The IMA Commission on New Minerals and Mineral Names: procedures and guidelines on mineral nomenclature, 1998. *Canadian Mineralogist* 36: 1–14.

Strunz H and Nickel EH (2001) *Strunz Mineralogical Tables*, 9th edn. E. Stuttgart: Schweizerbart'sche Verlagsbuchhandlung (Nägele u. Obermiller).

Amphiboles

R A Howie, Royal Holloway, University of London, London, UK

The minerals of the amphibole group differ from the pyroxenes (*see* **Minerals:** Pyroxenes) in having a double-chain silicate structure and in having hydroxyl ions as an essential constituent. Their name is from the Greek *amphibolos* (ambiguous) in allusion to the great variety of compositions and appearances shown by this mineral group. All have a perfect {210} cleavage, and typically range from white to yellow, green and dark green, or blue, the coloured varieties having variable to strong pleochroism.

As in the pyroxene group, both orthorhombic and monoclinic amphiboles occur; their flexibility of ionic replacement is due to their structure. The composition of the simplest calcium-rich amphibole, tremolite, may be expressed by the formula $Ca_2Mg_5[Si_8O_{22}](OH)_2$, the Ca atoms occupying the largest positions of between six- and eight-fold coordination. The general chemical formula for an amphibole can be given as $A_{0-1}B_2C_5[T_8O_{22}](OH,F)_2$ where A represents the larger cations, B the cations in *M*4, C the cations in *M*1, *M*2, and *M*3, and T those in the T1 and T2 sites (Figures 1 and 2).

The four main substitutions seen in the amphibole structures are:

i. Monovalent ion (Na or K) for vacancy (□) at A.
ii. Monovalent (Na,K) for divalent Ca at B (*M*4 sites).
iii. Trivalent (Al) for tetravalent (Si) at T sites.
iv. Trivalent (Al,Fe^{3+}) for divalent (Mg, Fe^{2+}, etc) at C (*M*1, *M*2, *M*3 sites).

To maintain charge balance, it is necessary for more than one of these substitutions to take place, e.g., in pargasite, $NaCa_2(Mg, Fe^{2+})_4Al[Si_6Al_2O_{22}](OH, F)_2$, where the Na ion is in the A position (type 1 substitution), Al is in the *M* sites (type 4 substitution) and the excess charge is balanced by the substitution of 2Al in the T site (type 3 substitution), the maximum substitution allowed for Si. At the same time, there may be complete substitution between Mg and Fe^{2+} to give ferro-pargasite. A complete range of possible amphibole substitutions was published by Leake and others, and also Deer and others in 1997.

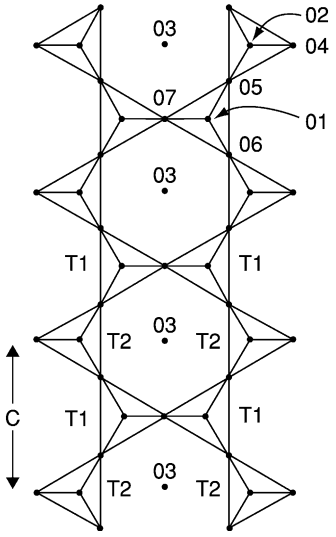


Figure 1 Double chain of linked $(Si,Al)O_4$ tetrahedra characteristic of all amphiboles.

There are, thus, four main groups with boundaries as follows:

i. Iron-magnesium-manganese-lithium amphiboles: $(Ca + Na)_B < 1.00$
ii. Calcic amphiboles: $(Ca + Na)_B \geq 1.00$ and $Na_B < 0.50$
iii. Sodic-calcic amphiboles: $(Ca + Na)_B \geq 1.$ and $0.50 \leq Na_B \leq 1.50$
iv. Sodic amphiboles: $(Na)_B \geq 1.50$

Because of excellent argon retention properties and the common incorporation of potassium in their structures, the amphiboles are particularly useful for K–Ar dating. Hornblende K–Ar ages have been used to date a wide variety of metamorphic and igneous rocks.

Leaving aside the lithium-bearing holmquistite, the two orthorhombic amphiboles are anthophyllite, $(Mg, Fe^{2+})_7[Si_8O_{22}](OH, F)_2$, and gedrite (Mg, Fe^{2+})

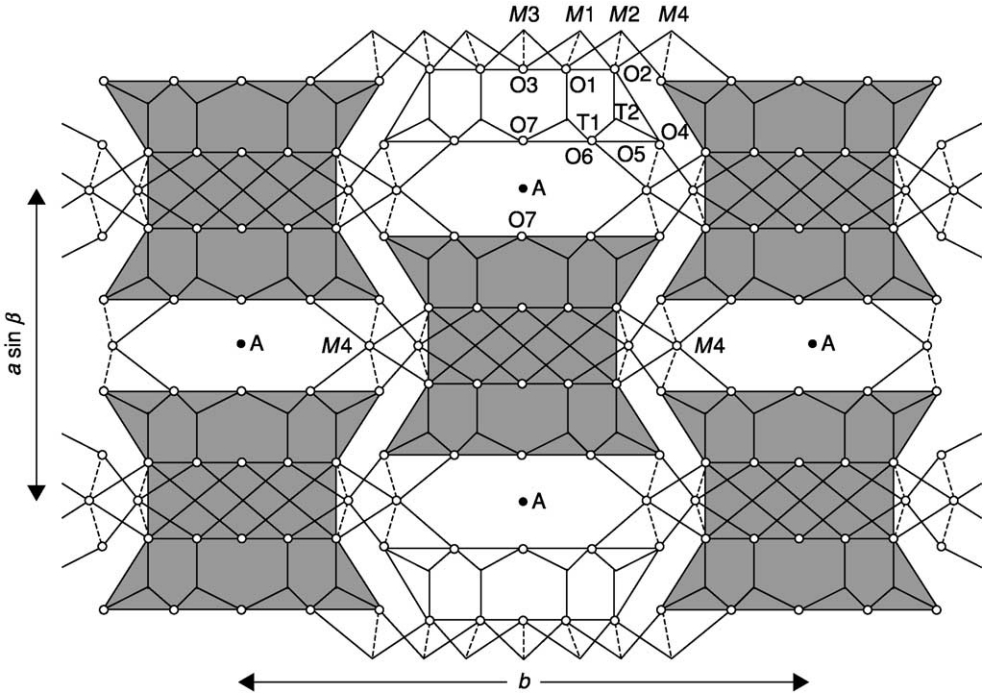


Figure 2 The crystal structure of an amphibole as viewed along *z*. Pairs of chains are linked by cations *M*1, *M*2, *M*3 and *M*4, and in some cases A.

$Al_2[Si_6Al_2O_{22}](OH,F)_2$. The replacement of Mg by Fe raises the refractive indices as does the substitution of Mg,Si by Al,Al; Mg-rich anthophyllites are optically negative, whereas the more Fe-rich anthophyllites and the gedrites are optically positive. In general, these orthorhombic amphiboles are virtually unknown in igneous rocks, but occur in a great variety of metamorphic and metasomatic rocks. The cummingtonite–grunerite series, $(Mg,Fe,Mn)[Si_8O_{22}](OH)_2$, are monoclinic and show characteristic multiple twinning on (100); cummingtonite is optically positive, but the more iron-rich grunerite is optically negative. Cummingtonite occurs in amphibolites derived by the regional metamorphism of basic igneous rocks and in hybrid rocks of intermediate composition. The more iron-rich (and sometimes manganese-rich) grunerites are typical of the banded iron formations of regional metamorphism, where they form a characteristic magnetite–grunerite–quartz association.

In the calcic amphiboles, the tremolite–ferro-actinolite series, $Ca_2(Mg,Fe^{2+})_5[Si_8O_{22}](OH,F)_2$, are colourless to yellow and green, and are essentially metamorphic minerals occurring in both contact and regionally metamorphosed rocks. In the metamorphism of siliceous dolomites, tremolite forms early on by reaction between dolomite and quartz:

$$\underset{\text{dolomite}}{5CaMg(CO_3)_2} + \underset{\text{quartz}}{8SiO_2} + H_2O \rightarrow$$

$$\underset{\text{tremolite}}{Ca_2Mg_5Si_8Oc\ \text{rocks}\ {}_{22}(OH)_2} + \underset{\text{calcite}}{3CaCO_3} + 7CO_2$$

Both tremolite and actinolite are characteristic minerals in low-grade regionally metamorphosed ultrabasic rocks, and actinolite is a common mineral in the greenschist facies.

Hornblende is used to describe a specific calcic amphibole, $Ca_2(Mg,Fe^{2+})_4(Al,Fe^{3+})[Si_7AlO_{22}](OH)_2$, but the continuous chemical variations towards pargasite (see above) and tschermakite, $Ca_2(Mg,Fe^{2+})_3(Fe^{3+}Al)_2[Si_6Al_2O_{22}](OH)_2$, as well as to Al-poor tremolite–ferro-actinolite tend to be called hornblendes in petrographic descriptions, when more appropriate names would pargasitic-, tschermakitic- or ferro-actinolitic-hornblende. This is illustrated in Figure 3, showing the typical chemical variations in analysed calcic amphiboles. The 'hornblendes' are typically pleochroic from pale green to yellow-brown or brown-green and have extinction angles in the range 12–34° (as compared with $\leq 45°$ for the clinopyroxenes). The hornblendes are typical minerals of intermediate plutonic rocks, and occur as products of primary crystallization of igneous rocks

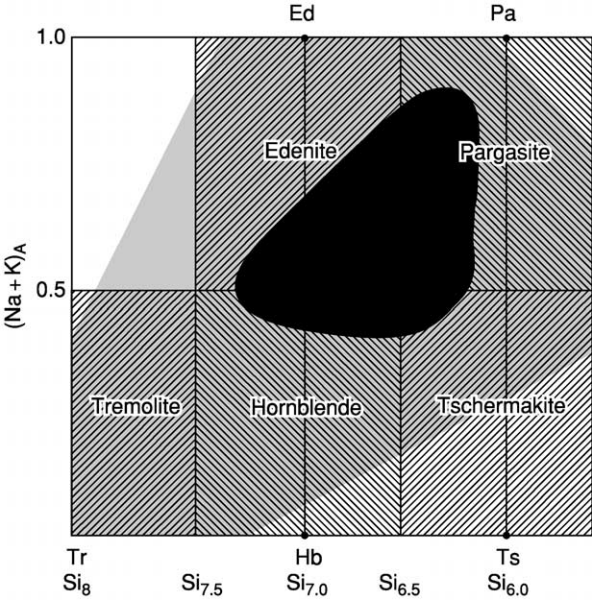


Figure 3 The chemical variations of calcic amphiboles expressed as the numbers of (Na + K) atoms in *A* sites and Si atoms per formula unit. Tr tremolite, Hb hornblende, Ed edenite, Pa pargasite, Ts tschermakite. The more densely stippled areas show the more commonly occurring compositions.

ranging from granites (*see* **Igneous Rocks:** Granite) and syenites to gabbros and ultrabasic rocks. They are neither as common or as abundant in volcanic rocks (due to the loss of volatiles from typical basalts and trachytes); they do, however, occur in a variety of andesites and dacites. Hornblende is also a dominant constituent in many regionally metamorphosed rocks, from the greenschist to the lower parts of the granulite facies.

Kaersutite, $NaCa_2(Mg,Fe^{2+})_4Ti[Si_6Al_2O_{22}](O,OH,F)_2$, is brown to reddish brown under the microscope and is characterized chemically by its high Ti content (TiO_2 5–10 wt.%, equivalent to 0.5–1.0 atom pfu). It is a typical constituent of alkali volcanic rocks, occurring as phenocrysts in trachybasalts, trachyandesites, trachytes, and alkali rhyolites.

Among the sodic amphiboles, glaucophane, in the glaucophane–riebeckite series, $Na_2(Mg,Fe^{2+})_3(Al,Fe^{3+})_2[Si_8O_{22}](OH,F)_2$, is the iron-poor end-member and characteristically shows blue–violet pleochroism in thin section, and may give a blue shade to its rocks, forming typical blueschists as the result of regional metamorphism of basaltic rocks. Riebeckite, the iron-rich equivalent, is an even deeper blue to almost black, and occurs in peralkaline granites, syenites, microgranites and in acid volcanic rocks. The eckermannite-arfvedsonite series, $NaNa_2(Mg,Fe^{2+},Mn)_4(Al,Fe^{3+})[Si_8O_{22}](OH,F)_2$, is even richer in sodium and ranges compositionally between both the Mg and Fe^{2+} and the Al and Fe^{3+}

end-members; they are generally associated with aegirine in peralkaline igneous rocks and have a bluish green to yellow-brown pleochroism.

See Also

Igneous Rocks: Granite. **Minerals:** Definition and Classification; Other Silicates; Pyroxenes. **Rocks and Their Classification**.

Further Reading

Deer WA, Howie RA, and Zussman J (1997) *Rock-Forming Minerals:* Vol. 2B. *Double-Chain Silicates*, 2nd edn. xii + 764 pp. London: Geological Society.

Leake BE, Woolley AR, Arps CES, *et al.* (1997) Nomenclature of Amphiboles: Report of the Subcommittee on Amphiboles of the International Mineralogical Association Commission on New Minerals and Mineral Names. *Mineral Magazine* 61: 295–321.

Arsenates

K Hudson-Edwards, University of London, London, UK

Crystal Structure

The arsenates are a subclass of the phosphate mineral class, which has a basic chemical unit of $[AO_4]$ with a negative three charge (-3) and a tetrahedral symmetry. In the case of the arsenates, the 'A' in the tetrahedron is the element arsenic (As). These tetrahedra are generally linked to one or more metal–oxygen, –hydroxide, and/or –water octahedra, which in turn are linked by edge- and corner-sharing arrangements. Other weaker bonds exist in the arsenate structures, such as H bonds between the octahedra and arsenate tetrahedra atoms. Some of the arsenates are sheet-like, with intersheet spaces occupied with, and bonded by, cations, halogens, or H_2O molecules. A good example of an arsenate structure is that of the mineral scorodite $[FeAsO_4 \cdot 2H_2O]$, which is composed of Fe(III)–O octahedra that share the oxygen atoms of arsenate tetrahedra. The arsenate tetrahedra are also weakly linked to the Fe(III)–O octahedra through H bonding.

Several dimorphous (same chemical formula, different structure) arsenates exist (e.g., roselite, monoclinic–beta-roselite, triclinic; symplesite, triclinic–parasymplesite, monoclinic).

Chemistry and Nomenclature

There are well over 180 arsenates. They form a diverse subclass of minerals, exhibiting widely varying chemistries and structures, and comprising a number of groups and solid-solution series. Table 1 lists examples of the most common arsenate groups, and their crystal system(s) and point group(s).

A common feature of the arsenates is their incorporation of water or hydroxide into their structure. The amount of water incorporated depends on temperature, vapor pressure, and crystal structure. Some of the more common hydrous and complex arsenates are summarized in Table 2. Those arsenates that have a phyllosilicate-like, layered structure, with interlayer water molecules, are often incorrectly termed 'micas' (e.g., the autunite arsenates are often referred to as 'uranyl micas').

Many of the arsenates form solid-solution series, with the As in the arsenate tetrahedron being replaced by P or V. In fact, many of the groups listed in Table 1 also include phosphates and vanadates (only the arsenate members of the groups are shown in Table 1). For example, the apatite group contains arsenates (e.g., mimetite $Pb_5(AsO_4)_3Cl$), phosphates (e.g., pyromorphite $Pb_5(PO_4)_3Cl$), and vanadates (e.g. vanadinite $Pb_5(VO_4)_3Cl$), and the mixite group contains arsenates (e.g., mixite, $BiCu_6(AsO_4)_3(OH)_6 \cdot 3H_2O$) and phosphate minerals (e.g., petersite $(Ca,Fe,Y,Ce)Cu_6(PO_4)_3(OH)_6 \cdot 3H_2O$). Solid solutions also exist between members of different arsenate groups, where metal cations substitute for one another (the 'X' positions in Table 1). In some cases, the solid solutions are complete, with no miscibility gaps (e.g., annabergite–erythrite), but in others the series are incomplete (e.g., annabergite-köttigite). At least 75 solid solution series involving arsenates have been described.

Physical Properties and Stability

Probably because of their wide-ranging chemistries, the arsenates exhibit a wide range of physical properties. Many of the arsenates are green or a variety of green (bluish, bright, bright apple, emerald, grass, grey, olive, pale, yellowish), but a wide number of colours have been reported, including blue, brown, grey, pink, orange, purple, red, yellow, white, and shades in-between. The colour is often dictated by the incorporation of transition metals such as Ni, Cu, U, and Co into the mineral structure. Similarly,

Table 1 Examples of common arsenate groups

Group	*General formula*	*Examples*	*Crystal system(s)*	*Point group(s)*
Adamite	$X(AsO_4)(OH)$	Adamite	Triclinic	bar 1
	X = Cu,Zn,Mn,Zn	Olivenite	Orthorhombic	2/m 2/m
		Paradamite		2/m
Adelite	$XYAsO_4OH$	Austenite	Orthorhombic	2/m 2/m
	X = Ca,Pb	Conichalcite		2/m
	Y = Cu,Zn	Duftite		2 2 2
Apatite-pyromorphite	$X_5(AsO_4)_3(F,Cl,OH)$	Clinomimetite	Hexagonal	6/m
	X = Ca,(Ba,Ca,Pb),(Ca,Pb),(Ca,Sr),Pb	Mimetite		
		Hedyphane		
Arthurite	$XFe_2(AsO_4)_2(O,OH)_2 \cdot 2H_2O$	Arthurite	Monoclinic	2/m
	X = Cu,Fe,Mn,Zn	Cobaltarthurite		
		Ojuelaite		
Autunite	$X(UO_2)_2(AsO_4)_2 \cdot 8–12H_2O$	Heinrichite	Monoclinic	2/m 2/m
	X = H,Mg,Fe,Cu,Ca,Ba,Mn,1/2(HAl)	Kahlerite	tetragonal	4/m 2/m
		Uranospinite		2/m
		Zuenerite		bar 4
Beudantite	$XY_3(AsO_4PO_4)SO_4(OH)_6$	Beudantite	Trigonal	
	X = Ca,Ba,Ce,Pb,Sr,H_3O	Gallobeudantite		
	Y = Fe,Al,Ga	Hidalgoite		
Crandallite	$XAl_3(AsO_4)_2(OH,F)_{1–6} \cdot 1–15H_2O$	Arsenocrandallite	Trigonal	3m 3m
	X = (Ca,Sr),(Sr,Ca,Ba),Ba,Pb,(Ce,La),Fe	Philipsbornite		3m 3
				2/m
Meta-autunite	$X(UO_2)_{1–2}(AsO_4)_2 \cdot 4–8H_2O$	Abernathyite	Tetragonal	4/m 2/m
	X = Mg,Fe,Co,Cu,Ca,Ba,K,Na,Zn	Metalodevite		2/m
		Metanovacekite		
Mixite	$XCu_6(AsO_4)_3(OH)_6 \cdot 3H_2O$	Agardite	Hexagonal	
	X = Bi,Al,Ca,Ce,La,Y,Nd,Th	Goudeyite		
		Mixite		
Pseuomalachite	$X_{3–5}(AsO_4)_2(OH)_{3–4}$	Cornwallite	Triclinic, monoclinic	1
	X = Cu,Mn	Arsenoclasite		2/m
				2 2 2
Roselite	$Ca_2X(AsO_4)2 \cdot 2H_2O$	Brandtite	Monoclinic	2/m
	X = Co,Mn,Mg,Zn	Roselite		
		Zincroselite		
Variscite	$XAsO4 \cdot 2H_2O$	Mansfieldite	Orthorhombic	2/m 2/m
	X = Al,Fe,In	Scorodite		2/m
		Yanomamite		
Vivianite	$X_3(AsO_4)_2 \cdot 8H_2O$	Annabergite	Monoclinic	2/m
	X = Co,Fe,Mg,Ni,Zn	Erythrite		
		Kottingite		
		Hörnesite		
		Parasymplesite		

mineral streak is often green or a shade of green (bluish, olive, pale), but grey, pale blue, brown red and yellow, orange yellow, and white are known.

Arsenate crystals tend to range from transparent to translucent, and some opaque varieties have also been documented. Luster is adamantine, dull (in massive or powdery varieties), greasy, pearly, resinous, silky, sub-adamantine, or vitreous. Hardness is generally low, ranging from 1.5 up to 4.5, and specific gravity tends to be slightly below average to heavy for translucent minerals (2.4 to 6.7).

A huge variety of crystal habits have been described for the arsenates. Examples include acicular, botryoidal, fibrous, foliated, micaceous, prismatic, radiating, striated blade and tabular single crystals, aggregates, clusters, or masses (Figure 1). Crystal faces exhibit diverse termination surfaces: domes with bipyramids, triangles, double triangles, pinacoids, and wedges with complex facets. Several varieties are twinned. Most commonly, however, the minerals occur as crusts, globular aggregates, nodules, or powdery masses with little crystal shape. Cleavage ranges from absent to poor or perfect in one or two directions. Fracture is described as conchoidal, fibrous, flaky, or lamellar.

The arsenate adamite ($Zn_2AsO_4(OH)$) is well known for its fluorescent properties. Arsenates belonging to the autinite group (e.g., meta-zeunerite, walpurgite, zeunerite) are radioactive due to the uranium in their structures.

Table 2 Examples of hydrous and complex arsenates

Mineral	*Formula*	*Crystal system*	*Point group*
Adamite	$Zn_2AsO_4(OH)$	Orthorhombic	2/m 2/m 2/m
Arsentsumebite	$Pb_2Cu(AsO_4)(SO_4)OH$	Monoclinic	
Bayldonite	$Cu_3Pb(AsO_4)_2 \cdot H_2O$	Monoclinic	
Carminite	$PbFe_2(AsO_4)_2(OH)_2$	Orthorhombic	2/m 2/m 2/m
Clinoclase	$Cu_3AsO_4(OH)_3$	Monoclinic	2/m
Cuproadamite	$(Cu,Zn)_2(AsO_4)(OH)$	Orthorhombic	2/m 2/m 2/m
Kankite	$Fe(AsO_4)_3 \cdot 5H_2O$	Monoclinic	
Legrandite	$Zn_2AsO_4(OH) \cdot H_2O$	Monoclinic	2/m
Liroconite	$Cu_2Al(AsO_4)(OH)_4 \cdot 4H_2O$	Monoclinic	2/m
Olivenite	$Cu_2AsO_4(OH)$	Orthorhombic	2/m 2/m 2/m
Paradamite	$Zn_2AsO_4(OH)$	Triclinic	bar 1
Pharmacolite	$Ca(AsO_3OH) \cdot 2H_2O$	Monoclinic	
Pharmacosiderite	$KFe_4(AsO_4)_3(OH)_4 \cdot 6–7H_2O$	Cubic	bar 4 3m
Picropharmacolite	$H_2Ca_4Mg(AsO_4)_4 \cdot 11H_2O$	Triclinic	bar 1
Symplesite	$Fe_3(AsO_4)_2 \cdot 8H_2O$	Triclinic	
Talmessite	$Ca_2Mg(AsO_4) \cdot 2H_2O$	Triclinic	bar 1
Tilasite	$CaMg(AsO_4)F$	Monoclinic	2/m
Tsumcorite	$Pb(Zn,Fe)_2(AsO_4)_2 \cdot H_2O$	Monoclinic	
Walpurgite	$(BiO)_4UO_2(AsO_4)_2 \cdot H_2O$	Trigonal	

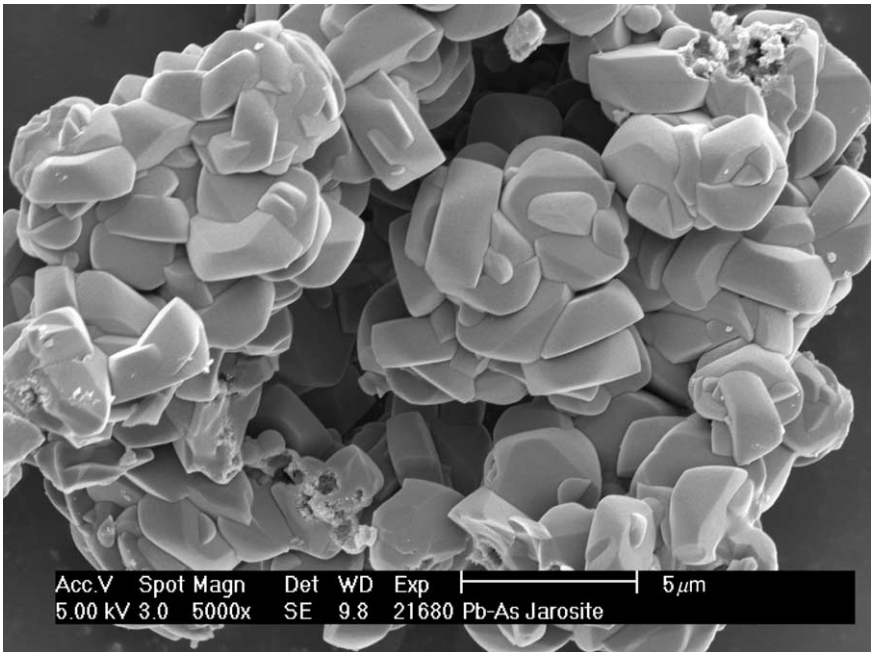


Figure 1 Photomicrograph of synthetic beudantite ($PbFe_3AsO_4SO_4(OH)_6$) crystals, exhibiting tabular cluster crystal shape. Image courtesy of AML Smith.

Solubility and Alteration

The solubilities of the arsenates are dependent on their degree of crystallinity and water content. Many have high log K_{sp} values under standard conditions, and are thus quite soluble (e.g., ferrarisite, guerinite, Table 3). Others, such as annabergite and scorodite, are less soluble and, as a result, tend to control the distribution of arsenic in soils and other environmental media. Scorodite is one of the most common arsenates in soils, but is only stable under acid (pH 4) oxidizing conditions. Under more neutral pH or reducing conditions, scorodite dissolves incongruently to form Fe oxides and soluble As species; amorphous scorodite is thought to be more soluble than crystalline forms (Table 3). Similarly, the arsenate beudantite ($PbFe_3AsO_4SO_4(OH)_6$) dissolves incongruently in acid and

Table 3 Solubilities of selected arsenates at 25°C, 1 bar

Mineral	*Dissociation reaction*	*Log K_{sp}*
Annabergite	$Ni_3(AsO_4)_2 \cdot 8H_2O = 3Ni^{2+} + 2AsO_4^{3-} + 8H_2O$	−28.38
TriCa arsenate	$Ca_3(AsO_4)_2 \cdot 4H_2O = 3Ca^{2+} + 2AsO_4^{3-} + 4H_2O$	−21.257
Calcium arsenate hydrate	$Ca_4(OH)_2(AsO_4)_2 \cdot 4H_2O = 4Ca^{2+} + 2H_3AsO_4 + 6H_2O - 8H^+$	40.2
Ferrarisite	$Ca_5(HAsO_4)_2(AsO_4)2 \cdot 9H_2O = 5Ca^{2+} + 4H_3AsO_4 + 9H_2O - 10\ H^+$	27.75
Guerinite	$Ca_5(HAsO_4)_2(AsO_4)2 \cdot 9H_2O = 5Ca^{2+} + 4H_3AsO_4 + 9H_2O - 10\ H^+$	28.55
Scorodite, amorphous	$FeAsO_4 \cdot 2H_2O = Fe^{3+} + AsO_4^{3-} + 2H_2O$	−22.89
Scorodite, crystalline	$FeAsO_4 \cdot 2H_2O = Fe^{3+} + AsO_4^{3-} + 2H_2O$	−25.89

alkali solutions, yielding As(V) ions, α-FeOOH, $Fe(OH)_3$, and $PbSO_4$.

One of the arsenates, meta-zeunerite, forms by dehydration and pseudomorphing of the parent arsenate zeunerite.

Occurrence

Most of the arsenates are rare to very rare. They are prized by mineral collectors because of their range of spectacular colours, lusters, and crystal habits. The arsenate balydonite, found in Cornwall, England, has even been used to craft cabochons (convex gems and beads) and cabinets.

Many arsenates are found in the oxidation zone of sulphide ore deposits or in oxidized mine waste, where they occur as discrete precipitates or coatings on other mineral grains. They are associated with sulphide minerals such as arsenopyrite, chalcopyrite, sphalerite, and galena, and with other arsenates and secondary ore minerals such as iron oxides, azurite, and malachite. The most common of these arsenates include annabergite, beudantite, conichalcite, hörnesite, parasymplesite, rauenthalite, and, particularly, scorodite. Although not common, two of the most famous of the oxidation zone arsenates are erythrite and annabergite, known by miners as 'cobalt bloom' and 'nickel bloom', respectively. These are used as indicator minerals of Co and Ni sulphide ores, and as ores of Co and Ni themselves. Several other arsenates are also minor ores of metals; examples are agardite (REE), beudantite (Pb), chalcopyllite (Cu), conichalcite (Cu), köttigite (Zn), mimetite (Pb), and walpurgite (U and Bi). These mine-related arsenates have been found at well-known mines in Devon and Cornwall, England; Mexico (Mapimi); Cobalt, Ontario, Canada; Sweden (Långban, Warmland); Nambia (Tsumeb); Alsace, France; Arizona; Romania; Russia; and Zaire. The arsenate nealite ($Pb_4Fe(AsO_4)_2Cl_4$) is also mine-related, but has been found in Greek and Roman slags in Lavrio, Greece, rather than outcrops, leading to some controversy over its classification as a mineral, since by definition, minerals should be natural, rather than man-made.

The arsenates scorodite and beudantite have been reported frequently in sulphide tailings. In uranium tailings, Ca–arsenates occur as coprecipitates with gypsum and as amorphous precipitates. $Ca_3(AsO_4)$ is the arsenates generally thought to control As solubility in aqueous systems where Ca^{2+} is present, but $Ca_4(OH)_2(AsO_4)_2 \cdot 4H_2O$ and $CaHAsO_4 \cdot H_2O$ have also been suggested to occur in tailings with high Ca/As ratios and alkaline pH (10) conditions, and low Ca/As ratios (0.88–1.00) and slightly acid pH (5.76–6.22) conditions, respectively.

Arsenic occurs most commonly in soil waters as the complexes $H_2AsO_4^-$ (low pH) and $HAsO_4^{2-}$ (higher pH). The species $H_3AsO_4^0$ and AsO_4^{3-} occur in extremely acidic and alkaline conditions, respectively. Sources of these complexes are many and varied, and include pesticides, mining, munitions waste, wood preservatives, and tannery wastes. The arsenate complex is sorbed onto common soil minerals, such as Fe oxides and clays, or forms discrete arsenates, many of which are listed in Tables 1 and 2. The type of arsenate formed in soils depends on the availability of other cations (e.g., Ca, K, Mg, H, Pb) in the soils. Many of the hydrous arsenates (Tables 1, 2) are found in soils (e.g., carminite, kankite, pharmacosiderite, talmessite, tilasite, and members of the vivianite group).

Arsenates such as scorodite have also been reported to occur as crusts in hot spring deposits. Due to its insolubility, scorodite is also used in metallurgy for the disposal of arsenic wastes from metallurgical effluents and flue dusts.

See Also

Economic Geology. **Environmental Geochemistry**. **Mineral Deposits and Their Genesis**. **Minerals:** Vanadates.

Further Reading

Bothe JV and Brown PW (1999) As immobilization by calcium arsenate formation. *Environmental Science Technology* 33: 3806–3811.

Bothe JV and Brown PW (1999) The stabilities of calcium arsenates. *Journal of Hazardous Materials* 69: 197–207.

Donahue R and Hendry MJ (2003) Geochemistry of arsenic in uranium mill tailings, Saskatchewan, Canada. *Applied Geochemistry* 18: 1733–1750.

Dove PS and Rimstidt JD (1985) The solubility and stability of scorodite, $FeAsO_4 \cdot 2H_2O$. *American Mineralogist* 70: 838–844.

Frost RL (2004) An infrared and Raman spectroscopic study of the uranyl micas. *Spectrochimica Acta A* 60: 1469–1480.

Frost RL, Kloprogge R, Weier ML, *et al.* (2003) Raman spectroscopy of selected arsenates—implications for soil remediation. *Spectrochimica Acta Part A* 59: 2241–2246.

Frost RL, Weier ML, Martents W, Kloprogge JT, and Ding Z (2003) Thermal decomposition of the vivianite arsenates—implications for soil remediation. *Thermochimica Acta* 403: 237–249.

Kitahama K, Kiriyawa R, and Baba Y (1975) Refinement of the crystal structure of scorodite. *Acta Crystallographica Section B* 31: 322–332.

Krause E and Ettel VA (1988) Solubility and stability of scorodite, $FeAsO_4 \cdot 2H_2O$: New data and further discussion. *American Mineralogist* 73: 850–854.

Krause E and Ettel VA (1989) Solubilities and stabilities of ferric arsenate compounds. *Hydrometallurgy* 22: 311–337.

Langmuir D and Mahoney J (1998) Appendix 3 McClean Lake Project, JEB Tailings Management Facility, Construction License Additional Information, Geochemistry, Sub-Appendix F Thermodynamic Data for Selected Species and Slids of Arsenic and Nickel. Cogema Resources Inc.

Mahoney J (2002) *The Corrected Solubility Product of Scorodite and Its Application to Arsenic Behaviour in Buried Mine Tailings*. GSA Program with Abstracts, Denver Annual Meeting, October 27–30, 2002, Paper 84-17.

Parkhurst DL and Appelo CAJ (1999) User's guide to PHREEQC—A computer program for speciation, batch reaction, 1D-transport and inverse geochemical calculations. US Geological Survey Water Resources Investigations Report 99-4259.

Rimstidt JD, Chermak JA, and Gagen PA (1994) Rates of reaction of galena, sphalerite, chalcopyrite, and arsenopyrite with Fe^{3+} in acidic solutions. In: Alpers CN and Blowes DW (eds.) *Environmental Geochemistry of Sulfide Oxidation*. p. 2–13. Washington, DC: American Chemical Society.

Robins RG (1981) The solubility of metal arsenates. *Metallurgical Transactions B* 12: 103–109.

Robins RG (1987) Solubility and stability of scorodite, $FeAsSO_4 \cdot 2H_2O$. *American Mineralogist* 72: 842–844.

Smith AML (2004) Mechanisms and products of the breakdown of contaminant element-bearing jarosites. PhD thesis, Birkbeck, University of London.

Borates

C Helvaci, Dokuz Eylül Üniversitesi, İzmir, Turkey

Introduction

Borate has a very long history. Derived from the Persian burah (boorak), borax was known to the Babylonians, who brought it from the Himalayas some 4000 years ago for use in the manufacture of rings, amulets, and bracelets. The Egyptians used borax in mummifying, and around AD 300 the Chinese were familiar with borax glazes, as were the Arabs three centuries later. Borax was first brought to Europe in the thirteenth century, presumably by Marco Polo, and it has been supplied since by traders from Tibet and Kashmir.

By the 1770s the French had developed a source of tincal, the old name for crude borax, in Purbet Province, India, and at about the same time natural boric acid (sassolite) was discovered in the hot springs in the Maremma region of Tuscany, Italy. The middle of the nineteenth century was a particularly active time for the discovery and commercial development of borate deposits. In particular, Chile started to mine the borate resources of the Salar de Ascotan in 1852, and within a few years output accounted for a quarter of the world's annual supply of 16 000 tonnes. In 1856 John Veatch discovered borax in Clear Lake, Lake County, California, and this led eventually to the formation of the California Borax Company in 1864 and to California's dominance of the borate industry. In Turkey modern borate mining began in 1865 when the Compagnie Industrialle des Mazures mined borates from the Aziziye Mine near Susurluk and shipped the ore to France for processing. Demand encouraged the exploitation of large-scale deposits in Turkey and the USA and overwhelmed more modest producers.

Borates are among the most interesting of the world's industrial minerals; they were used first in precious-metal working and later in ceramics. They are an unusually large group of minerals, but the number of commercially important borates is limited, and their chemistry and crystal structure are both unusual and complex. There are only a few large deposits, although there are many non-commercial occurrences in other rocks and brines. The accounts of the early exploration, mining, and processing of

borates are fascinating, because the remote locations of the deposits often led to unusual difficulties and hardships in recovering the desired products. These varied from workers wading into Himalayan lakes to harvest the 'floor' and then transporting the borax in saddlebags on sheep across the Himalayas to the markets, to the 'Dante's Inferno' of the Larderello boric-acid fumaroles, and the colourful 20 mule teams of the western USA. Such operations transformed borates from expensive minor minerals into the large-volume industrial commodities they are today. Boron's chemistry and reactivity are also fascinating because it forms oxygen compounds in an essentially unending variety of simple to exceedingly complex molecules. Determining the crystal structures of these compounds has given rise to a separate subfield of crystallography. The boron isotopes ^{10}B and ^{11}B have very different reactivities during both physical and chemical changes, which has allowed them to be used to understand many geological and other events, again forming a specialized field in geology.

A borate is defined as any compound that contains or supplies boric oxide (B_2O_3). A large number of minerals contain boric oxide, but the three that are most important from a worldwide commercial standpoint are borax, ulexite, and colemanite. These are produced in a limited number of countries (Figure 1), and production is dominated by the USA and Turkey, which together furnish about 90% of the world's borate supplies. Production in the USA is concentrated in the Mojave Desert of California: borax and kernite are mined from the large deposit at Boron. Borate-containing brines are pumped from Searles Lake, and a limited amount of colemanite is mined in Death Valley. Turkish production is controlled by Eti Maden (Eti mine), the national mining enterprise, which supplies most of the commercially traded ulexite and colemanite from mines in the Bigadiç and Emet districts, together with borax from the huge deposit at Kyrka.

Borate minerals have been used in many ways since at least the eighth century, when they were used primarily as a flux for assaying and refining gold and silver. Their valuable properties and relative rarity soon stimulated an international trade in borates. Borates were traded at relatively high prices for highly specialized applications until the late nineteenth century. At that time they were being used in medicines, food preservatives, ceramic glazes, and as metal fluxes. Borates are often defined and sold according to their boric oxide or B_2O_3 content, and most statistical data are listed in tonnes of B_2O_3. Borax pentahydrate and boric acid are the most commonly traded commodities. Boric acid plants are operated by all the major borate producers. Glass-fibre insulation is the major end use in the USA, followed by textile glass fibre and borosilicate glass, detergents, and ceramics. Detergents continue to be a major end use in Europe.

Geology and Mineralogy

Over 150 boron-bearing minerals have been identified, the most common being sodium, calcium, or magnesium salts. Table 1 gives a more complete list of the common borates. However, just four minerals – borax, ulexite, colemanite, and datolite – are commercially significant today. Borax or tincal, a natural sodium borate decahydrate, is the major commercial source of boron, with major supplies coming from the USA, Argentina, and Turkey. The principal commercial mixed sodium–calcium borate, ulexite, is produced in Turkey and several countries in South America, whereas large-scale production of the main calcium borate, colemanite, is restricted to Turkey.

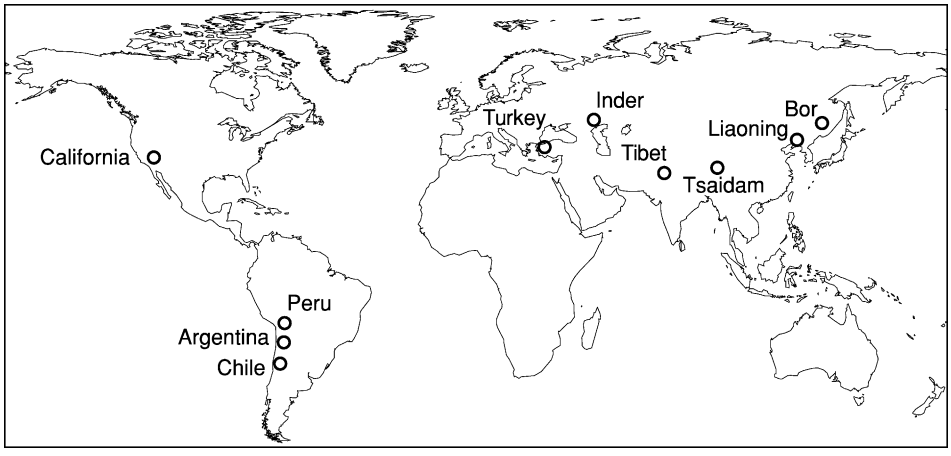


Figure 1 Current world resources of borates.

Table 1 Composition of principal boron minerals

Mineral	*Empirical formula*	*B_2O_3 Content (Weight %)*
Sassolite	$B(OH)_3$ or $B_2O_3 \cdot 3H_2O$	56.4
Borax (tincal)	$Na_2B_4O_7 \cdot 10H_2O$	36.5
Tincalconite	$Na_2B_4O_7 \cdot 5H_2O$	48.8
Kernite	$Na_2B_4O_7 \cdot 4H_2O$	51.0
Ulexite	$NaCaB_5O_9 \cdot 8H_2O$	43.0
Probertite	$NaCaB_5O_9 \cdot 5H_2O$	49.6
Priceite (Pandermite)	$Ca_4B_{10}O_{19} \cdot 7H_2O$	49.8
Inyoite	$Ca_2B_6O_{11} \cdot 13H_2O$	37.6
Meyerhofferite	$Ca_2B_6O_{11} \cdot 7H_2O$	46.7
Colemanite	$Ca_2B_6O_{11} \cdot 5H_2O$	50.8
Hydroboracite	$CaMgB_6O_{11} \cdot 6H_2O$	50.5
Inderborite	$CaMgB_6O_{11} \cdot 11H_2O$	41.5
Kurnakovite	$Mg_2B_6O_{11} \cdot 15H_2O$	37.3
Inderite	$Mg_2B_6O_{11} \cdot 15H_2O$	37.3
Szaibelyite (ascharite)	$Mg_2B_2O_5 \cdot H_2O$	41.4
Suanite	$Mg_2B_2O_5$	46.3
Kotoite	$Mg_3B_2O_6$	36.5
Pinnoite	$MgB_2O_4 \cdot 3H_2O$	42.5
Boracite (strassfurite)	$Mg_3B_7O_{13}Cl$	62.2
Datolite	$Ca_2B_2Si_2O_9 \cdot H_2O$	21.8
Cahnite	$Ca_2AsBO_6 \cdot 2H_2O$	11.7
Danburite	$CaB_2Si_2A_8$	28.3
Howlite	$Ca_4Si_2B_{10}O_{23} \cdot 5H_2O$	44.5
Vonsenite (paigeite)	$(Fe, Mg)_2FeBO_5$	10.3
Ludwigite	$(FeMg)_4Fe_2B_2O_7$	17.8
Tunnellite	$SrB_6O_{10} \cdot 4H_2O$	52.9

Production of datolite, a silicate mineral, is confined to Russia.

Boron minerals may be divided for convenience into three broad groups according to their origin and geological environments: first, skarn minerals related to intrusives (mainly silicates and iron oxides); second, magnesium oxides related to marine sediments; and, third, hydrated sodium and calcium borates related to continental sediments and volcanic activity (Figure 2).

In the first and second groups are the Russian and some of the Chinese sources, the major minerals being datolite and szaibelyite. In the third group are borax, kernite, colemanite, and ulexite, which provide the sources for most of the production from Turkey, and Argentina (Table 2). Minerals occurring in Turkey and the USA are shown in Table 3.

Table 2 Commercial borate minerals

Minerals	*Emprical formula*	*B_2O_3 content (weight %)*
Colemanite	$Ca_2B_6O_{11} \cdot 5H_2O$	50.8
Ulexite	$NaCaB_5O_9 \cdot 8H_2O$	43.0
Borax	$Na_2B_4O_7 \cdot 10H_2O$	35.5
Kernite	$Na_2B_{14}O_7 \cdot 4H_2O$	51.0
Pandermite	$Ca_4B_{10}O_9 \cdot 7H_2O$	49.8
Hydroboracite	$CaMgB_6O_{11} \cdot 6H_2O$	50.5

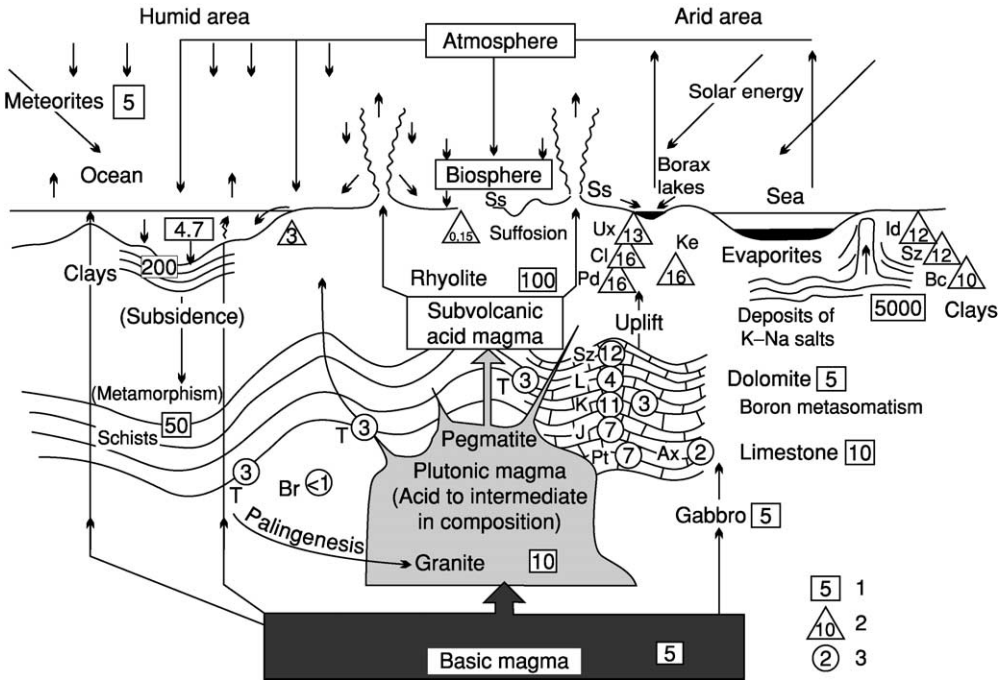


Figure 2 Scheme for the cycle and concentration of boron. The geochemical abundance of boron in rocks is cited according to Goldschmidt, Landergren and Harder; 1, average content of boron in rocks (g tonne^{-1}); 2, exogenic concentration of boron (%); 3, endogenic concentration of boron (%); Ss, sassolite; Ux, ulexite; Cl, colemanite; Pd, pandermite; Ke, kernite; In, inderite; Bc, boracite; Sz, szaibelyite; L, ludwigite; K, kotoite; J, jimboite ($Mn_3(BO_3)_2$); Dt, datolite; Ax, axinite; T, tourmaline; Br, braunite.

Table 3 Borate minerals occurring in Turkish and Argentine Miocene borate deposits

			Turkish borate deposits					*Argentine borate deposits*		
	Mineral	*Formula*	*Bigadiç*	*Sultançayir*	*Kestelek*	*Emet*	*Kirka*	*Tincalayu*	*Sijes*	*Loma Blanca*
Ca-borates										
	Inyoite	$Ca_2B_6O_{11} \cdot 13H_2O$	+	−	−	−	+	+	+	+
	Meyerhofferite	$Ca_2B_6O_{11} \cdot 7H_2O$	+	−	−	+	+	−	+	−
	Colemanite	$Ca_2B_6O_{11} \cdot 5H_2O$	+	+	+	+	+	−	+	+
	Tertschite	$Ca_4B_{10}O_{19} \cdot 20H_2O$	+	−	−	−	−	−	−	−
	Pandermite (priceite)	$Ca_4B_{10}O_{19} \cdot 7H_2O$	+	+	−	−	+	−	−	−
	Nobleite	$CaB_6O_{10} \cdot 4H_2O$	−	−	−	−	−	−	+	−
	Gowerite	$CaB_6O_{10} \cdot 5H_2O$	−	−	−	−	−	−	+	−
	Ginorite	$Ca_2B_{14}O_{23} \cdot 8H_2O$	−	−	−	−	−	+	−	−
Ca/Na-borates										
	Ulexite	$NaCaB_5O_9 \cdot 8H_2O$	+	−	+	+	+	+	+	+
	Probertite	$NaCaB_5O_9 \cdot 5H_2O$	+	−	+	−	−	+	+	−
Na-borates										
	Borax	$Na_2B_4O_7 \cdot 10H_2O$	−	−	−	−	+	+	−	+
	Tincalconite	$Na_2B_4O_7 \cdot 5H_2O$	−	−	−	−	+	+	−	+
	Kernite	$Na_2B_4O_7 \cdot 4H_2O$	−	−	−	−	+	+	−	−
	Ezcurrite	$Na_4B_{10}O_{17} \cdot 7H_2O$	−	−	−	−	−	+	−	−
	Ameghinite	$NaB_3O_5 \cdot 2H_2O$	−	−	−	−	−	+	−	−
Other borates										
(Mg)	Hydroboracite	$CaMgB_6O_{11} \cdot 6H_2O$	+	−	+	+	+	−	+	−
(Mg)	Inderborite	$CaMgB_6O_{11} \cdot 11H_2O$	−	−	−	−	+	−	+	−
(Mg)	Inderite	$Mg_2B_6O_{11} \cdot 15H_2O$	−	−	−	−	+	+	−	−
(Mg)	Kurnakovite	$Mg_2B_6O_{11} \cdot 15H_2O$	−	−	−	−	+	+	−	−
(Mg)	Rivadavite	$Na_6MgB_{24}O_{40} \cdot 22H_2O$	+	−	−	−	−	+	−	−
(Mg)	Mcallisterite	$Mg_2B_{12}O_{20} \cdot 15H_2O$	−	−	−	−	−	+	−	−
(Sr)	Tunellite	$SrB_6O_{10} \cdot 4H_2O$	+	−	−	+	+	−	−	−
(Sr)	Veatchite-A	$Sr_4B_{22}O_{37} \cdot 7H_2O$	−	−	−	+	−	−	−	−
(As)	Teruggite	$Ca_4MgAs_2B_{12}O_{28} \cdot 20H_2O$	−	−	−	+	−	−	−	+
(As)	Cahnite	$Ca_2BAsO_6 \cdot 2H_2O$	−	−	−	+	−	−	−	−
(Si)	Howlite	$Ca_4Si_2B_{10}O_{23} \cdot 5H_2O$	+	+	−	−	−	−	−	−
(Si)	Bakerite	$Ca_8B_{10}Si_6O_{35} \cdot 5H_2O$	−	+	−	−	−	−	−	−
(Si)	Searlesite	$NaBSi_2O_6 \cdot H_2O$	−	−	−	−	−	+	−	−
(Na–Mg)	Aristarainite	$Na_2MgB_{12}O_{20} \cdot 8H_2O$	−	−	−	−	−	+	−	−
(Ca–Sr)	Estroncioginorite	$(Sr,Ca)_2^-B_{10}O_{17} \cdot 7H_2O$	−	−	−	−	−	+	−	−

+, present; −, not present.

Borax is by far the most important mineral for the borate industry. It crushes freely and dissolves readily in water; its solubility and rate of solution increase with water temperature. Borax in large tonnages is present in the deposits at Boron (California), Kyrka (Turkey), and Tincalayu (Argentina). Kernite is present in minor amounts at Kyrka and Tincalayu, but it makes up about a third of the total reserve at Boron and has a higher B_2O_3 content than borax.

Colemanite is the calcium-bearing borate preferred by the non-sodium fibreglass industry. It has low solubility in water, although it dissolves readily in acid. Some colemanite is used in European chemical plants to produce boric acid because the supply from Turkey provides B_2O_3 at the lowest cost. Turkey is the world's major source of high-grade colemanite. The USA has important reserves in the Death Valley area, but only limited amounts are produced there at this time. Colemanite is not known to occur in major deposits outside Turkey and North America, although the higher hydrate, inyoite, is mined on a limited scale in Argentina.

Ulexite is the usual borate found on or near the surface, in playa lakes and marshes of Holocene to Quaternary age throughout the world, where it occurs as soft, often damp, masses of fibrous crystals. These 'cotton balls' or 'papas' are collected in large salars in South America and China. Ulexite of Neogene age is mined in Turkey and occurs at Boron and Death Valley in the USA. It is well lithified and, therefore, is hard, dense, and commonly well-bedded.

Szaibelyite (ascharite) is a major source of both Chinese and Russian borate. It is a magnesium borate and, like colemanite, has low solubility in water. Although it is less satisfactory, owing to its magnesium content, for most uses than either borax or colemanite, substantial tonnages are used in Eastern Europe, Russia, and Asia; it is not traded internationally as a mineral concentrate on a major scale.

The Russians also produce substantial amounts of borate from skarn borosilicates, mainly datolite, with some reports of minor amounts of danburite, ludwigite, and tourmaline. These minerals must be liberated, concentrated, and then dissolved in acid to make a usable product because their natural melting points exceed those of the other minerals used in common glass furnaces.

Pandermite (priceite) was mined in Turkey, and hydroboracite was mined in Russia and Argentina. Other minerals such as inyoite, howlite, meyerhofferite, and kurnakovite are found intimately associated with the major ores. Boracite was used in Germany prior to 1945, where it and minor magnesium borates were recovered as a by-product of potash mining.

A large number of silicates contain boron in their lattices in varying amounts. Axinite, suanite, kotoite, and others are listed in the literature as occurring with the Russian borosilicate ores. Sassolite has only a mineralogical interest at most occurrences, as the quantity found is generally very small. In the Lardarello region of Italy, however, natural steam carries boric acid recoverable as sassolite, and for a long period prior to 1965 several thousand tonnes per year were produced.

Chemistry

Boron is the fifth element of the periodic table and is the only electron-deficient non-metallic element. Thus, boron has a high affinity for oxygen, forming strong covalent boron–oxygen bonds in compounds known as borates. Boron is also the only light element with two abundant isotopes, B^{10} and B^{11}; the former has a large capture cross-section, which makes it an excellent neutron absorber. Within the borate group of minerals, BO_3 units can polymerize (in a similar way to the SiO_4 tetrahedral groups in the silicates) to form chains, sheets, and isolated multiple groups. This is possible because the small B^{3+} ion, which generally coordinates three oxygens in a triangular group, has bond strengths to each oxygen ion with an e.v. of 1; this is exactly half the bonding energy of the oxygen ion. This permits a single oxygen to be shared between two boron ions, linking the BO_3 triangles into expanded structural units (double triangles, triple rings, sheets, and chains). Because of the triangular coordination, boron is also found in tetrahedral groups. In addition to BO_3 and BO_4 groups, natural borates may contain complex ionic groups such as $[B_3O_3(OH)_5]^{2-}$, which consists of one triangle and two tetrahedra. In the structure of colemanite, $CaB_3O_4(OH)_3 \cdot H_2O$, complex finite chains of tetrahedra and triangles occur, and in borax, $Na_2B_4O_5(OH)_4 \cdot 8H_2O$, a complex ion, $[B_4O_5(OH)_4]^{2-}$, consisting of two tetrahedra and two triangles is found. Borates can be classified on the basis of the structural anionic linking into insular (independent single or double BO_3 or BO_4 groups), chain, sheet, and framework structures.

Although it is possible to prepare a three-dimensional framework made up of BO_3 triangles only and having the composition B_2O_3, such a configuration has a very low stability and disorders readily, yielding a glass. Because of its tendency to form somewhat disordered networks of BO_3 triangles, boron is regarded as a 'network former' in glass manufacture and is used in the preparation of special glasses with light weight and high transparency to energetic radiation.

The chemistry of borates is somewhat analogous to, but more complex than, that of silicates. The boron atom in borates can be either three- or four-coordinate bonded to oxygen, forming either planar trigonal BO_3 units or negatively charged tetrahedral BO_4^- units; crystalline sassolite or boric acid, $B(OH)_3$, has only BO_3 units. Other borate minerals are typically salts containing a mixture of BO_3 and BO_4^- units. The nomenclature for borates can be confusing because three different formula systems are in common use. For example, ulexite can be described by an empirical formula, $NaCaB_5O_9 \cdot 8H_2O$, an oxide formula, $Na_2O \cdot 2CaO \cdot 5B_2O_3 \cdot 16H_2O$, and a crystal-structure formula, $NaCa[B_5O_6(OH)] \cdot 5H_2O$. Only the crystal-structure formula conveys information about how the atoms are organized in the crystal.

Many borate minerals contain waters of hydration and, with the exception of borax, are stable at ambient conditions. Borax dehydrates to tincalconite under ambient conditions unless the relative humidity is high. Upon heating, borates lose their attached water molecules before fusing to form a melt. The common borate salts melt at low temperatures; the melts are excellent fluxes or solvents for dissolving other more refractory oxides.

Only alkali-metal borates, ammonium borates, and boric acid have appreciable solubility in water. In dilute aqueous solutions, borates exist as an equilibrium mixture of two species: non-ionized molecular boric acid molecules, $B(OH)_3$, and metaborate ions, $B(OH)_4^-$. For example, borax dissolves in water to give a 1:1 mixture of $B(OH)_3$ and $B(OH)_4^-$; it is a good buffer system at pH 9.4.

$$B(OH)_3 + 2H_2O \ \{\text{Revers React}\} \ B(OH)_4^- + H_3O^+$$

Most borate minerals dissolve in strong mineral acids to liberate the weaker boric acid; for example, boric acid can be produced by mixing ulexite with sulphuric acid. Boric acid is a weak Lewis acid and electron pair acceptor, with a solution pH of about 5.

Crystalline borate minerals are made up of both trigonal and tetrahedral borate units. These borate units can link through boron–oxygen–boron bonds to form chain polyborates, or they may join into rings, which can contain three or four boron atoms (i.e. triborates and tetraborates). These rings can in turn link together to form pentaborates, hexaborates, and higher borates.

Non-borate boron-containing compounds (those containing no boron–oxygen bond) are not found in nature. Such compounds include elemental boron, borides, boron halides, boranes, and organoboron compounds. In general, these manmade materials may be stable inert solids or may hydrolyze in the presence of moisture or oxidize in the presence of oxygen to form the more stable borates.

Borates are widely distributed in nature in low concentrations as alkali-metal and alkaline-earth borate and borosilicate minerals and less commonly as boric acid. They are typically found in soil and rock in concentrations of up to about 450 ppm total boron, distributed in over 150 minerals, primarily as salts of sodium, calcium, and magnesium. The average concentration of boron dissolved in land-surface water is about 0.1 ppm, and in seawater the concentration is about 4.6 ppm.

At low levels, water-soluble boron is an essential micronutrient for the growth and viability of plants; the range between insufficient boron and excess boron is narrow (0.25–15.0 ppm boron), but most soils fall within this range. Borate is found in animal tissues at about 1 ppm as a result of the ingestion of fruits and vegetables; it is not known to have an essential biochemical function, although it may play a role in the body's ability to use calcium. Borate transported in plants and animals is usually complexed with polyalcohols in the aqueous phase.

Depositional Setting and Formation of Borate Deposits

Boron is extremely dispersed in nature, averaging 0.1 ppm in land-surface water, 3 ppm in the Earth's crust, and 4.6 ppm in seawater. There are relatively few occurrences where the element is sufficiently concentrated to be economic. Where a degree of concentration does occur, it is usually a result of local volcanic activity (as a source of boron), a body of water such as a lake (to dissolve boron compounds), evaporative conditions (to concentrate the solution to the point of precipitation), and the deposition of a protective layer of sediment (to preserve the highly soluble borate minerals).

Major borate deposits throughout the world are found in tectonically active extensional regions associated with plate boundaries. Most of the commercial borate deposits in the USA, South America, and Turkey are thought to be associated with continental sediments and volcanism of Neogene age. Many of the older skarn deposits also appear to be related to continental volcanic sources. Marine borate deposits are apparently the product of evaporation of seawater in a restricted basin, probably associated with a seafloor borate source, and/or progressive decanting that preferentially concentrated the borates to many times natural seawater concentrations. Borates associated with igneous and some metamorphic rocks are thought to be an end phase of specialized magmatic segregation or leached from the intruded rocks by

associated hydrothermal fluids (*see* **Mining Geology:** Hydrothermal Ores).

Most of the South American deposits are associated with calcareous tuff, which occurs as a late-stage capping over the borates, and in some cases with halite and gypsum. Recent volcanic activity is indicated by basaltic to rhyolitic flows in adjacent areas, and a volcanic source for the borates is presumed. The Salar deposits of South America consist of beds and nodules of ulexite, with some borax or inyoite, associated with Holocene playa sediments, primarily mud, silt, halite, and gypsum.

Borates can be concentrated in different ways:

- by chemical precipitation in the neighbourhood of boron-bearing springs in playa-type basins (e.g. the Boron, Searles Lake, and Billie deposits of California, and the Kirka, Sultançayir, Bigadiç, and Emet deposits in Turkey);
- by precipitation from seawater in the closing phase of a salt-forming evaporate cycle (e.g. at Stassfurt in Germany);
- by contact metasomatism with dolomite or magnesite, forming magnesium borates such as ludwigite, kotoite, and ascharite (e.g. the Teazhnoe deposits at Yakutia in Russia, and the Hol-Kol deposit in North Korea);
- by contact metasomatism with limestone, forming boron silicates such as datolite and danburite (e.g. at Ak Akhdar, Pamir, and Dalnegorskoye, Primorsky, Russia); and
- by volcanic exhalations of boric acid, i.e. sublimates (e.g. at Clear Lake, Lake County, California, at Salar de Surire, Chile, and in the Maremma area, Tuscany, Italy).

The formation of borate deposits can be tentatively summarized as follows.

- Although boron is a rare element (average content in the Earth's crust is 10 ppm), extraordinary concentrations can be found in certain places.
- Over 150 minerals are known to contain boron, and they are found in various geological environments. These can be subdivided into three groups:
  1. a skarn group associated with intrusives and consisting of silicates and magnesium-iron oxides;
  2. a magnesium oxide group hosted by marine evaporitic sediments; and
  3. a sodium– and calcium–borate hydrates group associated with lacustrine (playa lake) sediments and explosive volcanic activity.
- The following conditions are essential for the formation of economically viable borate deposits in playa-lake volcanosedimentary rocks (Figure 3):

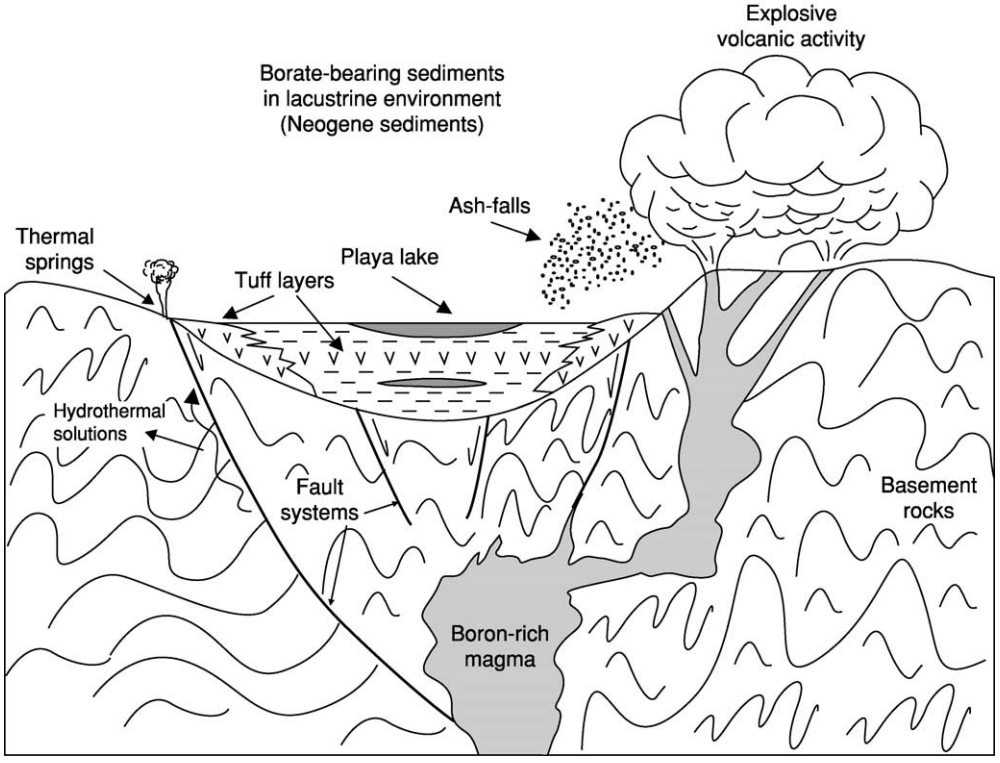


Figure 3 Generalized playa-lake depositional model, showing the formation of borate deposits in the Neogene basins of western Anatolia, Turkey.

1. formation of playa-lake environment;
2. concentration of boron in the playa lake, sourced from andesitic to rhyolitic volcanics, direct ash fall into the basin, or hydrothermal solutions along graben faults;
3. thermal springs near the area of volcanism;
4. arid to semi-arid climatic conditions; and
5. lake water with a pH of between 8.5 and 11.

Borate Deposits in Non-Marine Basins

The largest known borate deposits originated as chemical precipitates and are found interbedded with clays, mudstones, tuffs, limestones, and similar lacustrine sediments. There is evidence that most of these deposits were closely related in time to active volcanism. Thermal springs and hydrothermal solutions associated with this volcanic activity are therefore regarded as the most likely source of the boron.

Several South American springs in volcanically active areas are depositing borates, and the first borax discovered in the USA was found in the muds associated with warm springs at Clear Leak, California, a volcanically active area. The Italian (Tuscany) steam vents from which sassolite was recovered also represent an active volcanic source.

In addition to the concentrated source of the borates and a 'basin' in which they can collect, an arid to semiarid climate also seems to be essential during the deposition and concentration of economic amounts of soluble borates. These soluble borates can, in the long run, be preserved only by burial; however, the lack of deposits of soluble borates older than mid-Tertiary may indicate that even burial is not able to protect borates over long periods of geological time.

Hydrated borates may accumulate in several ways within a non-marine basin. They may be deposited in layers in a spring apron around a borate spring, with ulexite, borax, or inyoite as the primary borate mineral. Borates may also form in a pool dominantly fed by a borate spring, with borax crystals formed in bottom muds or at the intermittently dried margins (as at Clear Lake, and at Salar de Surire, Chile). If the spring flow is low or intermittent, evaporation develops a surface efflorescence or precipitate or an accumulation of crystals just below the surface (examples are the marsh or playa deposits that have been mined in California and Nevada and some of the salar deposits of South America). Finally, there are lake deposits, whose occurrence requires much more than seasonal flooding (examples are the borax deposits at Boron and Kirka, which were formed by chemical precipitation in a closed basin). There is some disagreement as to whether the thicker ulexite deposits, such as those in Turkey and Death Valley, are spring-apron or lake deposits. In fact, there is probably a gradation between spring-apron and small-lake deposits. These borate lakes are essentially monomineralic, in the sense that no other salts occur in major quantities. There is, however, another type of borate lake deposit, consisting of mixed salts and/or brine containing borates in sufficient quantity of justify recovery. Searles Lake in California has been cited as a type example of a multicomponent deposit formed by the evaporation of lake waters. Numerous studies of Searles Lake have concluded that boron and the other dissolved constituents originated along the eastern front of the Sierra Nevada, were concentrated and decanted in a series of up-drainage lakes, and were finally precipitated and preserved in Searles Lake itself. This scenario would also fit several of the Chinese and Tibetan lakes. The borates found in the large South American salars, such as Uyuni and Atacama, may also have formed by leaching of the surrounding rocks and subsequent evaporation, although the role of local mineralized spring waters containing boron has not been fully evaluated.

Marine Evaporites

Borates of marine origin have been found in commercial quantities only in Europe. These are magnesium borates associated with Permian salt deposits. They were produced in Germany, as a by-product of potash mining, and in the Inder region of Kazakhstan. The Inder deposits, where the borates occur as veins in the cap of a very large salt dome complex, are thought to have been remobilized and concentrated from the salt during the intrusion of the salt dome itself. Some of the Chinese deposits of the Liaoning Peninsula may be of similar origin, although they occur as veins in Precambrian metamorphosed limestone and magnesite.

The Inder Lake brines, which are also a source of Kazakhstan borate, appear to be simply a sump accumulation of borates leached from the huge Inder salt dome complex. The Kara-Bagaz-Gol Lagoon borates on the east shore of the Caspian Sea appear to have leached from marine brines.

Magmatic Sources

Pegmatites and contact-metamorphic rocks contain assemblages of various boron-containing minerals, such as datolite, ludwigite, paigeite, and tourmaline. These represent concentrations of boron that relate more or less directly to the crystallization of intrusive granitic magma. Analyses show that granites average about 10 ppm boron, with a few exceptions ranging up to 300 ppm. However, boron does not readily enter into the crystal structures of the common

rock-forming minerals; hence, when magma crystallizes, the boron is thought to leave with the released water. The combination of high temperatures (300–400°C) and fluids under high pressure at an intrusive contact could also extract some boron from the adjacent country rocks. These borate skarn deposits, some of which are associated with iron ores and magnesium deposits of commercial grade, are mined in both eastern Russia and China.

Exploration

Borate exploration consists of detailed prospecting of favourable areas followed by drilling, and uses all the tools available to the exploration geologist. The recognition of trends of favourable host rocks and structures is an important guide to areas that are of possible interest. Satellite imagery, both real and false-colour, and standard photographic interpretation can be successfully used under certain conditions.

In most parts of the world, the identification of a Cenozoic suite of non-marine fine-grained sediments and tuffs is the usual starting point for the field geologist, because most commercial borates are associated with these rocks. Any accumulation of salts in these sediments is examined closely. Because many borates are associated with volcanic rocks, volcanic centres, flows, ash deposits, and tuffs, particularly if they are zeolite-bearing, may also be favourable guides to borate prospecting.

Borate exploration begins with visits to Neogene basins that appear to be favourable for borate deposition. It should be remembered that the mobility of boron in the Earth's crust is very high, like chlorine, bromium, sulphur, etc. Because of its rapid alteration by water and carbon dioxide, borate can be altered to calcite and boric acid washed away as follows:

$$2CaO \cdot 3B_2O_3 \cdot 5H_2O(\text{colemanite}) + 2H_2CO_3(\text{carbonic acid}) + 2H_2O \rightarrow 2CaCO_3(\text{calcite}) + 6H_3BO_3(\text{boric acid})$$

The following criteria, based on the criteria outlined above in the section on the formation of borate deposits, are used during exploration in Turkey, the USA, and Argentina:

- contribution of Miocene volcanics and volcanoclastics to Neogene playa-lake basins;
- formation of capping limestone;
- claystone–limestone–marl intercalations;
- evidence of hot springs or hydrothermal solutions carrying boron into the playa lake; and
- evaporitic horizons or an indication of leaching of them.

The skarn borates of Eastern Europe and Asia were found by careful prospecting in geologically well-preserved fold belts where limy sediments are in contact with potassic to alkaline volcanics. Datolite and danburite occur in skarns where the limestone was originally rich in calcium and silica; axinite, kotoite, and ludwigite skarns are hosted by dolomitic limestones. The magnesium borates and borates associated with iron ores are generally of low grade. Marine borates are sought in tectonically stable areas with shallow to outcropping salt structures capped by gypsum, where near-surface borate deposits are identified by whitish soil cover and scanty vegetation.

In arid regions, ulexite often accumulates at, or just beneath, the current surface of salt flats and playas, indicating that boron is moving within the system. These recent crusts may also indicate brine deposits containing boron concentrations of interest. In either case, additional prospecting is usually justified; playas are usually prospected by pitting on a 100–500 m grid. Springs and recent spring deposits containing anomalous borates may also be used as a guide to the presence of ore in certain areas.

Geochemical surveys are useful for narrowing down prospect areas to a drill target. Both soil and rock-chip sampling techniques are used in exploration programmes, with boron, strontium, arsenic, and lithium as a common suite of elements. Beryllium is also used in the search for skarn borates in Russia, as are complex boron-magnesium-calcium-chlorine ratios. Sampling of both surface and well water may be useful. Certain plants are boron sensitive, and vegetation surveys may prove interesting.

Geophysical surveys, particularly gravity and magnetic, are used to outline target basins or structures beneath sedimentary basin fill. Resistivity and seismic surveys have been used to define basin structures and formations that may be associated with the borates in that area. Various down-hole well-logging techniques, including natural gamma and neutron probes, can indicate the approximate percentage of borates and clay in zones of special interest.

Geological mapping followed by drilling is still the definitive test in most areas of the world. While rotary drill methods may be used, cores are generally taken of the most prospective zones. Assays of B_2O_3 and other associated elements (arsenic, lithium, and strontium) are then run on the horizons that appear favourable for borates. Because the saline borates are water soluble, short core runs are used, but the common borates generally core well, with recoveries of more than 90%.

In areas of doubt, two easy field tests can determine the presence of borates. The original flame test, where the mineral is soaked with sulphuric acid and alcohol

and ignited ('she burns green, Rosie...') is still one of the most diagnostic. The turmeric test, where hydrochloric acid and turmeric solution turn the specimen red-brown, is also used. As chemical tests performed in the laboratory are more diagnostic and accurate, most field samples are sent to the laboratory for analysis.

Under current economic conditions, bedded deposits of borax, colemanite, and ulexite are not generally sought at depths greater than 500 m. Brines with a high borate content, particularly those associated with other salts of value, might be extracted from greater depths in certain circumstances. The skarn and magnesium borates are currently economical only from surface and near-surface excavations.

Mining and Mineral Processing

Most of the world's commercial borate deposits are mined by open pit methods. The world's major borate operations – the Boron mine of US Borax at Boron (Kramer) in California and the Kyrka mine of Eti Holding in Turkey – are huge open pit mines using large trucks and shovels and front-end-loader methods for ore mining and overburden removal. Ores and overburden are drilled and blasted for easier handling. The Boron operation uses a belt conveyor to move ore from the in-pit crusher to a coarse ore stockpile from which it is reclaimed by a bucket wheel that blends the ore before it is fed to the refinery. Kirka uses trucks to haul material to a crusher near the refinery, which is about 0.5 km from the current ore faces.

Smaller operations in Argentina, Chile, China, Turkey, and Russia use similar methods, but on a smaller scale appropriate to the scale of the operation. Some of the South American and Chinese salar operations use hand labour to mine the thin salar borates, generally after stripping the overburden with a small bulldozer or front-end loader.

Borates are mined by underground methods in the Liaoning area of north-eastern China and at the Billie and Gerstley mines in Death Valley, California. Borate brines are recovered at Searles Lake in California and in the Qinghai Basin of China; brines are also used in the Inder region of Kazakhstan. Borate-containing brines from several salars in South America are being considered for production. Also under study is the recovery of low-grade borates by in situ acid leaching, producing a crude boric acid for subsequent refining.

Processing techniques depend on both the scale of the operation and the ore type, with either the upgraded refined mineral (borax, colemanite, ulexite) or boric acid as the final product for most operations.

Borax–kernite ores (Boron, Kyrka, Tincalayu) are crushed to less than 2.5 cm and then dissolved in hot water or recycled borate liquor. The resultant strong liquor is clarified and concentrated in large countercurrent thickeners, filtered, fed to vacuum crystallizers, centrifuged, and dried. The final product is refined borax decahydrate or pentahydrate or fused anhydrous borax, or is used as feed for boric acid production.

The ulexite from most of the South American salars is air dried, screened, and bagged. It is then combined with locally available sulphuric acid to produce a relatively low-grade boric acid or exported as feed for boric acid plants elsewhere.

Colemanite concentrates are used directly in specific glass melts or as a feed for boric acid plants. The magnesium borates are generally concentrated, dissolved in acid to remove the magnesium, and converted to boric acid or sodium borates. The borosilicates of the Bor deposit in eastern Russia, with their relatively low B_2O_3 grades, are crushed and then run through a complex plant, which includes magnetic separators, heavy-media separators, and flotation cells. The concentrates are then dried, leached, and calcined before being converted to boric acid or a sodium borate.

Brines from Searles Lake, and presumably the Chinese sources, are recovered by either controlled evaporation or carbonation. In the latter process, carbon dioxide produced from lime kilns or flue gas is bubbled through the brine to crystallize sodium bicarbonate; borax is then crystallized in vacuum crystallizers. In the 'evaporation' process, a rapid controlled cooling electively crystallizes the various salts. The remaining borate liquor is fed to tanks containing borax seed crystals, which aid the recovery of borates from the liquor. The resultant slurry is filtered, washed, redissolved, and fed to vacuum crystallizers, which produce dehydrate borax products or boric acid.

Boric acid is one of the final products produced from most of the processes (Table 4). The world's largest boric acid facility is located adjacent to the Boron pit. It uses either refined borates or hydrated kernite as feed, with sulphuric acid as the reacting

Table 4 Commercial refined borate productions

Product	*Formula*	*%B_2O_3*
Borax decahydrate	$Na_2B_4O_7 \cdot 10H_2O$	30.5
Borax pentahydrate	$Na_2B_4O_7 \cdot 5H_2O$	47.8
Boric acid	H_3BO_3	56.3
Borax anhydrous	B_2O_3	100.0
Sodium perborate	$NaBO_3 \cdot 4H_2O$	22.0
Raw borax anhydrous	$Na_2B_2O_3$	69.2

agent. Other smaller facilities around the world use smelter acid or other locally available acid feed-stock to produce products acceptable to their local markets.

Uses

The principal uses of borates have not changed much in the past decade. Major uses include fibreglass, insulation, textile or continuous-filament glass fibres, glass, enamels and frits, and fertilizers (Figure 4). Bleaches and detergents are also a major end use; however, demand for borates for glass and glass fibres, including fibreglass, is increasing. Boron fibre-reinforced plastics are used in quantity for aerospace frame sheathing, where they combine flexibility and light weight with strength and ease of fabrication. Relatively minor uses that are expected to increase in the near future include fertilizers, wood preservatives, alloys and amorphous metals, fire and flame retardants, and insecticides. However, the promising field of boron–iron–silicon electrical transformers has not developed as rapidly as predicted owing to various cost factors.

Miscellaneous uses include pharmaceuticals, cosmetics, anticorrosion compounds, adhesives, abrasives, insecticides, metallurgical processes, and nuclear shielding. Research is still continuing in many areas. One of the more publicized applications is in super-magnets, where borates, combined with rare earths, nickel, and iron, produce an alloy that can be used to make electromagnets for computer drives, high-fidelity speakers, automobile starter motors, and various household appliances. Another new field is biological growth-control agents.

Borates can be used to protect the environment by aiding the conversion of heavy metals in industrial waste streams into recoverable free metals and by removing impurities from polymers used to bleach wood pulp for paper production. Borates also help to control the refractive index in optical fibres for medical research, where precise control is needed. Other medical applications include treatments for cancer, in which the ^{10}B isotope reacts with low-energy neutrons to give off short-range alpha particles that can be used for microsurgery in previously inoperable areas of the brain. Current tests on boron analogues indicate that they may be effective in reducing serum cholesterol and other disease-causing proteins.

Borates have become a relatively modestly priced industrial mineral commodity in recent years, following the development of the large deposits at Boron and, more recently, Kyrka. Prices are directly related to the cost of production, of which the major part is the cost of fuel for drying, dehydrating, and melting the refined ore into the products desired by industry. Industry prices for most products have increased only in line with inflation.

Both Western Europe and Japan, neither of which have local borate sources, are major markets for US and Turkish production. South America is largely self sufficient, and is producing an increasing excess, which is exported mainly to Europe and Japan. Russia and China both appear to be self sufficient in borates

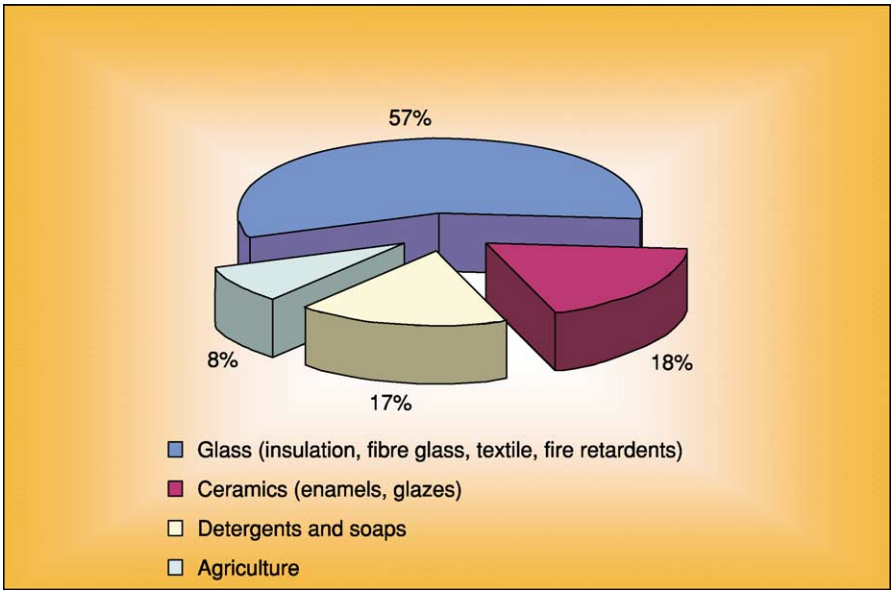


Figure 4 World borate end uses.

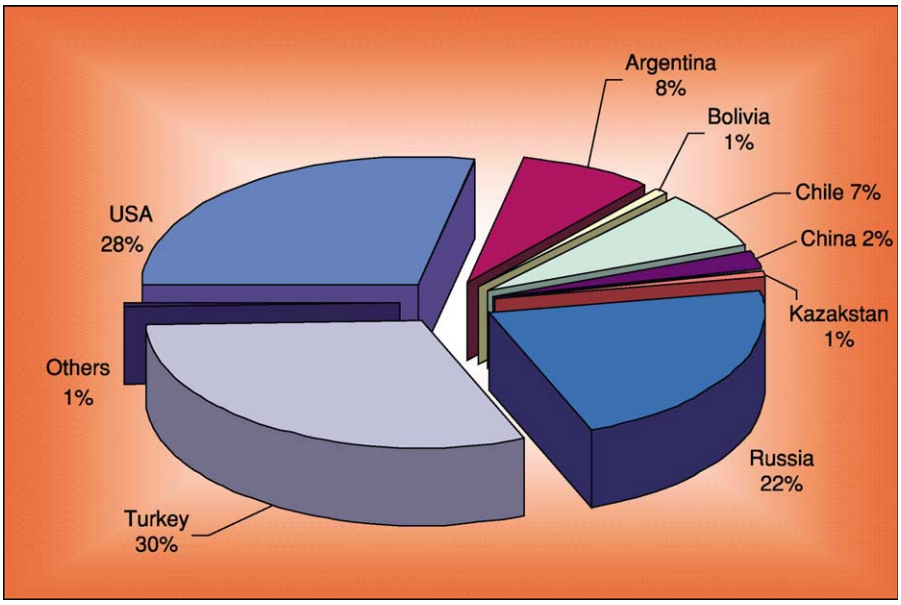


Figure 5 World borate production.

at this time, although their costs of production per tonne of B_2O_3 are thought to be relatively high.

Borates are a lightweight commodity and are generally sold in bulk by carload lots, in intermediate bulk containers or 'super-sacks', and by palletized bags. Overseas shipments are mostly made in bulk from special terminals at Los Angeles, California and from Bandirma on the Marmara Sea in Turkey to similar terminals in the Netherlands, Belgium, and the UK, from where they are moved by barge, as well as by rail and truck. Other imports to Europe arrive in Italy and Spain. Imports to the Far East are generally sold in small bag lots. Bulk imports to the USA (mainly colemanite) usually land in Charleston, South Carolina, where there are grinding facilities; this colemanite is then shipped to eastern fibreglass manufacturers. There is no import duty on borates brought into the USA.

Boron consumption is directly related to the use of glass, glass fibres, and ceramics. These materials, along with certain plastics that contain borate products, are seen as having a steady consumer demand in the construction and houseware markets well into the next century. Other major uses, such as detergents, plant foods, and wood preservation, are expected to show a slowly rising demand. Total world borate demand is expected to grow at about 3% per annum in the near future, based on industry forecasts.

Known reserves of borate minerals are large, particularly in Turkey, South America, and the USA, and production from Turkey and the USA will continue to dominate the world market. However, borates from other areas will probably take up an increasing share of the world market (Figure 5). This trend is already evident, with boric acid from Chile reaching the Far East and Europe, and both Russia and China beginning to export (Table 5).

Table 5 The reserves and life estimates of the world's borate deposits

Country	*Known economic reserve (million tonnes of B_2O_3)*	*Total reserve (million tonnes of B_2O_3)*	*Estimated life of known reserve (years)*	*Estimated life of total reserve (years)*
Turkey	224 000	563 000	155	389
USA	40 000	80 000	28	55
Russia	40 000	60 000	28	69
China	27 000	36 000	19	25
Chile	8000	41 000	6	28
Bolivia	4000	19 000	3	13
Peru	4000	22 000	3	15
Argentina	2000	9000	1	6
Kazakstan	14 000	15 000	10	10
Total	363 000	885 000	253	610

There are few substitutes for borates. In most applications, they provide unique chemical properties at a reasonable price; this is particularly true for glass fibres and in the field of heat- and impact-resistant glass. Borates are an essential part of certain plant foods. Their use in nuclear-reactor shielding and control is well documented. Future markets are difficult to predict. Based upon recent history, the

major world consumers of borates will continue to be the developed countries of North America, Europe, and Japan.

See Also

Geysers and Hot Springs. **Minerals:** Definition and Classification. **Mining Geology:** Exploration; Mineral Reserves; Hydrothermal Ores. **Sedimentary Environments:** Lake Processes and Deposits. **Sedimentary Rocks:** Evaporites.

Further Reading

Aristarain LF and Hurlbut CS Jr (1972) Boron minerals and deposits. *Mineralogical Record* 3: 165–172, 213–220.

Floyd PA, Helvaci C, and Mittwede SK (1997) Geochemical discrimination of volcanic rocks, associated with borate deposits: an exploration tool? *Journal of Geochemical Exploration* 60: 185–205.

Foshag W (1921) The origin of the colemanite deposits of California. *Economic Geology* 16: 194–214.

Garrett DE (1998) *Borates. Handbook of Deposits, Processing, Properties, and Use*. London: Academic Press.

Grew ES and Anovita LM (eds.) (1996) *Boron. Mineralogy, Petrology and Geochemistry*. Reviews in Mineralogy 33. Washington DC: Mineralogical Society of America.

Helvaci C (1978) A review of the mineralogy of the Turkish borate deposits. *Mercian Geology* 6: 257–270.

Helvaci C (1986) Geochemistry and origin of the Emet borate deposits, western Turkey. *Faculty of Engineering Bulletin, Cumhuriyet University, Series A. Earth Sciences* 3: 49–73.

Helvaci C (1984) Occurence of rare borate-minerals: veatchite-A, tunelite, terggite and cahnite in the Emet borate deposits, Turkey. *Mineralium Deposita* 19: 217–226.

Helvaci C (1989) A mineralogical approach to the mining, storing and marketing problems of the Turkish borate production. *Geological Engineering* 34/35: 5–17.

Helvaci C (1995) Stratigraphy, mineralogy, and genesis of the Bigadiç borate deposits, western Turkey. *Economic Geology* 90: 1237–1260.

Helvaci C and Orti F (1998) Sedimentology and diagnesis of Miocene colemanite-ulexite deposits (western Anatolia, Turkey). *Journal of Sedimentary Research* 68: 1021–1033.

Helvaci C and Alonso RN (2000) Borate deposits of Turkey and Argentina: a summary and geological comparison. *Turkish Journal of Earth Sciences* 24: 1–27.

Helvaci C and Firman RJ (1976) Geological setting and mineralogy of Emet borate deposit, Turkey. *Transactions/Section B, Institute of Mining and Metallurgy* 85: 142–152.

Kistler RB and Helvaci C (1994) Boron and borates. In: Carr DD (ed.) *Industrial Minerals and Rocks*, 6th edn, pp. 171–186. Littleton, CO: Society for Mining, Metallurgy and Exploration Inc.

Muessig S (1966) Recent South American borate deposits. In: Rau JL (ed.) *Proceedings of 2nd Symposium on Salt, Northern Ohio Geological Society, Cleveland*, vol. 1, pp. 151–159.

Palmer MR and Helvaci C (1995) The boron isotope geochemistry of the Kirka borate deposit, western Turkey. *Geochimica et Cosmochimica Acta* 59: 3599–3605.

Palmer MR and Helvaci C (1997) The boron isotope geochemistry of the Neogene borate deposits of wetsern Turkey. *Geochemica et Cosmochimica Acta* 61: 3161–3169.

Travis NJ and Cocks EJ (1984) *The Tincal Trail. A History of Borax*. London: Harrap.

Watanebe T (1964) Geochemical cycle and concentration of boron in the Earth's crust. *Verdenskii Institute Geochemistry and Analytical Chemistry USSR* 2: 167–177.

Carbonates

B Jones, University of Alberta, Edmonton, AB, Canada

Introduction

The term 'carbonates' refers to any sediments or sedimentary rocks that are formed primarily of $CaCO_3$ and/or $CaMg(CO_3)_2$. Carbonate sedimentary rocks, known from strata of all ages, house up to 50% of the world's oil and gas reservoirs and are commonly used for building stone, building aggregates, and cement production. Their chemical simplicity belies the bewildering array of textures evident in these rocks and the complexity of the processes that govern their formation. Indeed, the origins of many of these rocks, their constituents, and their fabrics are still matters of ongoing debate.

Carbonate sediments form and accumulate in a broad spectrum of depositional environments (Figure 1). Although most commonly associated with shallow tropical seas (Figure 2), carbonate sediments and precipitates can also form in oceans, lakes, streams, caves, and cold or hot springs. Animals and plants are multifaceted actors in the formation of these sediments – their calcareous skeletons contribute to the sediment, they promote precipitation through modification of the local environment, and they trap and bind sediments.

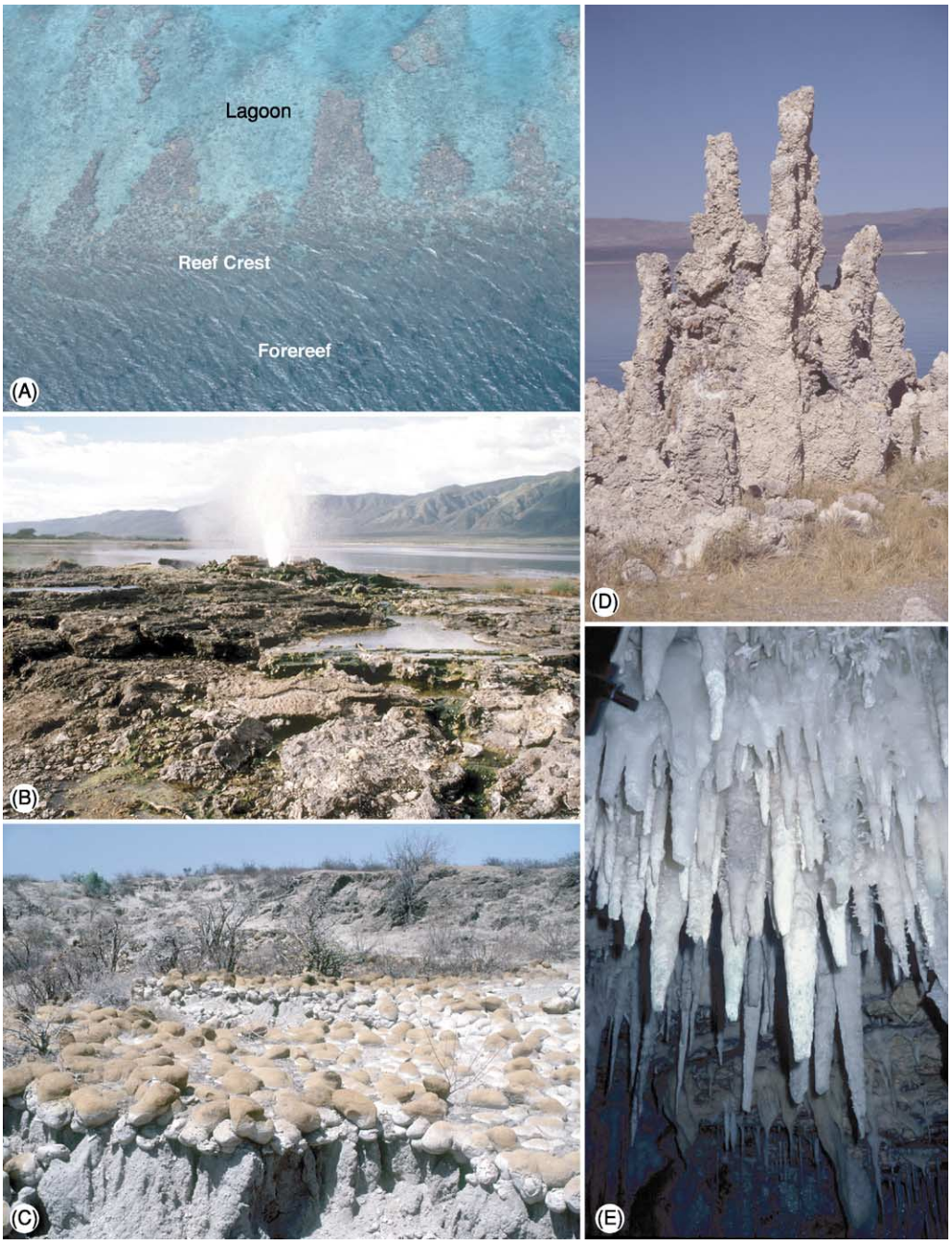


Figure 1 Carbonate depositional settings. (A) Marine environment with a marginal reef separating the forereef from a lagoon; oblique aerial view of a reef system on the coast of Caicos Island. (B) Hot-spring environment with carbonate deposits formed of dendritic calcite crystals; spring KL6 at Loburu, Lake Bogoria, in the Kenyan Rift Valley. Photo courtesy of Robin Renaut. (C) Lake environment, showing giant oncoids (20–30 cm) in the early Mid-Holocene Galana Boi Formation, Lake Turkana, Kenya; wind-blown sand covered the oncoids after the lake regressed. Photo courtesy of Robin Renaut. (D) Lake environment with large towers of tufa that formed around springs vents on a lake floor; the towers formed when the lake level was considerably higher than it is today (highest tower is ~5 m high); Mona Lake, California. (E) Cave environment, showing stalactites formed of calcite hanging from the roof of Crystal Cave, Bermuda.

Constituent Minerals

Solid $CaCO_3$ exists in three different crystal structures (polymorphs), namely, calcite (trigonal system), aragonite (orthorhombic system), and vaterite (hexagonal system). Vaterite is very rare and needs no further consideration. Aragonite is a common component of many carbonate sediments despite the fact that it is metastable at the low temperatures

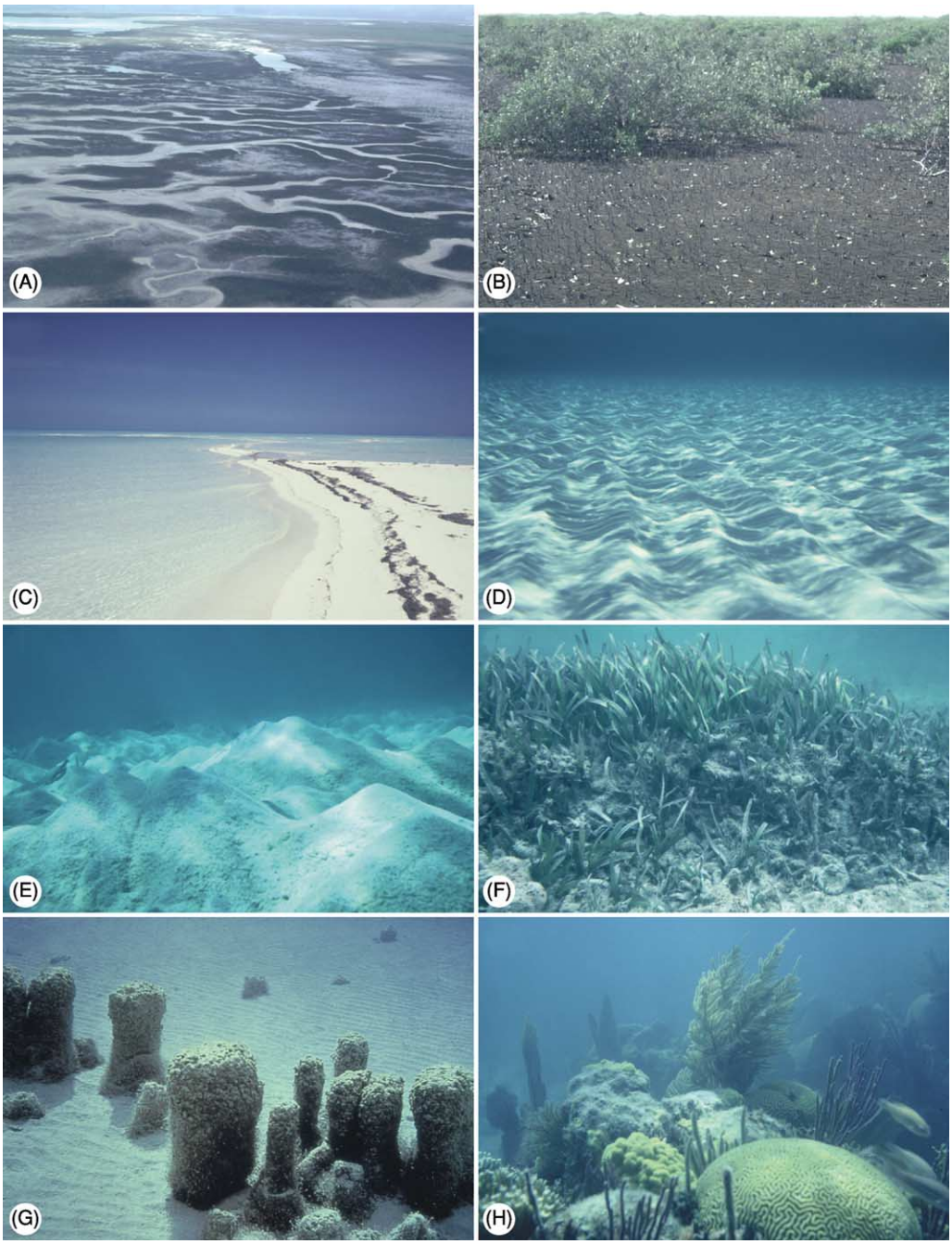


Figure 2 Marine carbonate depositional environments. (A) Intertidal and supratidal flats cut by numerous tidal creeks; aerial view, Caicos Island. (B) Intertidal flat covered by thick algae mat and mangroves; intertidal flat, Caicos Island. (C) Beach and shallow subtidal setting covered with ooid sands; Joulters Key, Bahamas. (D) Shallow subtidal seafloor covered with rippled ooid sands; Joulters Key, Bahamas. (E) Lagoonal sands with numerous mounds created by the burrowing ghost shrimp, *Callianassa*; water depth ~6 m, East Sound, Grand Cayman. (F) Shallow lagoon with *Thalassia* (seagrass), which has an extensive root system that binds sediment; water depth ~0.6 m, Pease Bay Lagoon, Grand Cayman. (G) Lagoon with stromatolites growing on the seafloor; columns ~1 m high, Shark's Bay, Australia. Photo courtesy of Ian Hunter. (H) Small patch reef formed of corals and other organisms; water depth ~5 m, near North Rock, Bermuda.

and pressures found at Earth's surface. With time, aragonite will invert to calcite, the more stable polymorph, by processes that are still poorly understood.

Dolomite, $CaMg(CO_3)_2$, a common component of many carbonate rocks, typically forms through the replacement of a precursor limestone. The origin of dolomite has been hotly debated ever since the term

was introduced over 200 years ago. The fact that it cannot be produced in the laboratory under low-temperature and low-pressure conditions has seriously impeded the understanding of its genesis. Calcite and dolomite are both prone to isomorphic substitutions. Magnesium commonly substitutes for calcium in calcite to produce magnesium calcite. Calcite, with <4 mol% $MgCO_3$, is referred to as low-magnesium calcite (LMC), whereas calcite, with >8 mol% $MgCO_3$, is referred to as high-magnesium calcite (HMC). Similarly, Fe commonly substitutes for Ca or Mg in calcite and dolomite to produce ferroan calcite and ferroan dolomite. Such substitutions are important because they affect crystal lattice structures, which can have a profound effect on the solubility and recrystallization of these minerals during diagenesis.

Components

Carbonate sediments are typically formed of various chemical or biochemical precipitates that are held in a matrix or cement (Figure 3). These precipitates, termed allochems, form within a basin and are organized into discrete aggregated bodies that have usually undergone some type of transport. A matrix refers to any mechanically deposited sediment (e.g., mud) that is deposited between the allochems, whereas cement refers to crystals that have been passively precipitated from pore fluids between the allochems. Non-marine carbonates typically, but not always, contain fewer allochems than do their marine cousins. Most cave deposits and hot-spring deposits, for example, are formed of crystalline calcite and aragonite that have been precipitated directly from water. Bizarre and unusual crystal morphologies commonly characterize these deposits.

Allochems may be coated grains (including ooids; Figure 2C and D), pellets/peloids, bioclasts, and lithoclasts. Although most coated grains are <2 mm in diameter (e.g., ooids), some of those found in lacustrine and spring settings are up to 0.5 m in diameter (Figure 1C). Depending on the environment in which the coated grain forms, the nucleus can be anything from a small sand grain to a large coral fragment to an animal bone. The cortex may be formed of radial crystals and/or laminae that are symmetrically or asymmetrically arranged around the nucleus. The cortical laminae may originate through organic processes and/or inorganic processes. The plethora of names applied to coated grains commonly reflects their size, internal structures, and/or the organism that mediated their formation. Common coated grains include ooids (<2 mm diameter), pisoids (>2 mm diameter), rhodolites (growth mediated by red algae), and oncolites (growth mediated by algae).

Faecal pellets, typically ovate and formed of micrite, are produced by animals (e.g., sea cucumbers, gastropods, worms) that extract organic material from the vast quantities of sediment that they ingest and process through their digestive tract. Although typically <1 mm long, some of the largest sea cucumbers produce faecal pellets up to 1 cm in diameter. Faecal pellets, which are coated with mucus when excreted by the animal, are delicate structures prone to breakdown and disintegration. Compaction can, for example, transform a faecal pellet grainstone into a featureless micrite. Not all ovate grains formed of micrite, however, are faecal pellets. Such grains, produced by micritization of skeletal calcite grains, are known as peloids.

Bioclasts are grains of any size that are derived from the calcareous skeletons of animals and plants (Figure 3A, B, and D). The composition and size of bioclasts depend on the skeletal architecture of the original animal/plant and/or the amount of postmortem physical and biological degradation that it underwent. Compositional variance arises because different animals and plants have skeletons formed of aragonite (e.g., corals, bivalves, calcareous algae), LMC (e.g., brachiopods, bryozoans, trilobites), or HMC (e.g., crinoids, echinoids, red algae). Some animals even have the ability to produce skeletons formed of alternating layers of aragonite and calcite. Tropical sediments, dominated by aragonite, are highly reactive, with inversion of the metastable aragonite to calcite being the dominant process. By comparison, temperate carbonates, which are dominated by LMC and HMC, are less prone to change because they contain only minor amounts of metastable aragonite. Bioclastic sediments typically have a polymodal grain size distribution because the constituent grains are derived from the skeletons of many different animals and plants.

Lithoclasts, derived from lithified or semi-lithified substrates, are important components of some carbonates. Tabular carbonate lithoclasts are commonly produced by storms and hurricanes that generate currents with power sufficient to rip up and transport lithified sediments from the seafloor.

The matrix in many carbonate rocks is microcrystalline calcite, or micrite, which commonly refers to any calcareous sediment that is formed of grains that are <4 μm long. Although originally defined as mechanically deposited sediment, micrite can also be formed as cement. Accordingly, the term 'micrite' is commonly used without any genetic connotation. Many carbonates, irrespective of their place of formation, are formed of allochems that are held in calcite and/or aragonite cements that were

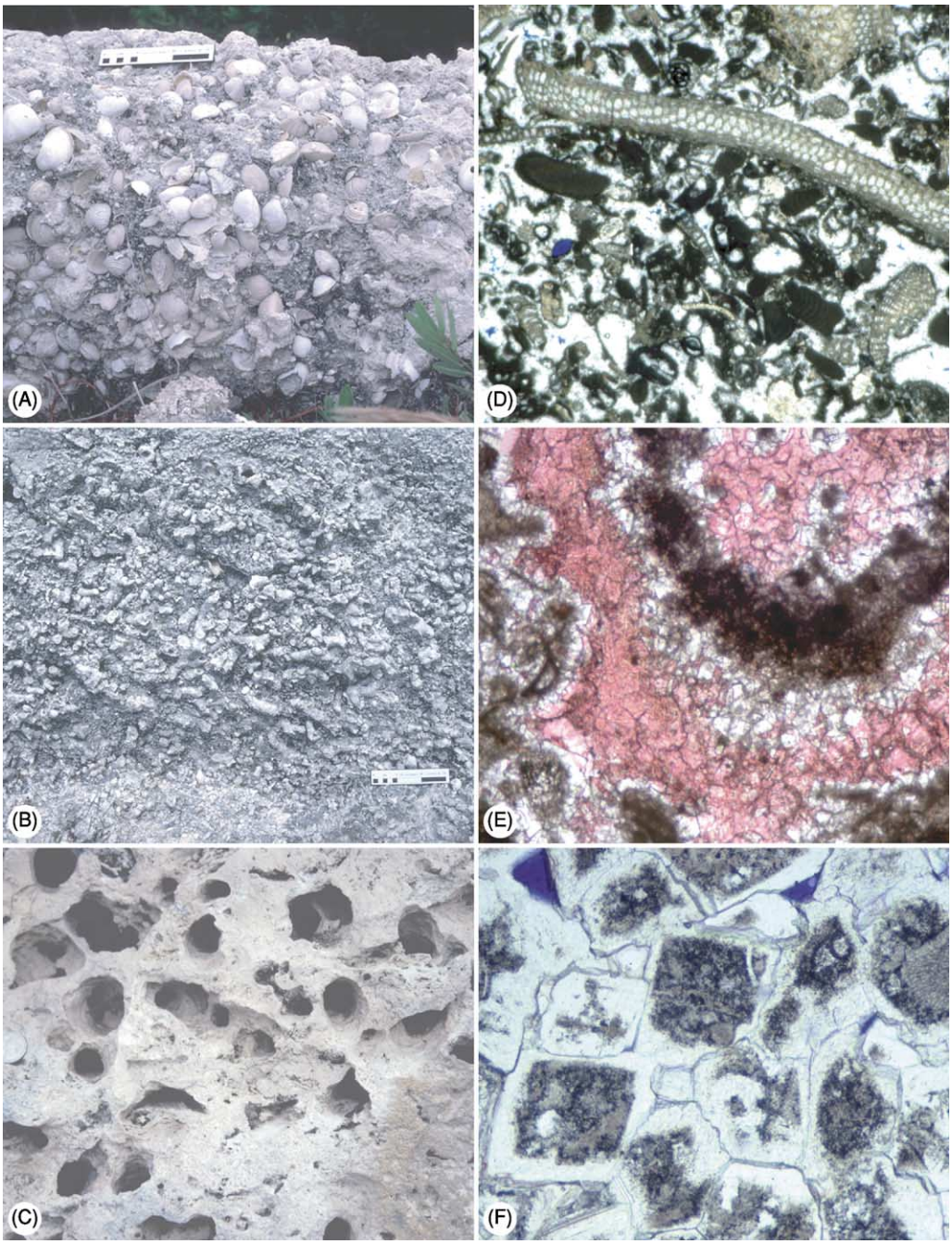


Figure 3 Lithified carbonates. (A) Articulated bivalves embedded in a packstone matrix; Ironshore Formation, Pleistocene, Grand Cayman. (B) Well-preserved patch reef formed of *Porites*; Ironshore Formation, Pleistocene, Grand Cayman. (C) Limestone with fossil-mouldic porosity (after corals); Pleistocene, Barbados. (D) Thin-section photomicrograph, showing grainstone formed of bioclasts embedded in spar calcite cement; Brac Formation, Oligocene, Cayman Brac. Field of view covers ~6 × 6 mm. (E) Finely crystalline dolostone with cavities lined with dolomite cement (white) and filled with calcite cement (stained red by Alizarin Red S solution); Cayman Formation, Oligocene, Cayman Brac. Field of view covers ~1 × 1 mm. (F) Sucrosic dolomite formed of euhedral dolomite crystals that have a dirty core surrounded by a clear rim; note ghost structures of skeletal fragments in cores, inherited from precursor limestone; Brac Formation, Oligocene, Cayman Brac. Field of view ~1.5 × 1 mm.

precipitated following their deposition (Figure 3D). Cementation may take place while the sediments are still in their original depositional setting (e.g., hardgrounds, reefs) or millions of years later, after the sediments have been buried.

Classification

Numerous classification schemes have been proposed for carbonate sediments and sedimentary rocks. The Folk and Dunham classification schemes, which

reflect different philosophical approaches to the naming and genesis of carbonate rocks, are the most commonly used schemes. Usage of one or the other of these schemes is a matter of personal preference and the type of study that is being undertaken. Fundamentally, names derived from Folk's scheme reflect the dominant allochem (e.g., ooid, pellet, bioclast, or lithoclast) and the micrite matrix or spar calcite cement found between the allochems (Figure 4). For example, an oomicrite refers to a rock formed of ooids embedded in micrite matrix, whereas an oosparite refers to a rock formed of ooids embedded in a spar calcite cement.

Dunham's scheme is based on the depositional texture of the rock and makes no specific reference to the allochems, the matrix, or the cement (Figure 5). The progression from mudstone to wackestones to packstone to grainstone reflects, for example, a progressive decrease in the amount of mud (i.e., micrite) and hence ever-increasing energy levels. This scheme was

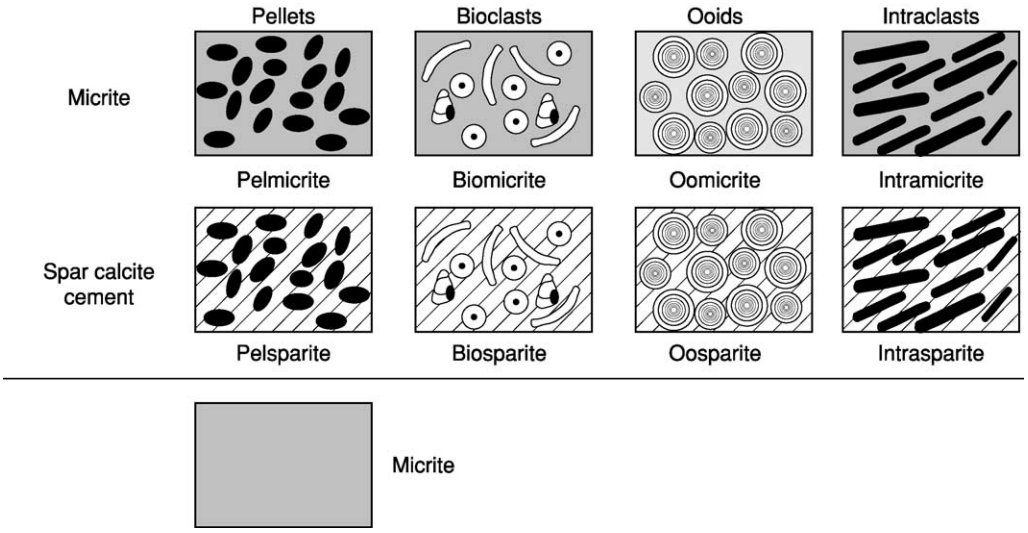


Figure 4 Folk's classification system for carbonate rocks.

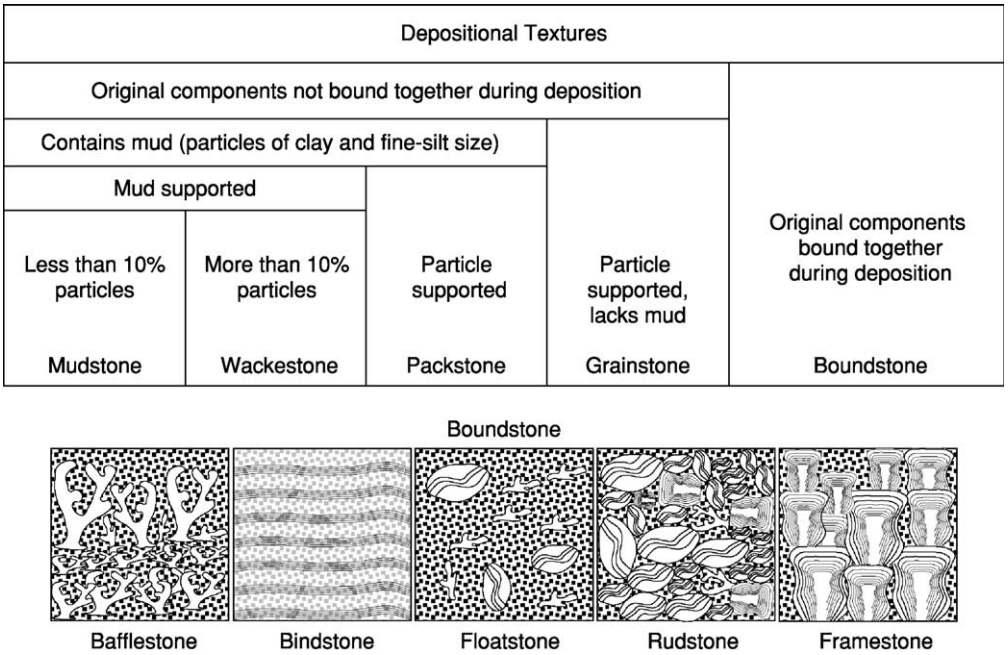


Figure 5 Dunham's classification system for carbonate rocks. The diagrams illustrating the various structures are labelled using the terminology subsequently introduced for different types of boundstones.

subsequently modified with the introduction of terms such as boundstone, bafflestone, and bindstone, which are typically applied to reefal limestones. In some cases, prefixes (e.g., skeletal grainstone) are used to signify the dominant allochem.

Marine Carbonate Deposits

Distribution and Composition

Depending on local conditions, many different minerals can be precipitated from marine waters. Seawater contains many salts, but under conditions of increasing temperature, $CaCO_3$ is one of the first compounds to be precipitated, either as calcite or aragonite. In modern oceans, carbonate sediments have the potential of forming anywhere and accumulating wherever the water depth is shallower than the carbonate compensation depth (CCD) and where there is little or no influx of siliciclastic sediments. In areas where large volumes of siliciclastic sediments are being introduced (e.g., where the Mississippi River empties into the Gulf of Mexico), the carbonate sediments, which form at a much slower rate, will be masked to the extent that they will appear to be non-existent. Extensive areas of carbonate sediment accumulation are found on shallow-water shelves (typically <30 m deep) that receive little or no influx of siliciclastic sediment, or on isolated banks that are surrounded by deep oceanic waters. In modern oceans, the mineralogy of the carbonate sediment is largely a function of the water temperature and pressure. Thus, aragonite dominates sediments forming in shallow tropical seas whereas calcite dominates temperate waters and the deeper, cooler waters in the tropics.

Depositional Systems

Marine carbonate sediments typically accumulate on a platform that develops as sediments accumulate in an area of subsidence. These depositional systems, which commonly cover hundreds of square kilometres, are referred to as a carbonate 'shelf' or a 'bank'. A carbonate shelf has a continental landmass on one side and deep oceanic waters on the other side, whereas a bank is completely surrounded by deep oceanic water (Figure 6). Carbonate shelves are classified as a ramp, an unrimmed shelf, or a rimmed shelf according to the slope of their depositional surface and the nature of their margins (Figure 6). A ramp slopes gradually (typically $<1°$) from the shoreline to the deep ocean with no appreciable change in slope. In contrast, rimmed and unrimmed shelves have a distinct break in slope on their oceanic margin (Figure 6). A rimmed shelf has islands, sand shoals, or actively growing reefs along their oceanic margins whereas an unrimmed shelf is devoid of such structures (Figure 6).

The margins of many modern tropical platforms are commonly characterized by large, robust coral reefs, especially on the windward coasts (Figure 1A). Spur and groove structures develop on the oceanward side of the reef crest. Unrimmed shelves or ramps are usually found on the leeward side of tropical islands and in temperate zones. The presence of islands, sand shoals, and/or reefs along the margin of a rimmed shelf significantly dampens the strength of any waves that encroach from the open ocean. As a result, low-energy environments and little current movement generally characterize the shelf. Only major storms and hurricanes will interfere with these tranquil settings, by generating high waves and strong currents for short periods of time. In stark contrast, ramps and unrimmed shelves have open communication with the ocean and therefore receive the full impact of onshore waves and currents. The fact that ocean waves move freely across the shelves with little or no impediment means that ramps and open shelves are characterized by high-energy conditions, complex arrays of nearshore facies, and physical transportation of sediment across the shelf.

The 'carbonate factory', where most carbonate sediment is generated, is typically located on the shallow, illuminated parts of the shelves. Conditions are ideal for the precipitation of carbonate sediment directly from seawater and for the growth of animals and plants that are the sources of aragonite and/or calcite skeletons (Figure 2F, G, and H). Most of the sediment stays in or close to the place where it formed. Nevertheless, storm-generated waves and currents can move sediment shoreward onto the

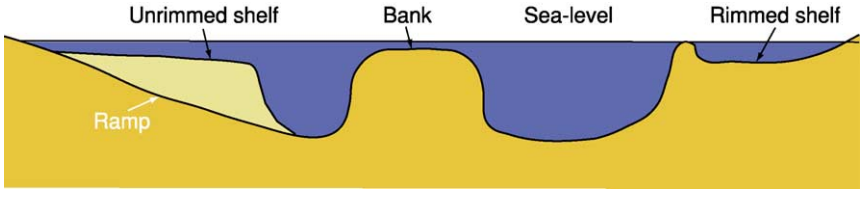


Figure 6 Schematic diagram showing differences between rimmed and unrimmed shelves, a ramp, and a bank.

intertidal and supratidal flats or seaward into the open ocean. Viewed in this context, it is apparent that the sediment production capacity of the carbonate factory carries important ramifications for sediment deposition throughout the carbonate depositional realm.

Factors critical to the health of the carbonate factory include water depth, water temperature, water circulation, oxygenation, salinity, and light penetration. In particular, these parameters control and strongly influence the growth and development of the plants and animals that contain the calcareous skeletons that will eventually become part of the sediment. Shallow, well-oxygenated, warm waters of normal salinity are, for example, ideal for the growth of many plants and animals, including calcareous algae, corals, bivalves, and gastropods. Conversely, many plants and animals cannot survive low or high salinities, water that is too cool or too hot, or poor circulation. Adverse environmental conditions commonly result in a biota formed of numerous individuals of a few taxa. Similarly, light penetration plays a major role in controlling the distribution of plants (e.g., calcareous algae and *Thalassia*) or animals (e.g., corals) that have a symbiotic relationship with plants (e.g., algae).

In many parts of the world, extensive carbonate deposits are accumulating in the intertidal (between low- and high-tide levels) and supratidal (above high-tide level) environments (**Figure 2A and B**). Conditions in these areas are highly variable because they are inundated by marine waters during high tide and storms, but are exposed to the atmosphere and rainwater during low tide. Such exposure means that climatic conditions play a critical role in the formation of sediments in these areas, especially in the supratidal area that is commonly exposed for extended periods of time. In hot, dry climates, evaporation dominates to the extent that evaporitic minerals, such as gypsum and halite, are produced. In humid climates, evaporates are rare and any that do form will dissolve during subsequent rainstorms. Under these conditions, the tidal flat commonly becomes covered with thick, leathery microbial mats and stromatolites that are formed of various consortiums of bacteria, fungi, and cyanobacteria (**Figure 2B**). The intertidal flats are complex depositional systems because meandering tidal creeks commonly cut them and isolated ponds are common (**Figure 2A**). Small-scale arrays of subtidal, intertidal, and supratidal microenvironments are associated with each creek and pond despite the fact that they are located on the intertidal flats. Such complex environmental diversity means that these depositional systems are usually characterized by complex and rapid vertical and lateral facies changes and a biota that is usually characterized by large numbers of only a few species. Little sediment (apart from evaporites) actually forms on the intertidal and supratidal flats. Instead, most of the sediment forms in the nearby lagoons and is carried onshore during severe storms and hurricanes.

Sediment that is transported in an oceanward direction will settle in the forereef area, where it may be transported downslope under the effects of gravity. As a result, carbonate sediments originally formed in shallow water may accumulate at considerable depth on the forereef slope or, in some cases, on the basin floor. Collapse of the reef, whether due to instability caused by overgrowth of the reef or to earthquake activity, will produce large blocks of reefal limestones (commonly the size of houses) that will eventually come to rest in deep water on the basin floor.

Sedimentation in marine systems is a dynamic process that is ultimately controlled by the balance between seafloor subsidence, sedimentation, and sea-level changes. The carbonate factory will continue to function while conditions conducive to plant and animal growth are maintained. A sudden increase in water depth, irrespective of its cause, will shut down the carbonate factory and in extreme cases will 'drown' the platform. Similarly, a decrease in water depth can have profound affects on the depositional system because it may lead to the exposure of some reefs and to the seaward extension of the tidal flats, making the subtidal conditions less conducive to the growth of plants and animals.

Reefs and Mounds

A reef, a positive seafloor structure constructed by animals and plants, is capable of withstanding strong currents and waves. In modern oceans, reefs are constructed mostly by coral-dominated communities and are typically found in shallow, well-illuminated tropical waters that are conducive to rapid coral growth. A mound, in contrast, is typically smaller and usually develops in more tranquil settings, with growth being mediated by fewer and more delicate creatures.

Modern reefs vary in size from small patch reefs (**Figure 2H and 3B**) that may be only a few metres in diameter and include few coral heads, to immense structures, such as the Great Barrier Reef off the north-east coast of Australia, that are thousands of kilometres long and are formed of numerous corals along with various other animals and plants. Such large reefs are important because they separate the open ocean from the landward shelf. Atolls, which are circular to ovate structures that enclose a central lagoon, are formed where coral reefs have colonized and grown upwards from the rim of a subsiding

volcano. The construction and rigidity of a reef, irrespective of its size and morphology, are controlled by the growth of the animals/plants and the deposition of sediment and/or precipitation of cements in cavities between those formative organisms. Destructive processes, which offset the constructive processes, include bioerosion and physical damage caused by high-energy events such as storms and hurricanes. Not all reefs have been formed by corals. Archaeocyathids (Cambrian), sponges (Ordovician), stromatoporoids (Silurian, Devonian), tabulate corals (Ordovician, Silurian), phylloid algae (Pennsylvanian), and rudist bivalves (Cretaceous), for example, have all been responsible for the construction of large reefal structures at different times throughout the Phanerozoic.

Mounds, which vary from low-relief lenses to mounds with slopes up to 40°, are generally categorized as microbial mounds, skeletal mounds, or mud mounds. Their genesis is poorly understood because the origin of the constituent mud is poorly constrained, the organisms that mediated their formation are typically poorly preserved or absent, and there are no modern examples of mud mounds. Growth and development of microbial mounds were typically mediated by cyanobacteria, algae, diatoms, and other micro-organisms that can become calcified, trap sediment, and/or induce $CaCO_3$ precipitation. Skeletal mounds formed where organisms such as bryozoa, corals, stromatoporoids, sponges, and rudist bivalves induced local sedimentation through current baffling, sediment trapping and binding, and sediment stabilization.

Marine Sedimentary Processes

The formation of carbonate deposits through deposition and/or precipitation depends on the depositional environment and, in particular, whether animals and/or plants are present. Plants play many different roles in the formation of marine carbonate sediment. Calcareous algae, seagrasses, and mangroves, for example, play important roles in carbonate sedimentation. Carbonate sedimentation is influenced by the binding action of their roots, the baffling action of their leaves, and the sediment production by the epibionts that live on their leaves and/or roots. The roles played by each plant vary. Calcareous algae, with their calcareous skeletons, are major sediment producers but play a minimal role in terms of current baffling or sediment binding. Conversely, *Thalassia* ('turtle grass') has long-bladed leaves that reduce current strengths and thereby cause deposition of any sediment suspended in that current, as well as long complex rhizomes that bind the sediment in place and prevent erosion except by the strongest of storm-driven currents. *Thalassia* is also a major sediment source because the calcareous skeletons of the small animals and plants that live on their leaves become part of the sediment load following the death and decay of the leaves (**Figure 2F**). Mangroves (**Figure 2B**), found in coastal regions, are characterized by extensive prop-root systems that bind sediment in place and are extremely effective at reducing currents, thereby causing sediment deposition. Sediment production, however, is generally limited to those plants and animals that attach themselves to the roots of trees.

Micro-organisms commonly play a critical role in carbonate sedimentation. Various consortia of microbes can form microbial mats that cover vast areas of the seafloor and the intertidal flats. In many areas, these microbial communities mediate the growth and construction of stromatolites (**Figure 2G**) by trapping and binding sediment to the seafloor or by acting as nucleation sites for aragonite and/or calcite precipitation. Removal of the mats commonly leads to erosion and transportation of the underlying sediment. In many areas, the microbial communities mediate the growth and construction of stromatolites, which are highly variable in terms of their size and morphology.

Non-Marine Carbonates

In non-marine settings, carbonate deposits will form from any water that is supersaturated with respect to $CaCO_3$. Rapid degassing of CO_2 and/or evaporation commonly trigger such precipitation (**Figure 1B** and **E**). In most situations, abiotic precipitation, in contrast to biologically influenced precipitation, is more important. Nevertheless, coated grains (**Figure 1C**), bioclasts, peloids, and lithoclasts can form in many of these settings. Indeed, it is commonly difficult to distinguish between allochems that form in marine and non-marine settings if their depositional context is unknown.

The precipitation of aragonite, as opposed to calcite, in non-marine settings has commonly been attributed to water temperature, with calcite being precipitated from water that is cooler than 40°C and aragonite being precipitated from water that is warmer than 40°C. This relationship, however, is not universally true. Exceptions are found when extreme CO_2 degassing leads to aragonite precipitation or when slow ion delivery, caused by microbial biofilms or high viscosity, leads to calcite precipitation. In the vents of some hot springs ($T > 90$°C), alternating aragonite and calcite precipitation can take place in response to variations in the rate of CO_2 degassing, even though the water chemistry remains essentially

constant. Precipitation under conditions of rapid CO_2 degassing and/or evaporation commonly produces bizarre and unusual crystals. Trigonal, dendritic, skeletal (i.e., hollow), or platy calcite crystals, for example, are found in many deposits that have formed around the vents of hot springs and geysers. Many of these crystals are composite crystals, with the constituent subcrystals being readily apparent in high-magnification, high-resolution scanning electron microscope images.

Compared to the role of plants, animals, and other eukaryotes in forming marine carbonates, their role in forming non-marine carbonates is relatively minor. The distribution of the biota in these settings is controlled primarily by water temperature, water acidity, and, in some spring systems, the presence of elements that may be toxic. One of the main controls over the distribution of the biota is the maximum water temperature that each taxon of can tolerate (e.g., 45–50°C for fungi). In the simplest sense, however, the diversity of plants and other organisms tends to decrease with increasing temperature. In some cold-water streams, lakes, and spring systems, however, calcite is commonly precipitated around charophytes (freshwater green algae) and bryophytes (mosses, liverworts, and hornworts). On the sides of valleys or tufa dams in some cold-water spring systems, calcite precipitation around these species can lead to the construction of large fan-shaped deposits that stretch from one side of a valley to another and cause ponding on the upstream side. Microbial communities formed of cyanobacteria, bacteria, and/or fungi are common in many freshwater lakes, streams, and spring systems. As in the marine environments, the microbes can play a major role in $CaCO_3$ accumulation by providing nucleation sites for crystal growth, by inducing precipitation through modification of the physiochemical conditions in their surrounding microenvironment, and/or by trapping and binding sediment to the substrate. Such activity commonly leads to the formation of microbial mats and/or stromatolites that are akin to those found in marine environments.

Diagenesis

Carbonate diagenesis is primarily driven by the fact that carbonate minerals are highly reactive to changes in temperature and pressure conditions, especially in the presence of vast quantities of water. Dissolution, cementation, inversion, recrystallization, and replacement, which are commonly triggered by pressure and temperature changes, may lead to the complete transformation of carbonate sediments and limestones. In many cases, the original components of the limestones will be obliterated, and porosity and permeability are either created or destroyed. Dolomitization, which is still poorly understood, is commonly pervasive, with thick successions of limestone being replaced by dolomite. Carbonate diagenesis will start on the seafloor and continue until the rocks have been buried to depths at which metamorphism takes place. In general, the factors that control diagenesis are poorly understood because they operate on a microscale in settings that cannot be directly observed or monitored. Much more is known about diagenesis on the seafloor and in near-surface settings than is known about diagenesis that takes place at depth, simply because the former settings are much more amenable to observation and monitoring. Diagenesis on the seafloor (e.g., hardground formation, reef lithification) and along the shoreline (e.g., beachrock) typically involves the precipitation of aragonite and/or calcite cements directly from seawater. Once exposed in the vadose zone (i.e., above the water table), carbonate sequences are prone to significant changes. Climate plays a critical role because maximum carbonate diagenesis takes place where vast quantities of water flow through the rocks under high-temperature conditions. Thus, carbonate rocks located in hot, humid climates tend to undergo more rapid diagenesis than do those in cool, dry climates. The surface and subsurface landforms associated with karst terrains provide clear evidence of the effects that surface and near-surface diagenesis has on carbonates. Dissolution is mediated by the weak carbonic acid that is formed as rainwater absorbs CO_2 from the atmosphere and from decaying vegetation on the ground. Aragonitic components are either dissolved, producing fossil-mouldic porosity (**Figure 3C**), or are transformed to calcite. Dissolution of aragonite and calcite means that the groundwaters commonly become supersaturated with respect to $CaCO_3$, allowing precipitation of aragonite and/or calcite cements to take place in cavities in which suitable physiochemical conditions exist. Diagenesis in the freshwater phreatic zone, which depends on water chemistry, may involve dissolution or cementation. Under conditions of deep burial, the rocks are subjected to considerable overburden pressures that may lead to dissolution and the formation of stylolites.

Many ancient carbonate successions have been pervasively dolomitized (**Figure 3F**) by processes that are still not well understood. All dolomite models, irrespective of their mechanics, require a source of Mg, a mechanism for transporting the Mg to the dolomitization site, and a dolomitization site that is physiochemically conducive to dolomite formation. Some models, such as the reflux model, involve near-surface

dolomitization that is mediated by brines with a high Mg:Ca ratio. Other models operate on the premise that dolomitization took place in the mixing zone, wherein mixtures of freshwater and seawater were responsible for dolomite formation. At the other extreme are models that have dolomitization taking place after deep burial. Tertiary 'island dolomites' are commonly attributed to dolomitization by slightly modified seawater in near-surface settings. It seems probable, however, that pervasive dolomitization can be achieved under many different conditions and that it is impossible to explain all dolomitization by means of a single model.

See Also

Biosediments and Biofilms. **Carbon Cycle**. **Diagenesis, Overview**. **Fossil Invertebrates:** Bryozoans; Corals and Other Cnidaria. **Fossil Plants:** Calcareous Algae. **Microfossils:** Foraminifera. **Sedimentary Environments:** Carbonate Shorelines and Shelves; Reefs ('Build-Ups'). **Sedimentary Rocks:** Mineralogy and Classification; Dolomites; Limestones.

Further Reading

Adams AE, MacKenzie WS, and Guilford C (1984) *Atlas of Sedimentary Rocks under the Microscope*. England: Longman.

Bathurst RGC (1975) *Carbonate Sediments and Their Diagenesis*. Amsterdam: Elsevier.

Budd DA, Saller AH, and Harris PM (1995) *Unconformities and Porosity in Carbonate Strata. American Association of Petroleum Geologists, Memoir 63*. Tulsa, OK: American Association of Petroleum Geologists.

Loucks G and Sarg JF (1993) *Carbonate Sequence Stratigraphy. American Association of Petroleum Geologists, Memoir 57*. Tulsa, OK: American Association of Petroleum Geologists.

Purser B, Tucker M, and Zenger D (1994) *Dolomites. A Volume in Honour of Dolomieu. International Association of Sedimentologists, Special Publication Number 21*. Oxford: Blackwell.

Scholle PA, Bebout DG, and Moore CH (1983) *Carbonate Depositional Environments. American Association of Petroleum Geologists, Memoir 33*. Tulsa, OK: American Association of Petroleum Geologists.

Tucker ME and Bathurst RGC (1990) *Carbonate Diagenesis. International Association of Sedimentologists Reprint Series*, vol. 1. Oxford: Blackwell.

Tucker ME and Wright VP (1990) *Carbonate Sedimentology*. Oxford: Blackwell.

Vacher HL and Quinn TM (eds.) (1997) *Geology and Hydrogeology of Carbonate Islands. Developments in Sedimentology 54*. Amsterdam: Elsevier.

Walker RG and James NP (1992) *Facies Models. Response to Sea Level Change*. Ottawa: Geological Association of Canada.

Wilson JE (1975) *Carbonate Facies in Geologic History*. New York: Springer-Verlag.

Chromates

P A Williams, University of Western Sydney, Penrith South DC, NSW, Australia

Introduction

Naturally occurring chromate minerals are confined to the oxidized zones of base metal orebodies and to extremely oxidizing desert environments. Under such conditions, chromium reaches its highest oxidation state (+6). The only commercially significant ore mineral of chromium is the primary spinel chromite, $FeCr_2O_4$, which contains Cr(III). Chromium (VI) can form under very oxidizing conditions, and nearly all of its minerals contain the simple tetraoxochromate(VI) or chromate ion, CrO_4^{2-} (see Table 1). Two exceptions are chrombismite, $Bi_{16}CrO_{27}$, a complex lattice oxide, and lopezite, naturally occurring potassium dichromate. The anion in the latter is generated reversibly by the acid-assisted dimerization of chromate, as shown in eqn [1]:

$$2CrO_4^{2-}(aq) + 2H^+(aq) \leftrightarrow Cr_2O_7^{2-}(aq) + 2H_2O \quad [1]$$

Geographical Distribution of Chromates

The chromates listed in Table 1 occur in two quite disparate settings: deserts and oxidized zones of a few base metals. Desert environments include those in Jordan (the only known locality for hashemite) and, more spectacularly, the Atacama Desert of Chile. In the latter case, chromates are associated with the caliche nitrate deposits, which besides nitrate contain other highly oxidized species, such as bromate, iodate, periodate, perchlorate, and native iodine. The persistence of highly oxidizing species in this environment is due to the extreme paucity of organic material with

Table 1 Chromate(VI) minerals

Type	*Name*	*Composition*
Simple chromates and dichromates		
	Chromatite	$CaCrO_4$
	Crocoite	$PbCrO_4$
	Hashemite	$BaCrO_4$
	Lopezite	$K_2Cr_2O_7$
	Tarapacaite	K_2CrO_4
Complex chromates		
	Cassedanneite	$Pb_5(CrO_4)(VO_4)_2 \cdot H_2O$
	Deanesmithite	$Hg_5S_2O_2(CrO_4)$
	Edoylerite	$Hg_3S_2(CrO_4)$
	Embreyite	$Pb_5(CrO_4)_2(PO_4)_2 \cdot H_2O$
	Dietzeite	$Ca_2(CrO_4)(IO_4)_2 \cdot H_2O$
	Fornacite	$CuPb_2[(Cr,As,P)O_4]_2OH$
	Hemihedrite	$Pb_{10}Zn(CrO_4)_6(SiO_4)_2F_2$
	Iranite	$Pb_{10}Cu(CrO_4)_6(SiO_4)_2(F,OH)_2$
	Macquartite	$Pb_7Cu_2(CrO_4)_4(SiO_4)_2(OH)_2$
	Phoenicochroite	Pb_2OCrO_4
	Santanaite[a]	$Pb_{11}O_{12}CrO_4$
	Vauquelinite	$Pb_2Cu(CrO_4)(PO_4)(OH)$
	Wattersite	$Hg_5O_2CrO_4$
	Yedlinite	$Pb_6Cl_6O_2CrO_4 \cdot 2H_2O$

[a]Contains both Pb(II) and Pb(IV).

which they would ordinarily react. Dietzeite, chromatite, lopezite, and tarapacaite have all been found in certain of the caliche deposits. These minerals are either freely soluble or decompose in water and they owe their preservation as well to the extreme aridity of the Atacama region.

Chromate minerals that form sparingly in the oxidized zones of base metal deposits form the bulk of the minerals listed in Table 1, and many of these chromate species are involved in complex solid solutions with both constituent cations and anions. Molybdate substitutes for chromate in the fornacite lattice and the alternate end-member molybdofornacite is recognized as a separate species. Nevertheless, all of the chromates are rare and none is a used as a source of chromium. One exception to this is the mineral crocoite, normal lead(II) chromate, which has been used as a paint pigment. The element chromium was first extracted and identified in 1796 by the French chemist Louis Vauquelin in samples of crocoite from Beresov, Russia. Fine specimens of crocoite together with the rarer phase, phoenicochroite, were recovered from Beresov, and the mineral has been found at a number of other localities worldwide. By far the most spectacular specimens have come from the Magnet mine and other deposits in the Dundas area of north-western Tasmania, Australia. Single crystals weighing several hundred grams each were recovered from the Red Lead mine, Dundas, and the neighbouring Adelaide mine has been a prolific producer of fine specimens for several decades. Crocoite in these deposits is formed by the interaction of chromium-bearing solutions derived from weathering serpentinites with oxidized, fault-hosted argentiferous galena ores. Associated with the crocoite is yellow cerussite, $PbCO_3$; the yellow colour is thought to arise from the substitution of minute amounts of chromate in the lattice.

See Also

Minerals: Definition and Classification; Molybdates; Nitrates; Tungstates.

Further Reading

Anthony JA, Bideaux RA, Bladh KW, and Nichols MC (2003) *Handbook of Mineralogy. Volume 5. Borates, Carbonates, Sulfates, Chromates, Germanates, Iodates, Molybdates, Tungstates, etc., and Organic Materials*. Tucson, AZ: Mineral Data Publishing.

Baes CF, Jr and Mesmer RE (1986) *The Hydrolysis of Cations*. Malabar, FL: Krieger Publishing Company.

Bard AJ, Parsons R, and Jordan J (1985) *Standard Potentials in Aqueous Solution*. New York: Marcel Dekker.

Ericksen GE (1981) *Origin of the Chilean Nitrate Deposits. United States Geological Survey Professional Paper, 1118*. Washington, DC: US Geological Survey.

Gaines RV, Skinner HCW, Foord EE, Mason B, and Rosenzweig A (eds.) (1997) *Dana's New Mineralogy: The System of Mineralogy of James Dwight Dana and Edward Salisbury Dana*, 8th edn. London: Wiley Europe.

Haupt J (1988) Minerals of western Tasmania. In: *The Mineralogical Record*, vol. 19, pp. 381–388. Tucson, AZ: Mineralogical Record Inc.

Mandarino JA (1999) *Fleischer's Glossary of Mineral Species 1999*, 8th edn. Tucson, AZ: Mineralogical Record Inc.

Roberts WL, Campbell TJ, and Rapp GR, Jr (1990) *Encyclopedia of Minerals*, 2nd edn. New York: Van Nostrand Reinhold.

Williams PA (1990) *Oxide Zone Geochemistry*. Chichester: Ellis Horwood.

Williams SA (1974) The naturally occurring chromates of lead. *Bulletin of the British Museum (Natural History)* 8: 379–418.

Feldspars

R A Howie, Royal Holloway, University of London, London, UK

The minerals of the feldspar group are the most abundant minerals in the Earth's crust and are also found on the Moon and in meteorites. Their ubiquity and their wide range of chemical composition has led to their use as a primary tool in the classification of the igneous rocks (*see* **Rocks and Their Classification**). In the great majority of these rocks, whether acid, alkaline, intermediate, or basic, the feldspars are the major constituents; they are absent only from some ultrabasic and rare alkaline rocks. They are important constituents of simple pegmatites and are common in mineral veins. Feldspars are major constituents of most gneisses and schists, and occur also in any thermally as well as regionally metamorphosed rocks. Although susceptible to alteration and weathering, the feldspars are second in abundance to quartz in the arenaceous sediments (*see* **Sedimentary Rocks:** Sandstones, Diagenesis and Porosity Evolution), in which they occur as detrital grains. It is only in the pelitic, and to a greater degree in the carbonate rocks that they are of relatively minor importance.

Nomenclature

Most feldspars may be classified chemically as members of the ternary system $NaAlSi_3O_8$ (albite, Ab)–$KAlSiO_8$ (orthoclase, Or)–$CaAl_2Si_2O_8$ (anorthite, An) (*see* **Sedimentary Rocks:** Mineralogy and Classification). Compositions between $NaAlSi_3O_8$ and $KAlSi_3O_8$ are referred to as alkali feldspars and those between $NaAlSi_3O_8$ and $CaAl_2Si_2O_8$ as plagioclase feldspars. These two series each generally contain less than 5–10% of the other. The distinction between alkali and plagioclase feldspars near the $NaAlSi_3O_8$ composition is an arbitrary one.

To achieve a proper understanding of feldspar relationships it is necessary to characterise them not only by chemical composition, but also according to their structural state, the latter depending on the temperature of crystallisation and on subsequent thermal history (*see* **Igneous Processes**). Feldspars that retain a crystal structure appropriate to their high-temperature formation are called high-temperature feldspars; most feldspars in volcanic rocks are of this type. Low-temperature feldspars are those with structures appropriate either to crystallization at lower temperatures, or to slow cooling from elevated temperatures as, for example, in plutonic igneous rocks.

The difference between high- and low-structural state in a feldspar can involve the degree of ordering of Al and Si atoms between distinct tetrahedral sites. Feldspars can thus be classified as:

i. homogeneous high-temperature disordered feldspars; or
ii. ordered feldspars, in most of which phase separation (i.e., phases with appreciably different Na, K, or Ca content) occurs on a macro-, micro- or submicroscopic scale.

The nomenclature of the high-temperature feldspars is based on crystal symmetry. The sanidines (Figure 1A), mostly Or_{50}–Or_{80}, have monoclinic symmetry; the anorthoclase field embraces those feldspars that are triclinic at room temperature but

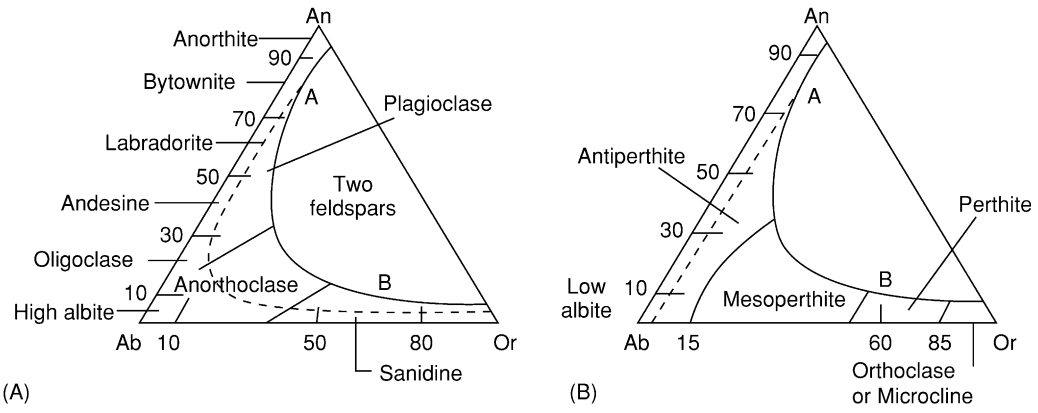


Figure 1 Feldspar nomenclature for (A) disordered ternary feldspars; (B) ordered ternary feldspars in which phase separation is at or below the resolution of the optical microscope. Composition in mol.%. Curve A–B limit of ternary solid solution.

which may have possessed monoclinic symmetry at crystallization temperatures. For pure $NaAlSi_3O_8$, both low- and high-temperature forms are triclinic, but the remainder of the plagioclases have triclinic symmetry both at room temperature and at their temperature of crystallization.

In the alkali feldspars Al/Si and Na/K ordering lead to marked changes in symmetry, optics, morphology, and exsolution textures. In the compositional range Or_{15}–Or_{85}, exsolution gives rise to cryptoperthite (in which exsolution is on a submicroscopic scale), or to perthite or microperthite (Figure 1B). Microcline is triclinic, and may show varying degrees of order. The name orthoclase is used to describe intermediately ordered K-rich feldspars. Adularia is a low-temperature ordered potassium feldspar with a characteristic habit (large prism faces) and paragenesis ('alpine' veins).

A purely chemical definition of a plagioclase can be given in terms of Ab–An 'molecular' percentages, but specific names are used to denote the six compositional ranges into which the series has been divided. Thus, albite, oligoclase, andesine, labradorite, bytownite, and anorthite refer to the An percentages 0–10, 10–30, 30–50, 50–70, 70–90 and 90–100, respectively. Position within a chemical range may be further indicated by such terms as sodium-rich (or sodic) oligioclase or calcic labradorite, etc. Many plagioclases consist of a fine lamellar intergrowth of two phases, one more Ca-rich than the other; those within the An_5–An_{20} range are called peristerites and their surfaces may exhibit iridescence. Other ranges for such intergrowths may be found at An_{45}–An_{70} (Bøggild) which shows labradorescence and at An_{75}–An_{85} (Huttenlocher), also with associated iridescence effects.

Most perthites (named from Perth, Quebec, an early locality) are exsolution intergrowths of sodium-rich feldspar in a potassium-rich feldspar host; an intergrowth of K-rich feldspar in a plagioclase host is called antiperthite. The name feldspar as originally given was feldtspat and is believed to be in reference to the presence of this spar (spath) in tilled fields (Swedish: feldt or fält) overlying granite (*see* **Igneous Rocks:** Granite), rather than to the German Fels, meaning rock.

Structure

The feldspars are framework silicates with tetrahedra of $(Si,Al)O_4$ linked to one another (by shared oxygens) in all directions rather than in chains or sheets. Although discrete chains of tetrahedra do not occur in the structure, its nature may be more readily understood by considering the atomic arrangement as the linking of chains in two directions perpendicular to their length. The chains themselves are formed by the linking of horizontal rings of four tetrahedra, as shown in Figure 2. When viewed in the direction of the chain axis, a horizontal ring appears approximately, as shown in Figure 2B, and this can be further simplified by its representation, as in Figure 2C. The configuration shown in Figure 2 is often described as a 'double crankshaft'; the upwards- and downwards-pointing tetrahedra in a horizontal ring are labelled U and D, respectively. In the actual structure, the rings are considerably distorted; successive horizontal rings of a chain are related by vertical glide planes passing through their centres; the view down the chain axis may be idealized, as in Figure 2D, the first, third, fifth, etc. rings being represented by thick and the even number rings by thin lines. At the level of the first ring, a network of oxygen linkages is formed, producing a plane of four-membered and eight-membered rings of tetrahedra. The eight-membered rings are of two kinds, UUUUDDDD and DUUDUDDU, characteristic of feldspars, as distinct from other framework silicates, (*see* **Minerals:** Other Silicates) such as zeolites.

Twinning in feldspars is common and can occur in three different ways: (i) as a primary phenomena during crystal growth; (ii) as glide twinning induced by deformation; or (iii) on thermal transformation to lower symmetry. There are several twin laws shown

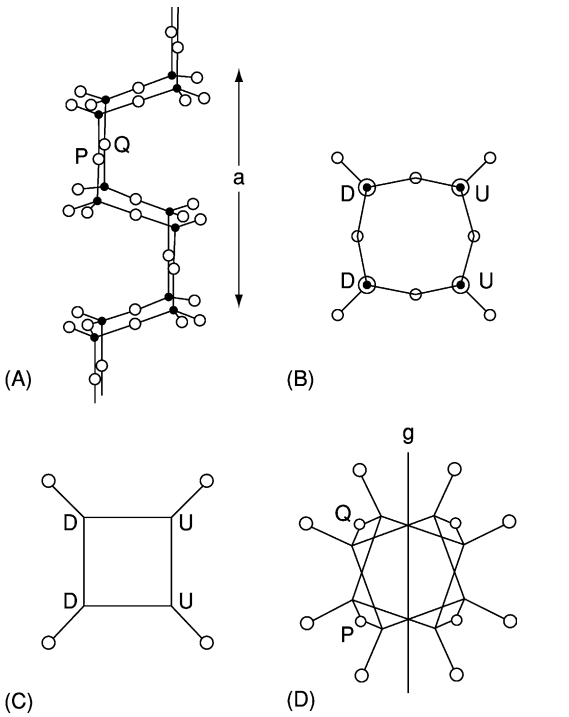


Figure 2 Idealized illustrations of the feldspar 'chain'. In (B) and (C) U and D indicate tetrahedra pointing up and down, respectively.

by feldspars; basically these may appear as single twins or as repeated or polysynthetic twins, an example of the latter being the very frequent twinning on the albite law in triclinic feldspars, giving the characteristic bands of alternating birefringence with an (010) composition plane. Measurement of the positions of optical extinction of these twins in plagioclase can be used to give an indication of the composition of the crystals.

Optical and Physical Properties

All the common members of the feldspars have D 2.55–2.5, with H 6–61/2. They are generally colourless, white or grey, but may be yellow, red, or green; colourless in thin section. Refractive indices are low, alkali feldspars having values <1.54 and plagioclase in the range 1.53–1.59 (the values increasing with increase in An content); the birefringence ranges 0.006–0.013. Single or multiple twins are common.

Alkali Feldspars

$(K,Na)[(AlSi_3O_8]$. Monoclinic or triclinic. K-feldspar melts incongruently at ~1150° to leucite ($KAlSi_2O_6$) under anhydrous condition at 1 atm (Figure 3). The eutectic temperature (E) between K-feldspar and tridymite is ~990°C. From a melt of composition A, leucite (B) is the first phase to crystallize, and continues to do so as the liquid composition moves along the liquidus curve to the reaction point (R). At this temperature, the liquid reacts with the leucite transforming it to K-feldspar. As the temperature continues to fall, the composition of the melt moves along the curve RE with K-feldspar continuing to crystallize; at the eutectic (E) crystallization is complete and the whole mass solidifies as a mixture of K-feldspar and tridymite. Later experimental work has shown that at 4 kbar the eutectic point is lowered to ~735°C.

The refractive indices for the alkali feldspar series increase steadily with increasing albite content (Figure 4), values for α and γ in the high-temperature series sanidine–anothoclase–high-albite being marginally lower. Microcline is distinguishable from most other alkali feldspars by its 'tartan' twinning in which two sets of twin lamellae are approximately at right angles to each other.

The K–Na feldspars are essential constituents of alkali and acid igneous rocks and are particularly abundant in syenites, granites, granodiorites, and their volcanic equivalents. In plutonic rocks, the alkali feldspar is usually orthoclase, microcline, microcline microperthite, or microcline perthite, whereas in volcanic rocks it is sanidine, anorthoclase cryptoperthite, or anorthoclase. Granophyric intergrowths, the bulk composition of which is close to the melting minimum in the orthoclase–albite–quartz–H_2O system, consisting of roughly equal amounts of quartz, K-feldspar, and Na-feldspar components, occur in the mesostasis of acid rocks of high-level origin.

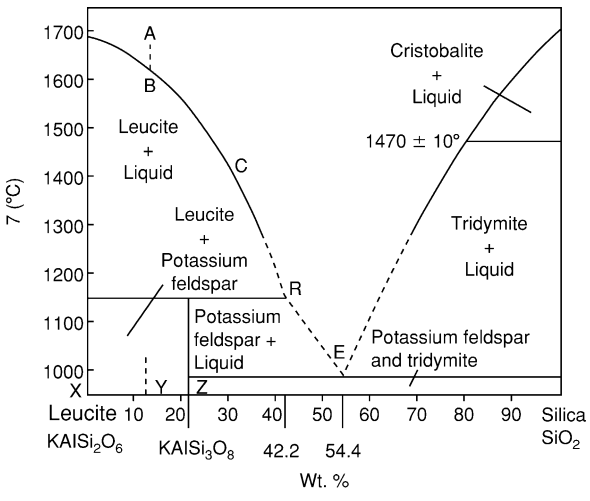


Figure 3 Equilibrium diagram of the binary system $KAlSi_2O_6$ (leucite)–SiO_2. Reaction point R represented approximately by $5\,KAlSi_2O_6(\text{leucite}) + 3\,KAlSi_3O_8 + 5\,SiO_2(\text{melt}) \rightarrow 8\,KAlSi_3O_8$.

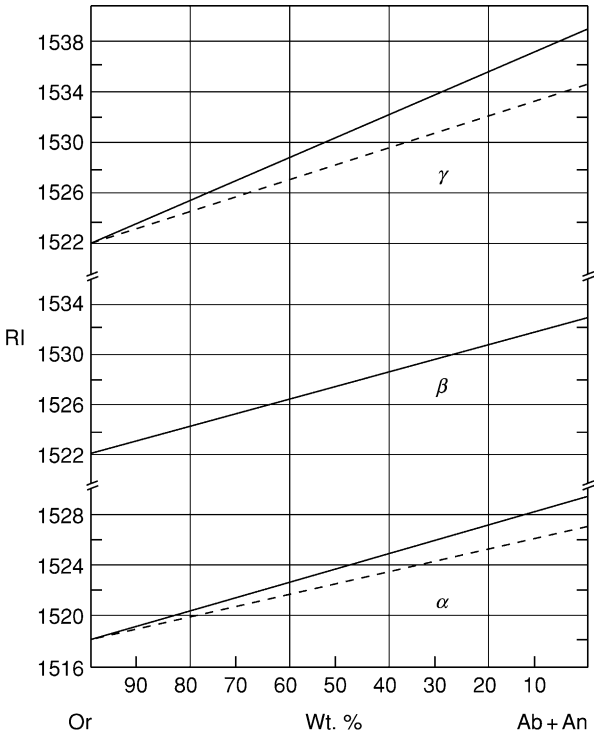


Figure 4 Variation of refractive indices with composition for the alkali feldspars; α and γ for the sanidine–anorthoclase–high albite series are shown as dashed lines, and those for the orthoclase (microcline)–low albite series as full lines. The β index is essentially the same in both series.

K-feldspar is a stable product of both high-grade thermal and regional metamorphism. It is a typical mineral of the sillimanite zone of metamorphism and, in rocks of argillaceous composition, does not appear at lower grades. Where K-feldspars released by weathering of igneous or metamorphic rocks have undergone rapid transportation and burial, they occur as relatively abundant detrital constituents in arkosic sediments. They may also develop as authigenic minerals occurring as rims on detrital feldspar, often in optical continuity.

Plagioclase

$Na[AlSi_3O_8]$—$Ca[Al_2Si_2O_8]$. Triclinic. The phase diagram for the plagioclase series was one of the first to be determined experimentally (Figure 5). It shows a solid solution without maximum or minimum, with the melting point of anorthite at 1550°C and that of albite at ~1100°C at $P = 1$ bar. A liquid of composition $Ab_{50}An_{50}$ (A) begins to crystallize at ~1450°C, the first crystal having the composition of ~$Ab_{82}An_{18}$(B). With further cooling under equilibrium conditions, both liquid and crystals change their composition along the liquidus and solidus, respectively, until at 1285°C the crystals reach a composition of $Ab_{50}An_{50}$(D) as the last of the liquid now of composition (C) is used up. This continuous change in composition of the plagioclase crystals with falling temperature occurs only if there is sufficient time for the earlier-formed crystals to react with the liquid: if there is insufficient time for this interchange of material, the crystals will be zoned. The resultant product will have an average composition of $Ab_{50}An_{50}$, but the inner core will be more calcic and the outer zones more sodic. With the addition of H_2O to the system, at $P_{H_2O} = 5$ kbar, the liquidus and solidus temperatures are over 300° lower than for the anhydrous system (Figure 5, curve 2). An even greater effect is shown on the addition of SiO_2 to the system at 2 kbar (curve 3).

The optical properties of the plagioclases are directly related to their anorthite content. Both the relief and birefringence are low and similar to quartz. A determinative chart is given in Figure 6; it is important to note that measurement of the refractive indices must be accurate to ±0.001 to obtain an accuracy of ± 2% An. The composition may also be determined by melting a small amount of material, quenching it to a glass and determining the refractive index of the glass, the refractive index for an isotropic glass being more easily determined than for a triclinic mineral and, for the plagioclases, having twice as great a rate of change with composition (Figure 6). This method can give the average composition for plagioclase which is strongly zoned or has exsolution intergrowths; also, the original structural state of the plagioclase is of no consequence.

In the majority of twinned plagioclase crystals, the composition plane is parallel to the crystal length.

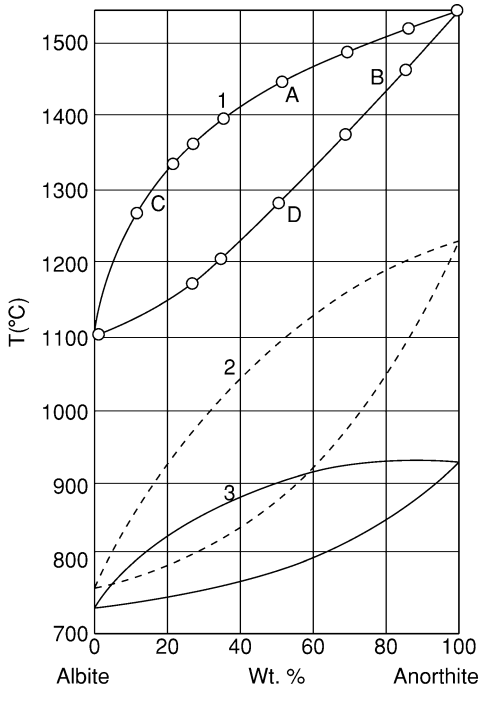


Figure 5 Equilibrium diagram of the plagioclase feldspars: (1) in anhydrous conditions, (2) at 5 kbar water pressure, (3) eutectic liquidus and solvus in the albite–anorthite–quartz–H_2O system at 2 kbar.

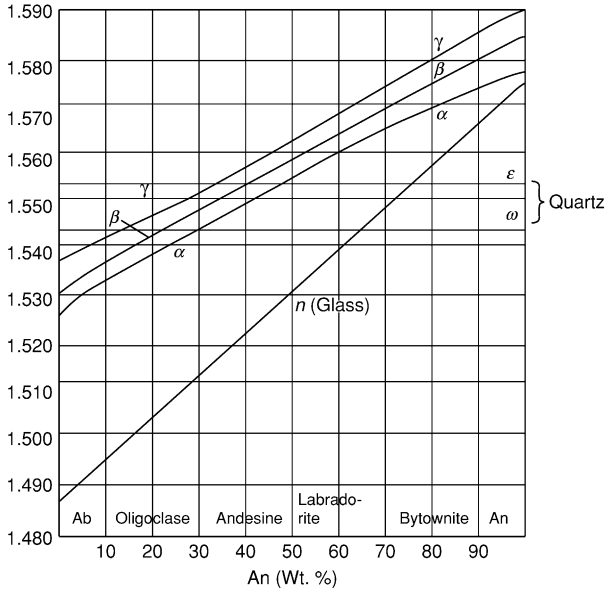


Figure 6 Refractive indices of the plagioclases. Values for glasses of plagioclase composition are also shown.

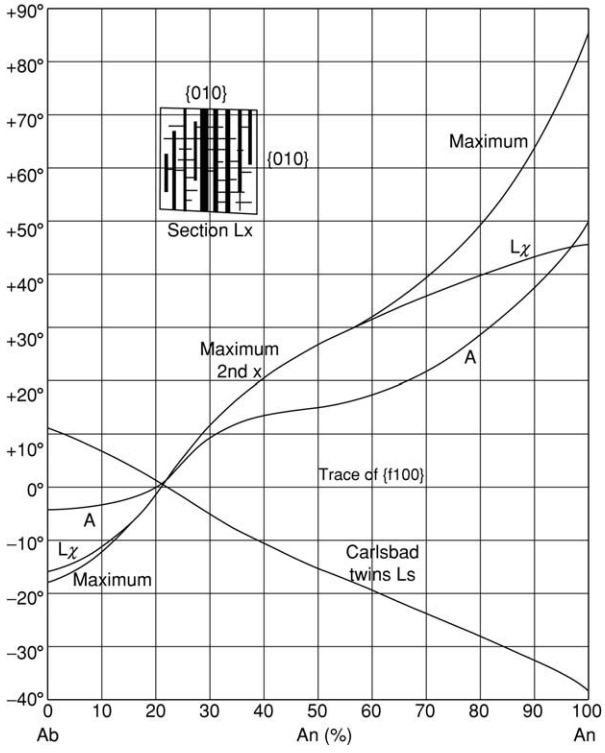


Figure 7 Extinction angles of plagioclase (with respect to α') in the 'symmetrical zone' and in sections normal to *x*.

The maximum extinction angles normal to {010} or in the so-called symmetrical extinction zone are diagnostic (Figure 7). In thin section, the alternate twin lamellae give symmetrical extinction angles on either side of the twin plane. Such sections may be recognised by the sharpness of the composition plane between albite twin lamellae, by the equal interference colours of the twin lamellae when the twin plane is parallel to the vibration direction of the polarizers, and by adjacent twin lamellae giving equal extinction angles on either side of the twin plane. Values which show more than a 5% divergence in the extinction angles for adjacent twin lamellae should be discarded, but lesser variations may be averaged; it is essential to take measurements from several (6–12) suitable grains, and the highest extinction angle must then be used.

Plagioclase is the most abundant mineral in the great majority of basic and intermediate lavas, in which it occurs both as phenocrysts and as a groundmass constituent. In basalts, the plagioclase phenocrysts usually have a wide homogeneous core of bytownite composition surrounded by a narrow zone of more sodic plagioclase. The broad cores of uniform composition indicate slow crystallization and it is clear that the growth of these crystals occurred before extrusion and final consolidation of the magma. Under plutonic conditions the first plagioclase to crystallize from most basic magmas, like that of basalts, has a bytownite composition. In basic igneous intrusions, plagioclase sometimes occurs in feldspar-rich bands formed by the accumulation of the primary precipitate plagioclase (the cumulus phase). In contrast, primary precipitate plagioclase may be absent from the ferromagnesian-rich bands in which the feldspar occurs as a product of the crystallization of intercumulus liquid. In the differentiated basic rocks of layered intrusions, the compositional range of the plagioclase is normally restricted to between An_{85} and An_{30}. In addition to the plagioclase-rich bands of layered intrusions, plagioclase also occurs as the only essential constituent of large masses of anorthosite; in these rocks the plagioclase may be bytownite, labradorite, or andesine in composition. Plagioclase feldspars are the main constituent of dolerites and many other hypabyssal rocks.

In metamorphic rocks, the composition of the plagioclase is generally related to the grade of the host rock. Thus, albite is the stable plagioclase in the chlorite and biotite zones of regional metamorphism, occurring in such rocks as chlorite-biotite-epidote-albite amphibolites and chlorite-albite schists; the anorthite component is not present in notable quantity until the garnet zone. Plagioclase between An_5 and An_{20} in composition is absent in low- to medium-grade schists; this compositional break (the peristerite gap) corresponds with the change in grade between the greenschist and the almandine amphibolite facies. In calc-silicate rocks of the amphibolite facies, clear compositional minima are seen corresponding also with the Bøggild and Huttenlocher gaps (Figure 8). The plagioclase feldspars of the intermediate and acid rocks of the granulite facies commonly are sodic andesines; thus in the rocks of the charnockite series, the majority of the plagioclase is between An_{30} and An_{35} in composition, although labradorite is present in some of the more basic charnockitic rocks. Plagioclase is not stable in the P–T environment of the eclogite facies; under these conditions the albite and anorthite components enter the compositions of omphacite and garnet, respectively. In the thermal metamorphism of calcareous sediments plagioclase again becomes increasingly calcic as the grade increases, with the production of anorthite marking a high grade of metamorphism.

In sedimentary rocks, plagioclase feldspars may occur via a variety of sources, including detrital minerals, authigenic minerals, and the products of volcaniclastic activity. In arkosic sediments, K-feldspar is usually dominant over plagioclase, but this varies according to the source rock and the weathering

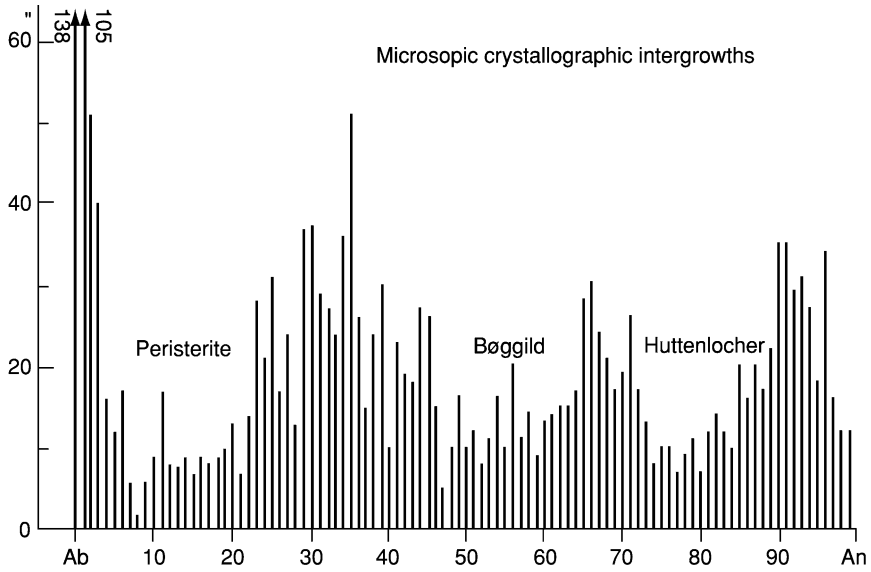


Figure 8 Frequency distribution of plagioclase compositions from 1883 microprobe results from 147 metamorphic calc-silicate rocks in the central Alps (after Wenk *et al.* (1991)).

conditions to which they have been subjected. In meteorites, a whole range of plagioclase compositions occur, but there are two main clusters at around An_{90} and An_{15}. In lunar rocks the great majority of feldspars are calcic plagioclase An_{90} to An_{95} representing the lunar anorthosites.

See Also

Igneous Processes. **Igneous Rocks:** Granite. **Minerals:** Definition and Classification; Other Silicates. **Rocks and Their Classification**. **Sedimentary Rocks:** Mineralogy and Classification; Sandstones, Diagenesis and Porosity Evolution. **Solar System:** Meteorites; Moon.

Further reading

Deer WA, Howie RA, and Zussman J (1992) *An Introduction to the Rock-Forming Minerals,* 2nd edn. London: Longman.

Wenk E, Schwander H, and Wenk H-R (1991) Microprobe analyses of plagioclases from metamorphic carbonate rocks of the central Alps. *European Journal of Minerals* 3: 181–191.

Feldspathoids

M D Welch, The Natural History Museum, London, UK

Structure

Feldspathoid structures consist of rings of AlO_4 and SiO_4 tetrahedra that are connected to form a fully polymerized three-dimensional framework enclosing cavities occupied by large interstitial monovalent or divalent cations (sodium, potassium, calcium, and caesium) and, in some cases, anions (chloride, sulphide, and sulphate). These cavities are aligned to form channels that pass through the structure. Compared with feldspars (*see* **Minerals:** Feldspars), the feldspathoids are silica poor.

Three main tetrahedral-framework topologies are found in the feldspathoid group, and these are represented by the minerals nepheline, leucite, and sodalite. They are distinguished primarily by their different ring types and by how these rings are connected to each other. The nepheline structure consists of rings of six tetrahedra, whereas the structures of leucite and sodalite consist of both four-membered and six-membered rings.

Nepheline Subgroup

Minerals in the nepheline subgroup have the chemical composition ABO_4, where A is Na, K, Ca, or vacant,

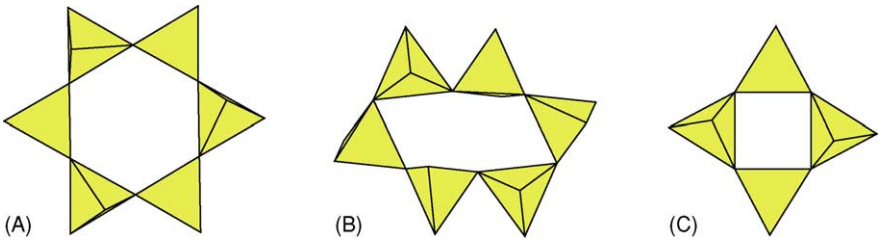


Figure 1 The different tetrahedral rings in feldspathoids. (A) The undistorted six-membered rings exhibited by the high-temperature form of tridymite. (B) The highly distorted six-membered rings exhibited by nepheline and the low-temperature form of tridymite. (C) The four-membered rings that occur in leucite and sodalite.

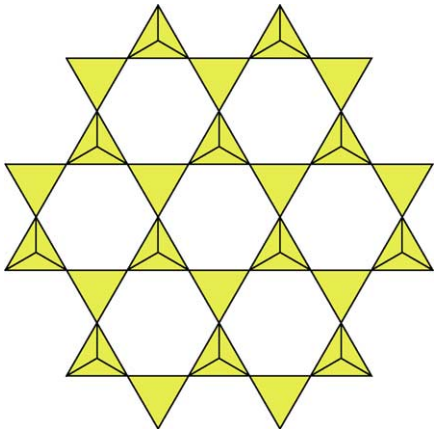


Figure 2 A sheet of undistorted hexagonal rings, as occur in the high-temperature form of tridymite.

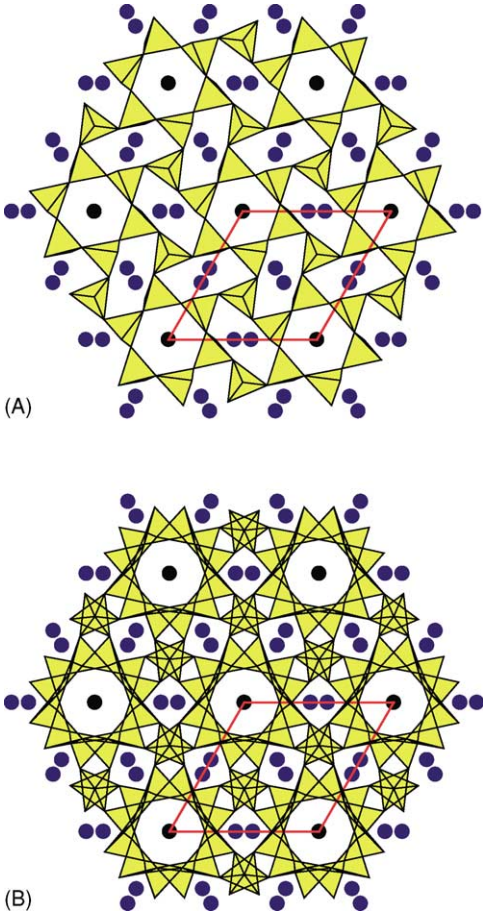


Figure 3 (A) The characteristic sheet of distorted and undistorted six-membered rings of AlO_4 and SiO_4 tetrahedra in the nepheline-type structure. Blue circles are sodium ions occupying smaller channel sites above and below the sheet, and black circles are potassium ions occupying the larger regular channels. The projection of the unit cell is shown in red. (B) A complete double sheet of the nepheline-type structure, showing the two kinds of channel.

and B is Al or Si. The tetrahedral framework of nepheline-type structures is related to that of the mineral tridymite (*see* **Minerals:** Quartz), which has two polymorphs: a high-temperature largely undistorted hexagonal framework collapses upon cooling to a low-temperature structure in which all the six-membered rings are highly distorted (Figure 1B). A sheet of undistorted hexagonal rings in the high-temperature form of tridymite is shown in Figure 2. Half the tetrahedra point up; the other half point down. In the nepheline structure these sheets are connected to each other by apical oxygen ions to form cavities that are aligned into channels running parallel to the crystallographic *c*-axis. Insertion of cations into the channels causes distortions of the framework that are similar to those associated with the thermally induced collapse of high-temperature tridymite to low-temperature tridymite. For this reason, minerals of the nepheline group are often referred to as having 'stuffed tridymite' structures. The common stoichiometry of natural nephelines is $Na_3KAl_4Si_4O_{16}$, which arises from the ordering of potassium and sodium in the two different channels (Figure 3).

Leucite and Sodalite Groups

The leucite and sodalite structures differ from the nepheline structure in that they also contain squares of four tetrahedra (Figure 1C). However, the ways in

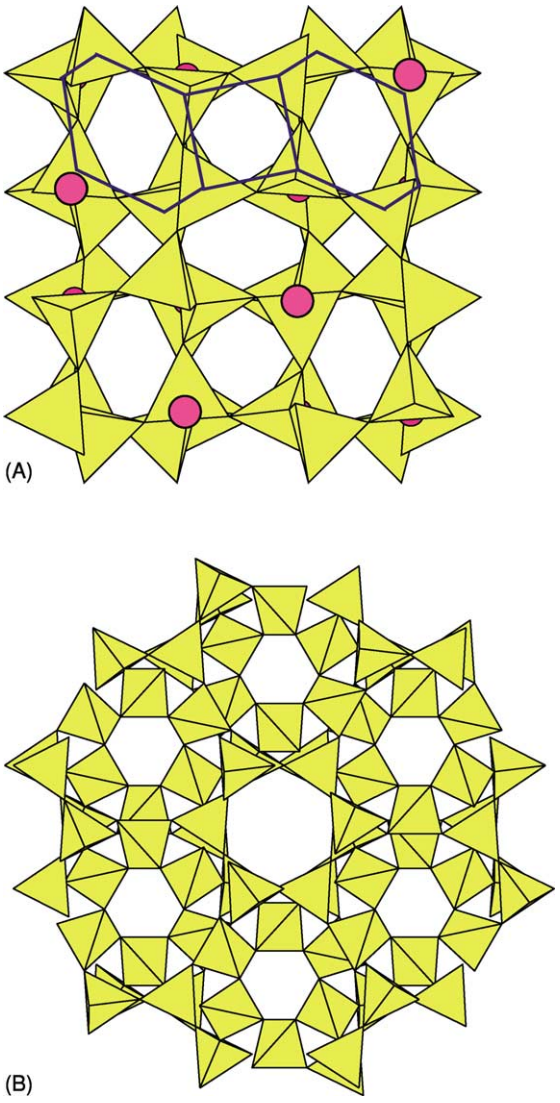


Figure 4 (A) The leucite structure is composed of connected four-membered and six-membered rings of tetrahedra. Two six-membered rings and one four-membered ring are shown in blue. Some potassium cations in the channels are shown in pink. (B) A view of the high-temperature cubic leucite structure looking down the <111> channels. Channel cations are omitted for clarity. In leucite these channels do not intersect, but in sodalite they do.

which the six-membered and four-membered rings are connected to each other differ in the leucite and sodalite structures. In leucite, the hexagonal rings are aligned to form a set of non-intersecting channels in four orientations corresponding to the triad axes of the cubic high-temperature phase. In sodalite, these channels intersect to form very large cavities. Figure 4 shows the structure of leucite, in which the four-membered and six-membered rings are connected and potassium ions occupy the channels. Natural leucites are usually extensively twinned as a result of a reduction in symmetry upon cooling from the high-temperature cubic polymorph to a tetragonal structure. In end-member leucite, $KAlSi_2O_6$, this transition occurs at about 625°C.

The leucite group has the general chemical formula $AB_3O_6 \cdot xH_2O$, where A is K, Na, or Cs, B is Al or Si, and x is between zero and one. The three main minerals of the group are leucite ($KAlSi_2O_6$), analcite ($NaAlSi_2O_6 \cdot H_2O$), and pollucite ($CsAlSi_2O_6 \cdot xH_2O$, where $x = 0$ to 1). Water molecules occur in the larger channels.

The structure of the sodalite group differs from the leucite structure primarily in that the channels intersect to form large cavities, which are occupied by chloride, sulphide, and sulphate anions. The hexagonal channels are occupied by sodium or calcium ions. The sodalite group has three main mineral types: sodalite (*sensu stricto*) ($Na_8Al_6Si_6O_{24}Cl_2$), nosean ($Na_8Al_6Si_6O_{24}SO_4$), and häuyne ($(Na,Ca)_{4-8}Al_6Si_6O_{24}(SO_4,S)_{1-2}$).

Natural Occurrences of Feldspathoids

Nepheline-group feldspathoids occur predominantly in silica-undersaturated igneous rocks such as syenites and phonolites, of which they are major diagnostic minerals. More rarely, nepheline can occur as a metasomatic alteration product in metamorphosed sedimentary rocks. Nepheline is used in the manufacture of ceramics and glasses.

Leucite-group and sodalite-group minerals occur only in alkaline volcanic rocks that are more siliceous than syenites and phonolites. Leucite itself is unstable and, upon slow cooling, can transform to pseudomorphs of potassium feldspar and nepheline.

See Also

Igneous Processes. **Minerals:** Definition and Classification; Feldspars; Quartz; Zeolites.

Further Reading

Deer WA, Howie RA, and Zussman J (2004) *Rock Forming Minerals*, vol. 4B, *Silica Minerals, Feldspathoids, and Zeolites*. London: Geological Society.

Palmer DC (1994) Stuffed derivatives of the silica polymorphs. *Reviews in Mineralogy* 29: 83–118.

Glauconites

J M Huggett, Petroclays, Ashtead, UK and The Natural History Museum, London, UK

Green clay can be Fe smectite, glauconitic-smectite, smectitic-glauconite, berthierine, odinite (formerly phyllite v), a ferric chlorite (phyllite c), chamosite, ferric illite, or celadonite. Although the environment of deposition can be a guide to identification, careful chemical and mineralogical analysis is required to unequivocally identify green clay minerals. The green granule-forming clays are glauconitic, odinite, berthierine, or rarely chamosite, all forming where the rate of deposition is slow. Glauconite is a green, ferric iron-rich micaceous mineral that forms in marine sediments. Odinite and ferric chlorite are the ferric iron-rich precursors of the ferrous iron-rich clays berthierine and chamosite. Together odinite and ferric chlorite comprise the verdine facies in recent sediments. Odinite is only known from marine sediments, but berthierine and chamosite can be non-marine. An assemblage of mainly glauconitic granules is referred to as the glaucony facies, whereas an assemblage of mainly iron-rich 7-Å granules is called the verdine facies (although the latter term is used for recent assemblages rather than ancient ones).

Composition

Glauconite (*sensu strictu*) is an Fe-rich, K-rich dioctahedral mica with tetrahedral Al (or Fe^{3+}) usually >0.2 atom per formula unit and octahedral R^{3+} correspondingly >1.2 atoms. Typically 5–12% of the total iron is ferrous (Table 1). The structural formula may be written as

$$K_{(x+y)}(Si_{4-x}, Al_x)_{\sim 4}(Fe^{3+}, Al, Mg, Fe^{2+})_{\sim 2} O_{10}(OH)_2$$

where x is 0.2–0.6 and y (the sum of the divalent octahedral cations) is 0.4–0.6.

It is distinguished chemically from ferric illite in having more total iron, and from celadonite in having greater substitution of aluminium for silicon in the tetrahedral layer and a higher octahedral charge (Table 1). Glauconitic granules that have replaced faecal pellets are almost always inhomogeneous, with inclusions of common detrital minerals: clays, quartz, feldspars, carbonates, or apatite. Glauconitic-smectite is a mixed layer clay that is less potassic than glauconite and which characterizes immature glauconitic granules. Fe smectite is a ferrous iron-bearing dioctahedral smectite, half way between montmorillonite and nontronite (iron-rich smectite). Odinite is a poorly ordered, ferric iron-rich, 7-Å 1:1 green clay with ~0.8 $Fe^{3+}/Fe^{2+} + Fe^3$ and low tetrahedra substitution of Al_2O_3 for SiO_2. It is believed to be the precursor of the poorly – ordered, ferrous iron-rich, 7-Å green clay berthierine, which has ≤0.2 $Fe^{3+}/Fe^{2+} + Fe^3$ and less Mg than odinite. Phyllite C has

Table 1 Chemical composition of various glauconites

	1	*2*	*3*	*4*	*5*	*6*
SiO_2	48.17	50.87	49.73	50.18	35.41	27.20
Al_2O_3	4.25	3.71	9.78	6.74	15.73	20.33
TiO2	0.07	0.07	0.05	0.00	0.00	0.00
Fe_2O_3	23.44	23.00	21.05	23.43	26.99	?
FeO[a]	2.34	2.30	2.10	2.34	1.30	37.96
MnO	0.04	0.04	0.03	0.01	0.14	0.00
MgO	3.59	3.93	4.34	2.77	4.95	2.97
CaO	0.09	0.29	0.22	1.10	0.00	0.00
Na_2O	0.41	0.52	0.15	0.56	0.00	0.00
K_2O	8.55	7.19	6.56	6.89	0.00	0.00
Total	90.95	91.92	94.01	94.02	84.52	88.46
Si	7.49	7.70	7.24	7.43	3.79	2.72
Al	0.51	0.30	0.76	0.57	0.21	1.28
Total tetrahedral	8.00	8.00	8.00	8.00	4.00	4.00
Al	0.27	0.36	0.92	0.60	1.76	1.12
Ti	0.01	0.01	0.01	0.00	0.00	0.00
Fe^{3+}	2.51	2.39	2.09	2.39	2.40	3.01
Fe^{2+} [a]	0.30	0.29	0.26	0.28	0.12	0.00
Mn	0.01	0.01	0.00	0.00	0.00	0.00
Mg	0.83	0.89	0.94	0.61	0.79	0.44
Total octahedral	3.93	3.95	4.21	3.89	5.07	4.57
Ca	0.02	0.05	0.03	0.17	—	—
Na	0.12	0.15	0.04	0.16	—	—
K	1.70	1.39	1.22	1.30	—	—
Total interlayer	1.84	1.59	1.29	1.63	—	—

1 = Shallow water glauconite (Pliocene) from ODP leg 174 A,
2 = deep water glauconite (Oligocene) from ODP leg 174 A,
3 = smectitic glacuonite from the Weches Formation (Eocene, Texas),
4 = smectitic glauconite from the Karai Formation (Cretaceous, India).
5 = odinite from Weches Formation (Eocene, Texas),
6 = berthierine. Interlayercations in berthierine are due to contamination by micaceous minerals. Number of cations are on the basis of O20(OH)4 for glauconitic clays and O10(OH)8 for odinite and berthierine.
[a]FeO is assumed to be 10% of the total iron for glauconite, the odinite FeO is <5% of total Fe (analysed by EELS).

chemistry similar to that of odinite but a 14-Å 2:1 lattice structure similar to that of chlorite; it is only found in recent verdine granules. Chamosite is chemically similar to berthierine, but has a 14-Å 2:1 clay structure; some chamosite forms by the replacement of berthierine.

Formation

Granules formed from faecal pellets are often referred to as pellets; where the origin is not clear, the term peloid may be used. Faecal pellets are the main substrate for glaucony formation in both ancient and recent sediments. For iron-rich, 7-Å clay they are the main substrate only in recent sediment; most ancient berthierine (and some chamosite) occurs as ooids or pore-lining or pore-filling clay. Other modes of occurrence are replacement of mica, quartz, chert, feldspar, calcite, dolomite, phosphate, and volcanic rock, cementation of microfossil cavities, and formation of thin films on limestone hardgrounds, quartz grains, and flint pebbles. The current consensus on the origin of iron-rich clays is that they form by direct precipitation, close to the sediment–water interface. The crucial component of the faecal pellets and microfossil cavities that predisposed them to replacement by iron-rich clay is the presence of labile organic matter, the composition of the clay-rich clastic component apparently having little if any affect upon the mineralization process. It is likely that microbial oxidation of the organic matter may be critical in creating redox conditions favourable to iron fixation and detrital clay dissolution. The formation of glaucony and iron-rich, 7-Å clays (verdine) is favoured by a semiconfined oxic microenvironment. This is indicated by the following observations:

1. Green clays form in microfossil tests.
2. Grains $<100\,\mu$m are commonly only slightly glauconitized, whereas grains of 100–500 μm are typically homogeneously glauconitized. Grains >1 mm have a zone of most intense glauconitization just inside the margin.
3. There is an absence of diffuse glaucony or iron-rich, 7-Å clay layers from sediment.
4. There is a concentration of green granules in invertebrate burrows.

Glauconitization

The iron in glauconite and its precursors is mainly ferric (unusual for a sedimentary silicate), but mobile iron is ferrous; hence, it is believed to form close to the redox boundary. The first stage in glauconitization of faecal pellets is the dissolution of carbonate, detrital clay and formation of Fe(III) smectite, while the organic matter content decreases. The overall Fe, and to a lesser extent K_2O (2–4%), content of the faecal pellets increases while the Al content decreases. At this stage the granules have the particulate or vermiform fabric of the detrital faecal pellets are ochre to light green in colour and soft; i.e., reworking will disperse them. This is the 'nascent' stage in the classification of Odin (see Further Reading). The 'slightly evolved stage' is characterized by the almost complete disappearance of detrital minerals (nonclay, noncarbonate minerals are the most resistant to replacement). Fe and K uptake continues through this phase, which is characterized by 4–6% K_2O, 20–28% Fe oxides, and a mixed layer expandable/10-Å clay composition. At this stage the pellets are olive green, and in backscatter SEM images still have a particulate or vermiform appearance (**Figure 1**). The 'evolved stage' is the result of recrystallization where dense, indurated dark green pellets, frequently cracked at the margins, are formed. The original fabric is destroyed and X-ray diffraction spectra show a strong reflection at 10 Å (diagnostic of potassium-bearing micas, including glauconite). At this stage there are few expandable layers remaining, K_2O content is 6–8%, and the iron content is no higher than that for the slightly evolved stage. In-filling of fractures with glauconite represents the highly evolved stage. By this stage the granule is made of end-member, *sensu strictu* glauconite with $>8\%$ K_2O. Note that iron is fixed in the octahedral layer before potassium fixation is complete. It is also potassium content that increases as the layer charge increases (due to loss of swelling layers and

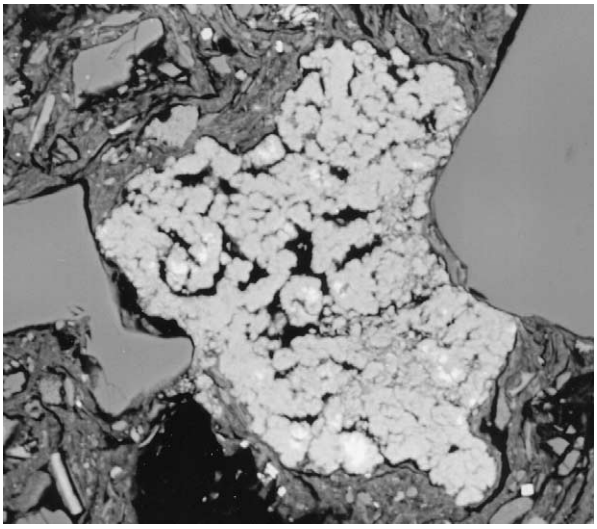


Figure 1 Thin section of vermiform (white in this picture) glauconite pellets in the ''slightly evolved'' stage of glauconization.

reduction of octahedral iron), and is consequently the only reliable chemical indicator of glauconitic maturity. If glauconitic clay is not buried, it will remain stable so long as there is not a major sea-level fall. In a very shallow marine, fluvial, or palaeosol environment, glauconite will oxidize to kaolinite and goethite. Recently it has been found that tunnelling into glauconitic sand can result in a sudden (and harmful to life) drop in oxygen level, due to oxidation of the ferrous iron in glauconitic clay to ferric oxides or hydroxides. This instability only affects incompletely evolved glaucony; the use of sandstone containing highly evolved glauconite as building stone that has remained green over several centuries demonstrates this. In a reducing environment glauconite may be replaced by phosphate. Burial to more than a few decimetres halts the glauconitization process. Hence, burial prior to the glaucony attaining the fully mature state will preserve the immature state. Burial diagenesis results in increasing substitution of Al for Si, a decrease in swelling (smectite) layers, and an increase in crystallinity; hence, the chemistry of ancient glauconite differs slightly from that of Recent and Quaternary glauconite.

Glauconite forms coatings on particles and as laterally extensive (many squared kilometres) surfaces on hardground. Coatings of glauconite on particles are most frequent on flint pebbles in chalk and on bioclasts. Thin section study of these films shows that the clay mainly develops along fissures, biogenic borings, and any other void of suitable size to provide the semiconfined microenvironment necessary to glauconite formation. The thickness of the film depends upon the substrate and extent of alteration: on silicates it is $\leq$3 mm for silicates and $\leq$1 cm for carbonates. Film glaucony is commonly present in Jurassic limestone and Late Cretaceous chalk and provides useful stratigraphic markers.

The origin of the elements necessary for the formation of glauconite has been the subject of much discussion. Experimental work has indicated that a high silica concentration is necessary for glauconite and iron-rich smectite formation, as at low silica concentration iron-rich, 7-Å clays will form instead. This may link glaucony formation to the presence of biogenic silica.

The concentrations of Fe, Al, and K in normal seawater are also too low for direct precipitation of glaucony, although in river water the concentrations of Fe and Si are sufficient. Reducing conditions develop as soon as all dissolved oxygen in the pore fluid is used up in microbial oxidation of organic matter. Subsequently iron-reducing bacteria become the major consumers of organic matter until the sediment is buried sufficiently deep for sulphate-reducing bacteria to become active. In this narrow depth zone, just below the sediment–water interface in oxic sediments, the concentration of dissolved (ferrous) iron increases. Where there is a low sedimentation rate, intraformational reworking, or winnowing, much of the dissolved iron will diffuse towards the oxic zone and be precipitated as ferric iron hydroxides. Such precipitates can sorb silica, Fe^{2+}, Al, K, Mg, B, and other ions from solution. Experimental work has shown that these poorly crystalline materials can reprecipitate as iron-rich smectite. Glauconite cannot be directly precipitated at the low K concentrations in natural pore water, but evolves gradually through increased interlayer charge and K fixation.

Studies of recent glaucony have provided information on the rate of glauconitization. On the Congolese continental shelf glaucony from depths $<$110 m is younger than 20 000 years and has an average K_2O content of 5%, indicating that it is slightly evolved, whereas the age of recent glaucony off Vancouver Island is estimated to be $<$3000 years. Exceptionally, very small mica grains in nonevolved faecal pellets may undergo K and Fe enrichment in less than 10 years. In general it is thought that a slightly evolved granule typically takes the order of 10^4 years to form, and a mature pellet 10^5–10^6 years. However, laser-argon probe dating suggests that the period of genesis may be as long as $\sim$5 million years. It is therefore possible to use glaucony maturity as an indication as to the length of a hiatus. However, the extent and intensity of glauconitization appears to depend upon the size of the host material.

Verdinization

Verdine granules are mostly odinite, with minor ferric chlorite and detrital minerals; they occur only in recent sediments and have never been buried. Also referred to as the process of *verdissement*, the replacement of faecal pellets by iron-rich, 7- and 14-Å minerals follows a maturation process similar to that for glauconite, although without the interlayer K it is less easy to monitor the degree of chemical evolution. Verdine formation requires oxic pore waters because both odinite and the less frequent ferric chlorite are rich in ferric iron. The formation of verdine is less well known than that of glauconite. It follows a similar morphological transformation from nascent to evolved, though detrital minerals may be more persistent (Figure 2). It has been suggested that the first part of the process is odinite formation, followed by formation of ferric chlorite, and finally replacement by ferrous iron-rich clays. However, as very few intermediate stages between recent odinite and diagenetically modified chlorite and berthierine deposits have

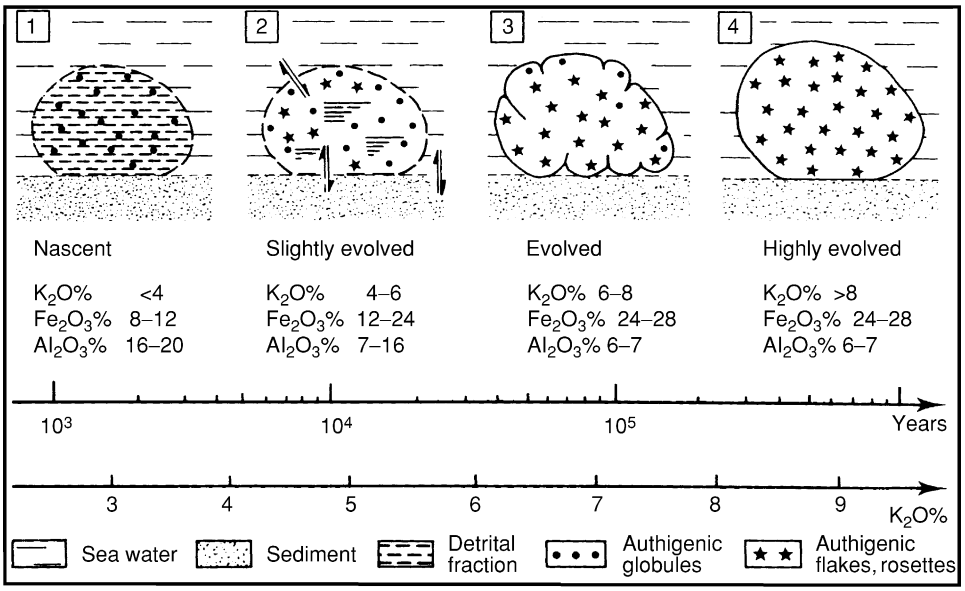


Figure 2 Morphological transformation from nascent to highly evolved glauconite.

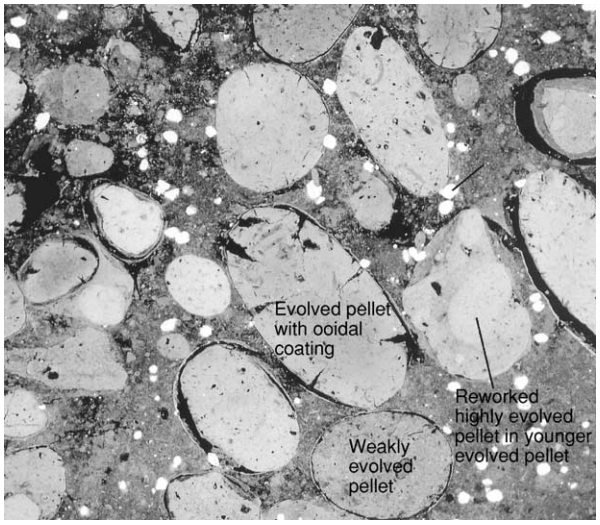


Figure 3 Thin section of faecal pellets with different stages of maturation towards full glauconization.

yet to be identified, the reaction pathway or pathways is not known with certainty. In some of the older literature it is suggested that berthierine forms by reaction of goethite and kaolinite. Although this may be true in some instances, it seems more likely that, like glaucony, formation of 7-Å, iron-rich clays is substrate independent. Verdinization may, like glaucony formation, follow a general pattern of iron-rich smectite formation followed by iron-rich, 7- and 14-Å clay formation (an example is the Claibourne Group of Texas, rich in pellets of an odinite-like mineral). Before it was realized that ferric iron-rich clays can be precursors of chlorite and berthierine it was commonly argued that these clays are indicative of a reducing environment at the time of sedimentation, rather than during diagenesis. It is possible to find many examples of this often-erroneous view in the literature.

Verdine is known only as granules and microfossil test in-fills. In contrast chamosite chlorite and berthierine frequently occur as ooidal coatings on grains, as pore-linings, replacement of detrital chlorite or discrete 'cryptocrystalline' clay, and only rarely as granules. This discrepancy has not been explained, although it might in part be due a widespread presumption that green granules are glauconitic. The odinite-like mineral found in the Claibourne Group of Texas does occur as ooidal coatings, but only rarely (Figure 3).

X-Ray Diffraction Identification of Green Clays

Optically the two various green granules are indistinguishable; hence, X-ray diffraction can assist in identifying the composition and purity of such granules. Due to their much higher Fe content, glauconite and ferric illite have a 002 (~5 Å) reflection much weaker than that of illite. Additionally, the iron-rich clays have a diagnostic 060 reflection at 1.51 Å. However, where both Fe-rich and-poor, 10-Å clays are present, this will not be so pronounced as that for the pure clays. All except the most evolved glauconitic clay will typically have some swelling clay layers, which can be identified by the broad area of reflection intensity between 10 and 14 Å and its shift to $\leq$17 Å

upon treatment with glycol and collapse to 10 Å upon heating at >400°C. Verdine granules yield a broad 001 reflection at ~7.2 Å, which is the strongest reflection of odinite, and also of berthierine, but the latter has a slightly smaller 002 reflection. Chlorites (both ferrous and ferric) are distinguished from other green clays by their 001 reflection at ~14 Å, which is unaffected by glycol treatment but shifts to a slightly smaller spacing upon heating at >500°C.

Distribution

In Time

Glaucony is absent from Precambrian sediments, but two major episodes of glauconite (and chamosite/berthierine) formation occurred during Phanerozoic time, the first in the Early Palaeozoic, and the second in the Middle-Late Mesozoic. Both episodes were characterized by temperate to subtropical climates, dispersed cratonic blocks, global sea-level rise (transgressive sediments), and periods of low shelf sedimentation. After the break-up of the Precambrian supercontinent about 600 My ago, glauconite became increasingly common through the Cambrian as cratonic blocks were rifted and dispersed from the Late Proterozoic supercontinent, the sea-level rose, and invertebrate faecal pellet makers became abundant. Glauconite formation in the Palaeozoic peaked in the Late Cambrian–Early Ordovician (*see* **Palaeozoic:** Ordovician). From then until the Cretaceous glauconite is relatively uncommon. The greatest period of glauconite formation occurred in the Cretaceous, notably along the continental margins of the widening North Atlantic Ocean, when global sea-level was again high (*see* **Mesozoic:** Cretaceous). Abundant glaucony continued until the Middle Cainozoic.

The earliest chamositic granules formed in the Early Proterozoic in association with greenalite. Chamosite ooids are not known until the Phanerozoic, and berthierine is only known from a few banded iron formations prior to the Mesozoic, whereas ancient odinite-like minerals are recorded only from the Eocene and younger rocks. The formation of ferrous clays in the Proterozoic was probably favoured by the lack of atmospheric oxygen. The major periods of chamosite/berthierine formation occurred when cratonic blocks were dispersed during the Ordovician, Devonian, and Jurassic–Cretaceous eras. Also global temperature may have effected the variation in abundance of chamosite and berthierine. In the Eocene a decrease in maturity of odinite-like mineral granules and an increase in Fe smectite may be linked to global cooling.

In Space

Recent glaucony apparently forms principally on the outer shelf at a water depth of 100–300 m, at a temperature of <20°C (possibly <13°C) and between 50° N and 50°S, although the true depth and latitude range may be masked by Late Pleistocene and Holocene sea-level fluctuations. Glaucony occurs where there are bottom currents that cause constant remobilization of sediment and winnowing. This association occurs because the intraformational reworking results in grains being exposed at the seafloor for sufficient time to permit glauconitization. Most recent glaucony occurs around the margins of the North and South Atlantic, and the Pacific west coast of the Americas, with smaller occurrences around the Mediterranean Sea and the Gulf of Oman. Glaucony is particularly abundant (up to 90% of the sediment) at the shelf-slope transition; however, apparently *in situ*, unweathered glauconite has been reported from ancient marine sediments deposited beneath 10–30 m water (e.g., in marginal marine to estuarine Tertiary sediments of the Hampshire basin, UK) and from clinoform toe sediments that must have been deposited beneath at least 600 m water. Very deep water glaucony (2000–3000 m) occurs in some recent sediments, as on the Chatham Rise, on the Scotia Ridge, and on both sides of Japan. The glauconite in the Chatham Rise and the Scotia Ridge sediments has not been investigated in detail, whereas that around Japan is associated with a fauna indicative of water shallower than the present depth. This could be due to either redeposition or rapid subsidence in this tectonically active area. Examples of ancient deep-water glauconitic sediments are the Miocene turbidites south-west of the Rockall Plateau (North Atlantic) and in Cretaceous turbidites in south-east France, e.g., the 'Petit Verolles'. The Petit Verolles is concentrated in coarser laminae, which suggests it was redeposited by turbidity currents along with the rest of the sediment with which it occurs. Glaucony in turbidites typically comprises fragments of mature pellets, because immature pellets are relatively soft and so become disaggregated during transport. It is likely that the precipitation temperature (which is reflected in the latitudinal constraints) and the availability of iron and potassium are ultimately more important than water depth. The importance of temperature is illustrated by the decrease in neoformed glaucony throughout the period of shallow marine sedimentation in the Eocene of the Hampshire basin (UK). This decrease coincides with a period of global

temperature decrease, whereas all other factors for glauconitization remained favourable. Indeed, the presence of the appropriate bacteria (which may be temperature-controlled) for organic matter degradation and iron reduction, plus faecal pellets (which are most abundant on the continental shelf in water <100 m deep), may prove to be the ultimate controls on glauconite formation.

Verdine in recent sediments is much more restricted than is glaucony. Its formation requires water temperatures >20°C; hence it is only known from the tropics. Verdine preferentially forms where the water depth is 20–60 m (but <5 m in the Casamance Estuary, Senegal, and up to 200 m elsewhere), where large amounts of land-derived iron compounds have been deposited in shallow marine sediments. Major recent examples are seaward of the Congo River, the Niger Delta, and between the mouths of the Amazon and Orinoco rivers. Recent verdine is also associated with iron derived from volcanic rocks, as in the case of offshore from New Caledonia. Berthierine, which in some instances has probably replaced odinite as the principal component of verdine, apparently precipitates over a much wider temperature range. For example, it is known from Arctic palaeosols and estuarine sediments of western Scotland. However, like glaucony, berthierine can be an indicator of hiatus, and in the case of the latter mineral this applies to nonmarine as well as to marine sediments. Hence berthierine and chamosite may be of use for onshore sequence stratigraphic interpretation.

The different types of substrate for replacement by green clays in a marine environment are typically found at different water depths. For example, on the Western African margin there is a succession of substrates from planktonic foraminiferans on the outer shelf margin, to benthic bioclasts, faecal pellets, mineral grains, and shell bioclasts in water <5 m deep. Of these the deeper water substrates are glauconitized, and the shallower ones are replaced by verdine. Such a succession is only possible in a period of stable sea-level. During a transgression, successive substrates will be glauconitized. Note that this produces a glauconitic sediment that may be used for facies correlation, although the age of the sediment will decrease up-dip.

Because green clay granules are slow to form, they are favoured by a low sedimentation rate, which may partially explain the depth constraint. The slow rate of formation also explains the common association of abundant glaucony with transgressive sediments. The frequent association of glaucony with transgressions and condensed sections predisposes it to reworking. Indeed, any glaucony found outside the range suggested above should be checked carefully for evidence of reworking. There have been several reports of nonmarine glauconite but detailed investigation has shown that it was reworked from underlying marine deposits.

The principal criteria for evidence of *in situ* granules are:

1. Concentration of glaucony in burrows rather than the host sediment; and
2. The presence of immature granules, which are soft and unlikely to withstand transport over more than a few metres; and
3. Deeply fractured mature granules (e.g., Plate), which would not survive transport without breaking up.

However, for reworked granules one or more of the following criteria may apply:

1. Fragments of granules;
2. Sand-size green granules in silt- or clay-dominated sediment;
3. Cross-bedding (which implies a high-energy environment), especially with laminae of green granules, whereas green clays are associated with slow sedimentation; and
4. Inclusions of material not representative of the host sediment; at the immature stage glaucony or verdine has a mineralogy similar to that of the enclosing sediment.

Age Dating

Potassium-rich glaucony provides ~40% of the absolute age dates for the geological time-scale of the past 250 million years. However, there are potential problems with this method of K/Ar radiogenic isotopic age dating:

1. Argon loss.
2. Residual detrital illite or K feldspar will result in incorrect (too old) dates. Both K/Ar and Rb–Sr apparent ages progressively decrease as glauconitization proceeds, but even the most evolved grains may retain around 10% of the initial substrate, which may affect the apparent age, depending upon its mineralogy. This is true for both ancient and recent glaucony.
3. Potassium-rich glaucony, being the most resistant to weathering, is the most likely to survive reworking; hence, ages greater than time of sedimentation may be measured.

4. A single glaucony assemblage may represent a cluster of glauconitization events; hence, the bulk age will be an average of these.

To get a meaningful, if not completely accurate date, the glauconite should be *in situ*, then sieved, mixed with distilled water, and subjected to one hour in an ultrasonic bath to break up the less-evolved granules. The remaining intact granules can then be separated off by washing, and if necessary any feldspar can be separated by magnetic separation. *In situ* glaucony with $>7\%$ K_2O gives the most reliable dates. Argon loss can be avoided by the microencapsulation technique. And clustered glauconitization events can be individually identified by single-grain age dating.

Ferric Illite

Ferric illite (sometimes misleadingly called nonmarine glauconite) most often forms green illite-rich claystones rather than granules. The iron content is much less, and the aluminium content is higher than that in marine glauconitic minerals. It forms in palaeosols and hypersaline lacustrine environments, but is not known to originate from marine environments. In lacustrine environments it is often associated with limestone and evaporite minerals, where alkaline, potassium-rich pore water favours illite precipitation. In palaeosols illite forms by the wetting and drying of smectite. Ferric iron in the smectite is reduced while the clay is wet, resulting in increased layer charge, loss of swelling capacity, and K fixation in the interlayer sites. Note that green clay in palaeosols can also be berthierine, but the latter is not associated with wetting and drying.

See Also

Analytical Methods: Geochronological Techniques. **Clay Minerals**. **Mesozoic:** Cretaceous. **Palaeozoic:** Ordovician. **Sedimentary Environments:** Depositional Systems and Facies. **Sedimentary Rocks:** Ironstones; Sandstones, Diagenesis and Porosity Evolution.

Further Reading

Amorosi A (1995) Glaucony and sequence stratigraphy: A conceptual framework of distribution in siliciclastic sequences. *Journal of Sedimentary Research B* 65: 419–425.

Bailey SW (1988) Odinite, a new dioctahedral-trioctahedral Fe^{3+}-rich 1:1 clay mineral. *Clay Minerals* 23: 237–248.

Brindley GW (1982) Chemical compositions of berthierines—a review. *Clays and Clay Minerals* 30: 153–155.

Chamley H (1989) *Clay Sedimentology.* Berlin: Springer-Verlag.

Clauer N, Keppens E, and Stille P (1992) Sr isotopic constraints on the process of glauconitization. *Geology* 20: 133–136.

Hesselbo PS and Huggett JM (2001) Glaucony in ocean-margin sequence stratigraphy (mid-Cenozoic, offshore New Jersey, USA, ODP, Leg 174A). *Journal of Sedimentary Petrology* 74: 599–607.

Huggett JM and Gale AS (1997) Petrology and palaeoenvironmental significance of glaucony in the Eocene succession at Whitecliff Bay, Hampshire Basin, UK. *Journal of the Geological Society, London* 154: 897–912.

Ireland BJ, Curtis CD, and Whiteman JA (1983) Compositional variation within some glauconites and illites and implications for their stability and origins. *Sedimentology* 30: 769–786.

Odin GS (1988) *Green Marine Clays,* Developments in Sedimentology, 45. Elsevier: Amsterdam.

Rao VP, Thamban M, and Lamboy M (1995) Verdine and glaucony facies from surficial sediments of the eastern continental margin of India. *Marine Geology* 127: 105–113.

Stille P and Clauer N (1994) The process of glauconitization: chemical and isotopic evidence. *Contributions to Mineralogy and Petrology* 117: 253–262.

Toth TA and Fritz SJ (1997) An Fe-berthierine from a Cretaceous Laterite: Part 1. Characterization. *Clays and Clay Minerals* 45: 564–579.

Van Houten FB and Purucker ME (1984) Glauconitic peloids, and chamositic ooids—Favorable factors, constraints, and problems. *Earth Science Reviews* 20: 211–243.

Micas

R A Howie, Royal Holloway, University of London, London, UK

Introduction

The minerals of the mica group show considerable variation in their chemical and physical properties, but all are characterized by a platy morphology and perfect basal cleavage which is a consequence of their layered atomic structure. They all have a negative optic sign and have the optic direction α approximately perpendicular to their perfect cleavage. The most common micas are muscovite (or white mica), the phlogopite-biotite series (brown or black mica), and lepidolite (lithium mica), of which muscovite, phlogopite, and lepidolite are of economic importance. The classification of

micas has been reviewed recently by an international committee.

Crystal structure

The basic structural feature of mica is a composite sheet in which a layer of octahedrally coordinated *Y* cations is sandwiched between two identical layers of linked $(Si,Al)O_4$ tetrahedra. Two of these tetrahedral sheets, with a composition $(Si,Al)_2O_5$, are illustrated in Figure 1. On the left is a sheet in which all tetrahedra are pointing upwards (see the elevation view), and on the right is a sheet of tetrahedra which all point downwards. These two sheets are superimposed and are linked by a plane of octahedrally coordinated cations (Figure 2). Additional hydroxyl ions, together with the apical oxygens of the inward-pointing tetrahedra, complete the octahedral coordination of the sandwiched cations. The central *Y* ions determine the positions of the two tetrahedral sheets so that they are displaced relative to one another by *a*/3 in the [001] direction. Thus the micas, although superficially hexagonal, have a monoclinic unit cell (outlined in Figure 2). The perfect cleavage is along the weakly bound sheet of *X* ions (K, Na, etc.) lying between the tetrahedral layers.

The tetrahedral hexagons may be superimposed in six different ways. Thus one hexagon may be related to the next by rotation through 0° or by multiples of 60°; this, in combination with the stagger of *a*/3 introduced by the *Y* layer (Figure 2), determines the location of corresponding atoms in successive cells. Various sequences of layer rotations are possible, and when repeated regularly these build up unit cells with one, two, three, or more layers. The commonest stacking sequences lead to either one- or two-layered monoclinic polytypes (symbols 1M, $2M_1$), a different two-layered monoclinic ($2M_2$) or a three-layered trigonal (3T) polytype.

Chemistry

The general formula, which may be used to describe the chemical composition of micas, is $X_2Y_{4-6}Z_8O_{20}(OH,F)_4$, where:

X is mainly K, Na or Ca, but also Ba, Rb, Cs, etc.
Y is mainly Al, Mg or Fe, but also Mn, Cr, Ti, Li, etc.
Z is mainly Si or Al, but perhaps also Fe^{3+} and Ti

The micas can be subdivided into di-octahedral and tri-octahedral classes in which the number of *Y* ions is 4 and 6, respectively. Thus muscovite $K_2Al_4[Si_6Al_2O_{20}(OH,F)_4$ is di-octahedral whereas phlogopite $K_2(Mg,Fe^{2+})_6[Si_6Al_2O_{20}](OH,F)_4$ is tri-octahedral.

Di-octahedral			
	X	*Y*	*Z*
Muscovite	K_2	Al_4	Si_6Al_2
Paragonite	Na_2	Al_4	Si_6Al_2
Glauconite	$(K,Na)_{1.2–2.0}$	$(Fe,Mg,Al)_4$	$Si_{7–7.6}Al_{1.0–0.4}$
Tri-octahedral			
	X	*Y*	*Z*
Phlogopite	K_2	$(Mg,Fe^{2+})_6$	Si_6Al_2
Biotite	K_2	$(Mg,Fe^{2+},Al)_6$	$Si_{6–8}Al_{2–3}$
Zinwaldite	K_2	$(Fe, Li,Al)_6$	$Si_{6–7}Al_{2–1}$
Lepidolite	K_2	$(Li,Al)_6$	$Si_{6–5}Al_{2–3}$

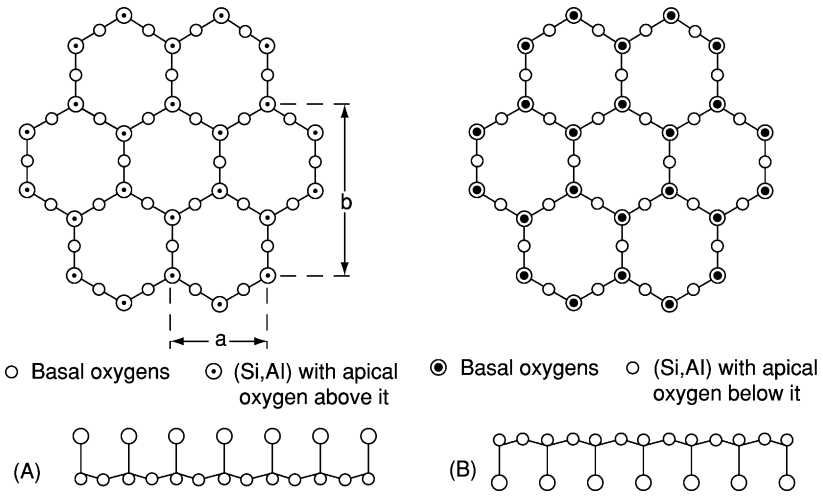


Figure 1 Tetrahedral layer $[(Si,Al)_4O_{10}]$ in an ideal mica structure viewed in *z*- and *y*-axis projection with tetrahedra pointing (A) upwards and (B) downwards. (After Deer *et al.* (1992).)

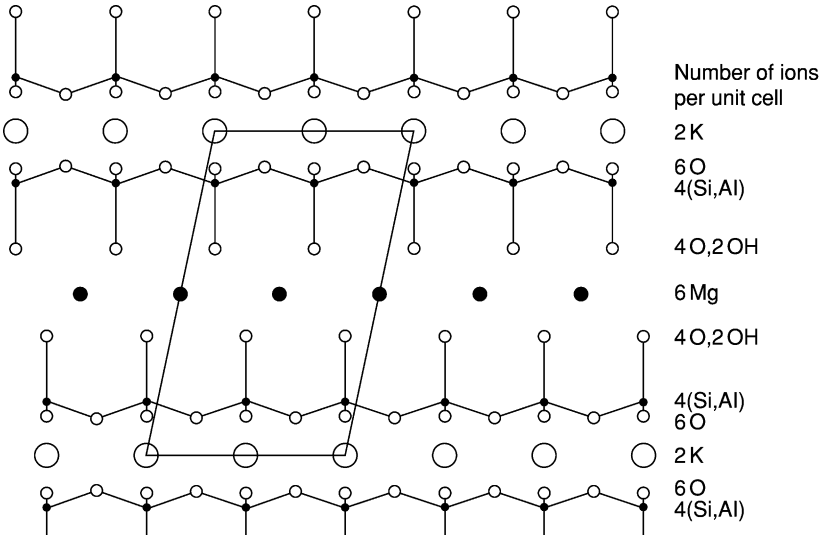


Figure 2 The crystal structure of phlogopite [$KMg_3AlSi_3O_{10}(OH)_2$] viewed in *y*-axis projection, highlighting the three-layer (tetrahedral–octahedral–octahedral [*t–o–t*] sandwich unit, and the layer of interlayer cations (K^+). (After Deer *et al.* (1992).)

Some 40 mica species are recognized, but the more common micas are be listed as below. For muscovite and paragonite, a Si/Al ratio greater than 6:2 can be balanced by equivalent substitution of divalent ions for Al in the *Y* sites.

A chemical feature in most micas is their water content, which is generally around 4–5 wt% H_2O^+, except for those with a high fluorine content. The F/(F + OH) ratio rises in acid rocks, particularly in late-stage granites (*see* **Igneous Rocks:** Granite) and in the micas of pegmatites.

Optical and physical properties

The optical properties of micas cover a wide range but all have negative optic sign, low 2V, and have the α optic direction approximately perpendicular to their perfect cleavage. Birefringence is generally very weak in the plane of cleavage flakes but strong in transverse sections. Pleochroism is strong in coloured micas such as biotite, where it may range from yellow in the α direction to dark reddish-brown perpendicular to this direction. The perfect {001} cleavage of all micas is a useful characteristic. The sheet-like nature of the micas makes muscovite and phlogopite of economic interest in electrical insulation applications

Paragenesis

Minerals of the mica group occur in igneous, metamorphic, and sedimentary rocks. The principal occurrences are as follows:

i. *Igneous rocks*
   - Muscovite: granites, granitic pegmatites, and aplites.
   - Phlogopite: peridotites and kimberlites.
   - Biotite: gabbros, norites, diorites, syenites, granites, and pegmatites.
   - Lepidolite and zinnwaldite: pegmatites and high-temperature veins.

ii. *Metamorphic rocks*
   - Muscovite, paragonite, and biotite: phyllites, schists, and gneisses.
   - Phlogopite: metamorphosed limestones and dolomites.

iii. *Sedimentary rocks*
   - Muscovite and paragonite: detrital and authigenic sediments.
   - Glauconite: greensands.

See Also

Clay Minerals. **Igneous Rocks:** Granite. **Minerals:** Definition and Classification; Amphiboles; Other Silicates; Pyroxenes.

Further Reading

Deer WA, Howie RA, and Zussman J (1992) *An Introduction to the Rock-Forming Minerals*, 2nd Ed. London: Longman.

Fleet M (2003) *The Micas. Rock-Forming Minerals*, 2nd edn. Vol. 3A. London: Geological Society.

Rieder M, Cavazzini G, D'Yakonov YS, *et al.* (1999) Nomenclature of the micas. *Mineralogical Magazine* 63: 267–279. *Clays and Clay.*

Smith JV and Yoder HS, Jr (1956) Experimental and theoretical studies of the mica polymorphs. *Mineralogical Magazine* 31: 209–234.

Molybdates

P A Williams, University of Western Sydney, Penrith South DC, NSW, Australia

Inroduction

Molybdates are compounds containing negatively charged oxymolybdenum ions. For the most part, mineralogically, the simple tetraoxomolybdate(VI), or molybdate ion, MoO_4^{2-}, is present. Under acid conditions, molybdate ions polymerize and this process can incorporate other chemical entities. Resulting heteropolymolybdates are represented in the mineral kingdom, but are rare. Indeed, only a few minerals containing essential molybdate are known, and overwhelmingly, molybdate minerals contain the simple molybdate ion. In this connection, the chemistry of molybdenum displays some differences compared to that of its congeners, chromium and tungsten. The former does not polymerize to any significant extent in its highest oxidation state, aside from the formation of dichromate(VI), $Cr_2O_7^{2-}$, and tungsten and molybdenum minerals in the primary zones of deposits are markedly differentiated by virtue of redox differences. Standard potentials at 25°C are –0.78 and –1.26 V for eqns [1] and [2], respectively. This accounts for the fact that tungstates are favoured even in the primary environment; tungstenite, WS_2, is an extremely rare species.

$$MoO_4^{2-} + 2H_2O + 2e^- \rightarrow MoO_2 + 4OH \qquad [1]$$

$$WO_4^{2-} + 2H_2O + 2e^- \rightarrow WO_2 + 4OH \qquad [2]$$

Primary Molybdates

The primary mineralogy of molybdenum is dominated by molybdenite, MoS_2. Aside from the substitution of molybdate in primary scheelite, $CaWO_4$, molybdates are generally confined to the supergene environment. A remarkable exception to this pattern of paragenesis concerns the rare occurrence of primary powellite, $CaMoO_4$, in centimetre-sized crystals from basalt-hosted deposits of the Deccan Trap in India. In primary or other reducing environments, lower molybdenum valencies are stabilized in various complex oxides. Some molybdenum oxides are termed 'fumarolic condensates', and the more complex mineral, kamiokite, $Fe_2Mo_3O_8$, found in such settings contains Mo(IV). Sedovite is thought to be a U(IV) molybdate, and mourite, $UMo_5O_{12}(OH)_{10}$, may also contain U(IV). Alternatively, it may be a mixed-valency Mo(V,VI) molybdate or polymolybdate of the uranyl ion, but further studies are required to characterize this species fully.

Secondary Molybdates

Apart from the preceding examples, molybdates are mostly confined to the oxidized zones of base metal orebodies. Here molybdate is derived from the oxidation of molybdenite. In such environments, a number of simple salts are found, together with rarer double salts and polymolybdates, as listed in Table 1. For simple salts such as powellite, substitution by tungstate is common and solid solution over the full range of compositions to scheelite is known. The end-member composition is shown in Table 1, as is the case with other examples (aspects of solid solution are dealt with later, where they are significant). Where solid solution is more complex, stoichiometries reflect compositions of known materials.

In similar fashion to the case with the powellite–scheelite pair, complete solid solution between wulfenite, $PbMoO_4$, and stolzite, $PbWO_4$, can occur, although most naturally occurring material contains minor tungstate. Other ions substitute in the lattice as well, including chromate, arsenate, and vanadate, the latter involving some other charge compensation mechanism. Molybdofornacite, $CuPb_2[(Mo,Cr)O_4][(As,P)O_4]OH$, is structurally related to the brackebuschite group of monoclinic arsenates, phosphates, sulphates, and vanadates of general formula $AB_2(XO_4)_2(OH,H_2O)$; A = Al, Cu, Fe, Mn, Zn; B = Ba, Ca, Pb, Sr; X = As, P, S, V. Chromate substitution is pronounced in this species. However, not all of the species in Table 1, even those of comparatively simple stoichiometry, contain isolated molybdate ions. Koechlinite is an example of a layered-structure compound; this type of chemical and structural feature is more pronounced in related minerals of W(VI). For example, there is no Mo(VI) analogue of the dimorph of stolzite, raspite, that contains chains of edge-linked WO_6 octahedra (*see* **Minerals:** Tungstates). The heteropolymolybdates betpakdalite, sodium betpakdalite, and mendozavilite are all rare minerals, although betpakdalite is commonly found in small amounts in the oxidized zones of molybdenite deposits. It is possible that the very rare minerals

Table 1 Molybdate(VI) minerals

Mineral	*Chemical composition*
Simple molybdates	
Powellite	$CaMoO_4$
Wulfenite	$PbMoO_4$
Umohoite	$UO_2MoO_4 \cdot 4H_2O$
Sedovite	$U(MoO_4)_2$
Ferrimolybdite	$Fe_2(MoO_4)_3 \cdot 8H_2O$
Basic double salts	
Koechlinite	Bi_2MoO_6
Lindgrenite	$Cu_3(MoO_4)_2(OH)_2$
Szenicsite	$Cu_3MoO_4(OH)_4$
Parkinsonite	$Pb_6O_4(MoO_4)Cl_2$
Molybdofornacite	$CuPb_2[(Mo,Cr)O_4][(As,P)O_4]OH$
Complex uranium salts	
Deloryite	$Cu_4UO_2(MoO_4)_2(OH)_6$
Calcurmolite	$Ca(UO_2)_3(MoO_4)_3(OH)_2 \cdot 11H_2O$
Tengchongite	$Ca(UO_2)_6(MoO_4)_2(OH)_{10} \cdot 7H_2O$
Moluranite	$H_4U(UO_2)_3(MoO_4)_7 \cdot 18H_2O(?)$
Cousinite	$Mg(UO_2)_2(MoO_4)_2(OH)_2 \cdot 5H_2O(?)$
Mourite	$UMo_5O_{12}(OH)_{10}(?)$
Polymolybdates	
Iriginite	$UO_2(Mo_2O_7) \cdot 3H_2O$
Betpakdalite	$MgCa_2[Mo_8As_2Fe_3O_{36}(OH)](H_2O)_{23}$
Sodium betpakdalite	$MgCa_2[Mo_8As_2Fe_3O_{36}(OH)](H_2O)_{23}$
Mendozavilite	$Na(Ca,Mg)_2Fe_6(PO_4)_2(PMo_{11}O_{39})(OH,Cl)_{10} \cdot 33H_2O$
Other complex species	
Obradovicite	$H_4(K,Na)CuFe_2AsO_4(MoO_4)_5 \cdot 12H_2O$
Melkovite	$CaFeH_6PO_4(MoO_4)_4 \cdot 6H_2O$
Chiluite	$Bi_6(TeO_4)_2(MoO_4)_2O_5$

obradovicite, melkovite, and chiluite also belong to this class of heteropolymolybdate.

Molybdate minerals have little commercial significance, although wulfenite assumed an important role in this regard between the two World Wars. The Mammoth–St Anthony mine at Tiger, Arizona, USA, was the principal producer of molybdenum during this period, all from oxidized ores carrying wulfenite. Today, however, the mineral is of interest only to collectors. Nevertheless, it is a common mineral in many deposits, the most noteworthy of which are found in the south-western states of the United States and in Mexico.

See Also

Minerals: Definition and Classification; Chromates; Tungstates.

Further Reading

Anthony JA, Bideaux RA, Bladh KW, and Nichols MC (2003) *Handbook of Mineralogy. Volume 5. Borates, Carbonates, Sulfates, Chromates, Germanates, Iodates, Molybdates, Tungstates, etc., and Organic Materials.* Tucson, AZ: Mineral Data Publishing.

Anthony JA, Williams SA, Bideaux RA, and Grant RW (1995) *Mineralogy of Arizona,* 3rd edn. Tucson, AZ: University of Arizona Press.

Baes CF Jr and Mesmer RE (1986) *The Hydrolysis of Cations.* Malabar, FL: Krieger Publishing Company.

Bard AJ, Parsons R, and Jordan J (1985) *Standard Potentials in Aqueous Solution.* New York: Marcel Dekker.

Bideaux RA (1980) Famous mineral localities: Tiger, Arizona. In: *The Mineralogical Record,* vol. 11, pp. 155–180. Tucson, AZ: Mineralogical Record Inc.

Gaines RV, Skinner HCW, Foord EE, Mason B, and Rosenzweig A (eds.) (1997) *Dana's New Mineralogy: The System of Mineralogy of James Dwight Dana and Edward Salisbury Dana,* 8th edn. London: Wiley Europe.

Mandarino JA (1999) *Fleischer's Glossary of Mineral Species 1999,* 8th edn. Tucson, AZ: Mineralogical Record Inc.

Ottens B (2003) Minerals of the Deccan Traps, India. In: *The Mineralogical Record,* vol. 34, pp. 1–82. Tucson, AZ: Mineralogical Record Inc.

Roberts WL, Campbell TJ, and Rapp GR Jr (1990) *Encyclopedia of Minerals,* 2nd edn. New York: Van Nostrand Reinhold.

Williams PA (1990) *Oxide Zone Geochemistry.* Chichester: Ellis Horwood.

Native Elements

P A Williams, University of Western Sydney, Parramata, Australia

Introduction

A surprising number of the elements of the Periodic Table occur naturally, in the so-called native state. These range from the familiar coinage metals – copper, silver, and gold – to quite reactive metals, such as iron, zinc, and perhaps aluminium. Non-metallic elements such as sulphur and carbon are also well represented in the mineral kingdom (Table 1). The amounts of native elements present in rocks and ores vary widely, from minute grains to bodies weighing millions of tonnes. A number of these occurrences play extremely important economic roles, yet others are no more than mineralogical curiosities. The elements and their associations in widely disparate environments are reviewed, but ordered intermetallic species are not included.

Gases and Liquids

The majority of the atmosphere is made up of two gaseous elements, nitrogen and oxygen. Traces of hydrogen are known to be present. All of the noble gases – helium, neon, argon, krypton, xenon, and radon – are present in the atmosphere, the latter in small amounts, particularly in soil gas, where it is produced as the result of the radioactive decay of certain elements. No doubt other elements, including astatine, not considered in detail below, exist transiently in the free state as members of various radioactive decay series, but this aspect of the existence of the elements is rather arcane and is beyond the scope of this article. The third most abundant gas in the atmosphere is argon. Its apparently abnormal distribution is due to ^{40}Ar being a stable decay product of ^{40}K. While these gaseous elements are ubiquitous on Earth, albeit in highly variable amounts, they are not considered to be minerals. Another element present in volcanic gases is chlorine.

Under ambient conditions, only one element is found on Earth as a liquid. Native mercury is widespread, usually in small amounts, in low-temperature hydrothermal deposits and in hot springs. One exceptional occurrence is in the celebrated mercury deposits of Almadén, Spain, where it occurs in sufficient amounts to make it an economic mineral.

Solid Metals

The frequent occurrence of the coinage metals – copper, silver, and gold – gave rise to their use from early times. Native gold (*see* **Gold**) is overwhelmingly the commonest mineral form of the element and

Table 1 The naturally occurring native elements

Gases	
Argon	Ar
Astatine	At
Chlorine	Cl_2
Helium	He
Hydrogen	H_2
Iodine	I_2
Krypton	Kr
Neon	Ne
Nitrogen	N_2
Oxygen	O_2
Radon	Rn
Xenon	Xe
Liquids	
Mercury	Hg
Solids	
Allargentum	$Ag_{1-x}Sb_x$ ($x = 0.09$–0.16)
Aluminium	Al
Antimony	Sb
Arsenic	As
Cadmium	Cd
Copper	Cu
Chromium	Cr
Diamond	C
Gold	Au
Graphite	C
Indium	In
Iodine	I_2
Iridium	Ir
Iron	Fe
Kamacite	(Ni, Fe)
Lead	Pb
Osmium	Os
Palladium	Pd
Phosphorus	P_n
Platinum	Pt
Rhodium	Rh
Rosickýite	S_8; γ-sulphur
Ruthenium	Ru
Selenium	Se
Silver	Ag
Stibarsen	SbAs
Sulphur	S_8; α-sulphur
Taenite	(Ni, Fe)
Tellurium	Te
Tin	Sn
Unnamed	(Ni, Fe); unnamed tetragonal phase
Unnamed	S_8; β-sulphur
Unnamed	S_n; amorphous sulphur
Zinc	Zn

remains its economically most important ore mineral. It is frequently alloyed with silver (electrum) or mercury (an amalgam) and usually contains small amounts of other alloyed elements. While silver and gold form a continuous solid solution, the mutual solubilities of copper and silver are quite restricted. Gold of very high fineness is known from supergene deposits. Silver is a comparatively common native element, forming a range of solid solutions with other elements. It also is known to occur as ordered alloys and intermetallic compounds, which are not considered here. Allargentum is a hexagonal mineral that may be considered to be the hexagonal form of silver, at least in part. Native silver is present in both primary and secondary ores, but it is now chiefly produced by the refining of other metals or concentrates, especially those of copper, lead, and zinc. Native copper is a very common mineral in the oxidized zone of base metal ore bodies. It is also found in the form of masses amounting to many tonnes in weight in the famous Keweenaw Peninsula of northern Michigan, USA. Native copper remains a significant ore mineral, especially in the oxidized zones of base metal deposits.

The other precious metals – iridium, osmium, palladium, platinum, rhodium, and ruthenium – are well known minerals in primary and placer deposits. They also form alloys with each other and with base and noble metals. They remain economically important species in layered deposits of platinum-group elements such as those in the Bushveldt (South Africa), Stillwater (Montana, USA), and related igneous complexes. Platinum was originally extracted from extensive placer and alluvial deposits in South America, the Urals, and elsewhere, but these are not now important sources of the element.

Many other metallic elements are known to exist in the natural environment in amounts varying from isolated grains to large masses. Native iron is a frequent accessory mineral of basic rocks, and masses weighing many tonnes have been recovered from weathering basalts on Disko Island, Greenland. It is also found in meteoritic materials, alloyed with nickel and small amounts of cobalt and other elements. Native nickel is a minor component of nickel-rich ultramafic rocks. Taenite (Ni, Fe) and kamacite (Ni Fe) are ordered and disordered cubic minerals that have been recovered from Ni–Fe meteorites, and a third tetragonal phase is known. Elemental chromium has been reported from unspecified localities in Russia and China. A number of high- and low-temperature hydrothermal deposits are known to host native antimony, arsenic, and bismuth. The last is a significant ore mineral in some deposits. The intermetallic compound stibarsen (SbAs) is known to occur naturally, and both native selenium and tellurium, included here with the metals, are frequent accessories in certain epithermal deposits. Native indium has been reported to occur in several unspecified Russian deposits.

Very rarely, metals whose natural existence would be unexpected because of their redox potentials have been found in small amounts. Aluminium, zinc, and cadmium have been found in highly reduced skarn and intrusive igneous deposits. Native zinc is the most common of these and has been found in other settings. Elemental tin has been found in a number of igneous deposits. Grains of tin have also been recovered from placer deposits in northern New South Wales, Australia. It is thought that these may be the result of the reduction of cassiterite, SnO_2, by carbon during bush fires. Some doubt has attended many reports of native lead because a number of cases may be due to the recovery of shot or other human artefacts. However, it has been reliably reported from a dozen or so deposits, and fine crystal groups have been found at Långbarn and the Harstigen mine, Sweden.

Solid Non-Metals

Only four solid non-metals, other than perhaps selenium, occur naturally as elements. These are carbon, sulphur, phosphorus, and iodine. Two allotropes of carbon – diamond and graphite – are recognized as minerals. No doubt the buckminsterfullerines and related compounds occur in nature as well, but they have not yet been formally described. Deposits of millions of tonnes of α-sulphur (the mineral sulphur) occur in so-called 'salt dome' deposits, and these are the most important commercial source of the element, which is recovered by the Frasch process. Sulphur is found in other sedimentary rocks, in the oxidized zones of many sulphide ore bodies, as a product of various micro-organisms, and very frequently as a sublimate associated with volcanic fumaroles, where it is sometimes associated with selenium. Two further modifications of sulphur have been reported to occur as minerals. Rosickýite is the natural counterpart of the well-known allotrope γ-sulphur. It occurs as a volcanic sublimate at Vulcano in the Mediterranean Sea and in sediments in Czechoslovakia and the USA. Naturally occurring β-sulphur has been reported from volcanic fumaroles on Vulcano and Mount Vesuvius, Italy. Amorphous or 'plastic' sulphur occurs in the Kobui sulphur mine, Japan, but it crystallizes over time to α-sulphur. Native phosphorus has been found in the Saline Township meteorite from Kansas, USA. Finally, the occurrence of elemental iodine is worthy of note. Iodine-rich caliche from certain

deposits of the Atacama Desert, Chile, is coloured blue-purple by elemental iodine. In addition, specimens of marshite (CuI) frequently possess the characteristic odour of iodine. This is perhaps not surprising given the mode of formation of marshite by the interaction of aqueous solutions carrying Cu^{2+} and I^- ions:

$$Cu^{2+}(aq) + 2I^-(aq) \rightarrow CuI(s) + 0.5I_2(s)$$

See Also

Gold. **Mining Geology:** Hydrothermal Ores; Magmatic Ores. **Solar System:** Meteorites. **Tectonics:** Hydrothermal Activity.

Further Reading

Anthony JA, Bideaux RA, Bladh KW, and Nichols MC (1990) *Handbook of Mineralogy. Volume 1. Elements, Sulfides, Sulfosalts*. Tucson: Mineral Data Publishing.

Bard AJ, Parsons R, and Jordan J (1985) *Standard Potentials in Aqueous Solution*. New York: Marcel Dekker.

Boyle RW (1968) *The Geochemistry of Silver and its Deposits*. Quebec: Geological Survey of Canada.

Boyle RW (1987) *Gold: History and Genesis of Deposits*. New York: Van Nostrand Reinhold.

Cabri LJ (ed.) (1981) *Platinum-Group Elements Mineralogy, Geology, Recovery*. Quebec: Canadian Institute of Mining and Metallurgy.

Cox PA (1995) *The Elements on Earth*. New York: Oxford University Press.

Emsley J (1991) *The Elements*, 2nd edn. Oxford: Clarendon Press.

Gaines RV, Skinner HCW, Foord EE, Mason B, and Rosenzweig A (eds.) (1997) *Dana's New Mineralogy: The System of Mineralogy of James Dwight Dana and Edward Salisbury Dana*, 8th edn. London: Wiley Europe.

Gasparrini C (1993) *Gold and Other Precious Metals*. Berlin: Springer-Verlag.

Guilbert JM and Park CF Jr (1986) *The Geology of Ore Deposits*. New York: W H Freeman.

Mandarino JA (1999) *Fleischer's Glossary of Mineral Species 1999*, 8th edn. Tucson: Mineralogical Record Inc.

Williams PA (1990) *Oxide Zone Geochemistry*. Chichester: Ellis Horwood.

Nitrates

P A Williams, University of Western Sydney, Penrith South DC, NSW, Australia

Introduction

Naturally occurring nitrate mineral deposits are found in a number of different environments, but most are found in arid climates or in areas in which evaporation rates are high. This is because nitrate salts are generally highly soluble in water; exceptions are the metastable nitrate-bearing members of the hydrotalcite groups and certain basic copper salts.

Geographical Distribution of Nitrate Minerals

Despite the high degree of solubility of most nitrates in water, aside from basic salts, they are quite widely distributed in the natural environment, given appropriate conditions for their crystallization. Simple nitrates (Table 1), including gwihabaite, nitratine, niter, nitrobarite, nitrocalcite, and nitromagnesite, are found in quantities ranging up to thousands of tonnes in dry caves, where rates of evaporation are significant. The nitrate in this case (together with the ammonium ion in the case of gwihabaite) is usually derived from the decomposition of bat guano or is derived from the leaching of nitrate-bearing volcanic or sedimentary rocks. Some of the simple anhydrous species are also formed as ephemeral efflorescences on rock faces and in playa lakes in arid climates. Aside from this, both simple and complex nitrates (again, excluding the basic salts) are present in quantities amounting to millions of tonnes in the celebrated Chilean nitrate deposits of the Atacama Desert. These represent a significant source of nitrate fertilizer and have been mined as such for centuries. The rare mineral ungemachite was first found in during mining operations in the oxidized zone of the Chuquicamata copper deposit together with clinoungemachite; the mine is located in the Atacama region and ungemachite has been identified elsewhere only in a single specimen from the oxidized zone of the New Cobar copper–gold deposit near Cobar, New South Wales, Australia. The caliche-hosted nitrate deposits of the Atacama Desert contain other highly oxidized species (besides nitrate, with nitrogen in its highest oxidation state), including bromate, iodate, periodate, perchlorate, and native iodine.

Three nitrate-bearing members of the hydrotalcite group containing nickel are confined to a single locality, the Mbobo Mkulu cave in South Africa. Nitrate is formed from the leaching of nitrate-bearing volcanic or sedimentary rocks, and nickel and copper are

Table 1 Nitrate(v) minerals

Type	*Name*	*Composition*
Simple nitrates	Gwihabaite	$(NH_4,K)NO_3$
	Nitratine	$NaNO_3$
	Niter	KNO_3
	Nitrobarite	$Ba(NO_3)_2$
	Nitrocalcite	$Ca(NO_3)_2 \cdot 4H_2O$
	Nitromagnesite	$Mg(NO_3)_2 \cdot 6H_2O$
Hydrotalcite-related phases	Hydrombobomkulite	$(Ni,Cu)Al_4[(NO_3)_2,(SO_4)](OH)_{12} \cdot 12–14H_2O$
	Mbobomkulite	$(Ni,Cu)Al_4[(NO_3)_2,(SO_4)](OH)_{12} \cdot 3H_2O$
	Nickelalumite	$(Ni,Cu)Al_4[(SO_4),(NO_3)_2](OH)_{12} \cdot 3H_2O$
	Sveite	$KAl_7(NO_3)_4Cl_2(OH)_{16} \cdot 8H_2O$
Compound nitrate-sulphates	Darapskite	$Na_3(NO_3)(SO_4) \cdot H_2O$
	Humberstonite	$K_3Na_7Mg_2(NO_3)_2(SO_4)_6 \cdot 6H_2O$
	Ungemachite	$K_3Na_8Fe(NO_3)_2(SO_4)_6 \cdot 6H_2O$
Basic copper nitrates	Gerhardtite	$Cu_2NO_3(OH)_3$
	Rouaite	$Cu_2NO_3(OH)_3$
	Likasite	$Cu_3NO_3(OH)_5 \cdot 2H_2O$
	Buttgenbachite	ca $Cu_{36}(NO_3)_2Cl_8(OH)_{62} \cdot 4–10H_2O$

derived from oxidizing sulphide minerals in associated shale horizons. Sveite is also a cave mineral; it has been found in the Autana cave, Venezuela, and has been reported as a coating on rock faces at an unspecified California locality. All of these minerals are exceptionally rare and are based on the hydrotalcite structure, $Mg_6Al_2CO_3(OH)_{16} \cdot 4H_2O$, consisting of an $M(OH)_2$ layer that sandwiches layers containing anions, metal ions, and water molecules. The minerals are metastable phases with respect to Al(III) oxyhydroxides and other soluble components.

The basic copper(II) nitrates of Table 1 are an exception to the pattern of nitrate mineral distribution in arid regions in that they are found sparingly in the oxidized zones of a number of copper orebodies. Perhaps the most common of the minerals is buttgenbachite. The composition of this mineral is complicated by the fact that it is involved in a complex solid-solution series involving the sulphate-chloride analogue, connellite. Various amounts of chloride and hydroxide ions are found in the lattice, depending on the composition of the solution from which the mineral crystallizes and with respect to requirements for charge compensation. Connellite, the sulphate-dominant analogue, almost invariably carries some nitrate, although the nitrate-free end-member can be synthesized.

The origin and persistence of nitrate in the natural environment are both complex phenomena. Biological activity serves both to produce and to consume nitrate, depending on the micro-organisms involved. The accumulation of vast amounts of nitrate in the Chilean deposits is due to the extreme paucity of living matter that would ordinarily consume it. Large amounts of nitrate minerals can accumulate only in the absence of significant biological activity or in a steady-state environment when losses due to metabolic processes are balanced by biological nitrate production or inputs from other sources. In this connection, aside from contributions from leaching of rocks, nitrate is formed in significant amounts by electrical discharge in the atmosphere during thunderstorm activity.

Despite the large quantities of some of the world's nitrate deposits, including those found in caves, nitrates generally are rare minerals and are confined to special environments. Aside from the few simple minerals that dominate the exploitable resources, nitrate minerals are of interest only in the academic sense or to collectors.

See Also

Minerals: Definition and Classification; Chromates. **Sedimentary Rocks:** Evaporites.

Further Reading

Anthony JA, Bideaux RA, Bladh KW, and Nichols MC (2003) *Handbook of Mineralogy. Volume 5. Borates, Carbonates, Sulfates, Chromates, Germanates, Iodates, Molybdates, Tungstates, etc., and Organic Materials.* Tucson, AZ: Mineral Data Publishing.

Blackburn TH (1983) The microbial nitrogen cycle. In: Krumbein WE (ed.) *Microbial Geochemistry*, pp. 63–90. Oxford: Blackwell Scientific Publications.

Böhlke JK, Ericksen GE, and Revesz K (1997) Stable isotope evidence for an atmospheric origin of desert nitrate deposits in northern Chile and southern California, USA. *Chemical Geology* 136: 135–152.

Doner HE and Lynn WC (1989) Carbonate, halide, sulfate, and sulfide minerals. In: Dixon JB and Weed SB (eds.) *Minerals in Soil Environments*, 2nd edn. pp. 331–378. Madison, WI: Soil Science Society of America.

Ericksen GE (1981) *Origin of the Chilean Nitrate Deposits. United States Geological Survey Professional Paper, 1118*. Washington, DC: US Geological Survey.

Fenchel T, King GM, and Blackburn TH (1998) *Bacterial Biogeochemistry – The Ecophysiology of Mineral Cycling*, 2nd edn. New York: Academic Press.

Gaines RV, Skinner HCW, Foord EE, Mason B, and Rosenzweig A (eds.) (1997) *Dana's New Mineralogy: The System of Mineralogy of James Dwight Dana and Edward Salisbury Dana*, 8th edn. London: Wiley Europe.

Hill C and Forti P (eds.) (1997) *Cave Minerals of the World*, 2nd edn. Huntsville, AL: National Speleological Society.

Mandarino JA (1999) *Fleischer's Glossary of Mineral Species 1999*, 8th edn. Tucson, AZ: Mineralogical Record Inc.

Roberts WL, Campbell TJ, and Rapp GR, Jr (1990) *Encyclopedia of Minerals*, 2nd edn. New York: Van Nostrand Reinhold.

Williams PA (1990) *Oxide Zone Geochemistry*. Chichester: Ellis Horwood.

Olivines

G Cressey, The Natural History Museum, London, UK
R A Howie, Royal Holloway, University of London, London, UK

Introduction

Olivine is an important rock-forming mineral and occurs in mafic igneous rocks such as peridotite, gabbro, and basalt, along with pyroxene and plagioclase. The ultramafic rock dunite consists almost entirely of olivine, and is the result of the fractional crystallisation of peridotite-composition melts to form cumulate olivine. Magnesium-rich olivine is abundant in the majority of mantle-derived ultramafic nodules in kimberlite pipes. Olivine is common in the Earth, Moon, (*see* **Solar System:** Moon) and in the stony meteorites, and much research is conducted into the physical, chemical, and mechanical properties of olivine in order to understand the physical state of the interior of the Earth, the origin of meteorites, and the history and processes on planetary bodies (*see* **Solar System:** Mars).

Crystal Chemistry of the Olivine Group

Crystals of olivine possess orthorhombic symmetry and have a structure consisting of an array of individual (SiO_4) tetrahedral units that do not share linking oxygen atoms (unlike the pyroxenes, amphiboles, or feldspars which form linked SiO_4 polymers). The oxygen atoms that form the individual tetrahedral units of the olivine structure are arranged in planes perpendicular to the x-axis; in each plane the oxygen atoms are in approximate hexagonal array, but they are not touching. The tetrahedra point alternately up and down in the x and y directions of the unit cell, and there are two non-equivalent sets of cation sites linking the tetrahedra. Half of the cation sites lie on a centre of symmetry between the tetrahedral bases (M1 sites), and half lie on mirror planes between edges and bases of tetrahedra (M2 sites). Although in each of these sites the cations are in six-coordination with oxygen atoms of neighbouring SiO_4 tetrahedra, the cation sites are geometrically distinct (**Figure 1**). At elevated temperatures, the cation–oxygen bond lengths of both M1 and M2 sites increase, but the SiO_4 tetrahedra remain the same, thus reflecting the greater bond strength of Si–O compared with that between other cations and oxygen. In most natural olivines, Mg^{2+} or Fe^{2+} occupy the cation sites. Because of their similar ionic size, Mg^{2+} (ionic radius 0.72 Å in 6-coordination with oxygen) or Fe^{2+} (ionic radius 0.78 Å) can be accommodated in either cation site, and there exists a complete solid solution series between these two end-members having compositions Mg_2SiO_4 (forsterite) and Fe_2SiO_4 (fayalite). Almost all natural olivines have compositions intermediate between these two extremes, the exact composition reflecting the availability of Mg and Fe during crystallisation. Mg^{2+} and Fe^{2+} can occupy M1 and M2 with almost equal preference, so Mg^{2+} and Fe^{2+} remain disordered over these two sites down to low temperatures, and the solid solution is close to ideal with no tendency for ordering or exsolution to occur. However, it has been established that there is a slight tendency for Fe^{2+} to occupy the M1 site rather than the M2 site (this is paradoxical since M1 is the smaller site and Fe^{2+} the larger cation), therefore the Mg–Fe distribution is not completely random, but this effect is only slight.

Larnite, Ca_2SiO_4, does not have an olivine-type structure, but Ca^{2+} ions can be accommodated in the olivine structure on half the cation sites, i.e.,

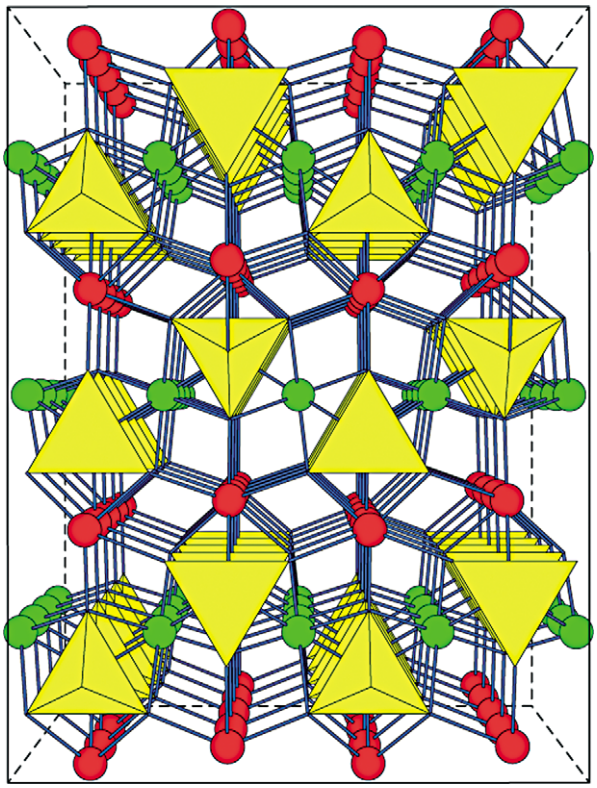


Figure 1 The olivine structure (view looking down the *x*-axis of the unit cell): SiO_4 tetrahedral units (yellow) point alternately either way in both the *x* and *y* directions, and there are two different cation sites, M1 (green) and M2 (red).

$CaMgSiO_4$, forming the mineral monticellite. The Fe analogue, $CaFeSiO_4$ (kirschsteinite) is known only in synthetic systems, but complete solid solution is possible between montcellite and kirschsteinite as Fe replaces Mg in the structure. The most iron-rich naturally occurring examples in the $CaMgSiO_4$–$CaFeSiO_4$ series contain about 70% $CaFeSiO_4$ component. However, there is very limited solid solution between the montcellite-kirschsteinite series and the forsterite-fayalite series, because the size of the Ca^{2+} ion (ionic radius 1.00 Å) is substantially larger than Mg^{2+} and Fe^{2+} ions. Substitution of the large Ca^{2+} ion into the olivine structure causes strain that distorts the SiO_4 tetrahedra.

The olivine structure cannot tolerate the degree of distortion that would be produced by a random occupation of the cation sites by ions of such different sizes, so very little solid solution can exist between the Ca-olivines and the Mg, Fe-olivines, even at high temperatures. In the Ca-bearing olivines, the Ca and Mg, Fe ions are completely ordered in the structure; Ca^{2+} occupies the (larger) M2 site, while Mg^{2+} and Fe^{2+} are randomly distributed on the M1 sites. In many natural Mg-Fe olivines, particularly those rich in Fe, there is often a small proportion of Ca and Mn present in the structure. The substitution of Mn^{2+} (ionic radius 0.83 in 6-coordination) for Fe^{2+} in fayalite also occurs when Mn is present during crystallization, and a complete solid solution series exists between Fe_2SiO_4 (fayalite) and Mn_2SiO_4 (tephroite). In natural occurrences, the low Mg-content of fayalite-tephroite phases is more likely to be related to the limited amounts of Mg available in such crystallization environments, rather than the inability of the structure to tolerate differences between these cations sizes. Zn-bearing tephroites have been reported, and $CaMnSiO_4$ (glaucochroite), although rare in nature, has an olivine structure.

Other compositions that also possess the olivine structure, such as Ni_2SiO_4 and Mg_2GeO_4, have been prepared by laboratory synthesis. Small amounts of Ni^{2+} (ionic radius 0.69 Å in 6-coordination) are commonly present in the structure of natural Mg-rich olivines. However, the presence of Cr^{3+} and/or Fe^{3+} (which would produce unfavourable charge imbalances in the olivine structure) is often found to be associated with exsolved sub-microscopic crystallites of chromite or magnetite inside the olivine. Fe^{3+} is also likely to be present in the oxidation products commonly formed during the hydrothermal alteration of olivine.

Nomenclature

Compositions within the forsterite-fayalite isomorphous solid solution series are the most abundant of the naturally occurring members of the olivine group, and the term olivine has come to signify compositions between these two end members, with general formula $(Mg,Fe)_2SiO_4$. In the past, different names have been assigned to specific ranges of Mg:Fe ratio (e.g., chrysolite, hyalosiderite, hortonolite, ferrohortonolite), but it is preferable to indicate the composition by giving the mole proportion of either the forsterite (Fo) or the fayalite (Fa) component in the solid solution. This can be expressed in different ways; for example, an olivine with specific composition $Mg_{1.8}Fe_{0.2}$ SiO_4 can be expressed as Fo_{90} (or Fa_{10}), meaning that the olivine contains 90% of forsterite component (and 10% of fayalite component). Alternatively, the cation site occupancy in this olivine structure can be expressed as $X_{Mg} = 0.90$, $X_{Fe} = 0.10$, where $X_{Mg} = Mg/(Mg + Fe)$, $X_{Fe} = Fe/(Mg + Fe)$, and $X_{Fe} + X_{Mg} = 1$. This nomenclature can be extended to any number

of components, if small amounts of Mn, Ca, Ni, Zn, etc. are present.

Physical Properties and Stability

The densities of olivine solid solutions vary smoothly and linearly from forsterite (3.22 gm cm^{-3}) to fayalite (4.39 gm cm^{-3}). Similarly, their unit cell dimensions and optical properties vary systematically with composition. At atmospheric pressure the melting point of pure forsterite occurs at 1890°C; under anhydrous conditions the melting point increases with pressure, and under water-saturated conditions the melting point decreases with pressure. Pure fayalite melts at a much lower temperature (1205°C at 1 atm) than forsterite, because the Fe^{2+}-oxygen bonds are weaker than the shorter Mg^{2+}-oxygen bonds. For cations of the same charge, bonds to oxygen are weaker for the larger cation, so as more of the larger cations enter the structure there is a progressive reduction in the melting points of intermediate compositions. This is reflected in the crystallisation behaviour of olivine solid solutions; the first olivine crystals to separate from a liquid of given composition are richer in Mg than those of later in the crystallisation sequence, and consequently the larger Fe^{2+} ions are concentrated in the residual liquid.

Igneous olivines formed during a slow, undisturbed cooling regime will have sufficient opportunity to re-equilibrate their Mg-Fe composition with the coexisting liquid as the temperature gradually falls, so that the final crystals have a single composition. However, rapid cooling can often result in zoned olivine crystals that have a Mg-rich core and progressively more Fe towards their rims; the core represents the initial composition to form which was subsequently prevented from equilibrating with the remaining liquid by a protective overgrowth. In very rapidly cooled magmas, olivines sometimes form dendritic crystals. The morphology of olivine crystals may be equant, tabular, acicular, or dendritic; these can, in general, be related to the rate of cooling of the magma during crystallisation. In a natural silicate melt, the melting point of Mg-rich olivine is much depressed, such that crystals of composition Fo_{88} will melt in basalt magma at about 1250°C.

At high pressures in the deep earth, the olivine structure transforms to a denser spinel-type structure. For forsterite-rich compositions, the transformation to a distorted spinel phase occurs at about 140 kbar pressure at a temperature of 1000°C. This structural transition is of geophysical significance, as it is believed to be the cause of the abrupt change in seismic velocity in the Earth's mantle at about 400–500 km depth.

Alteration

Olivine is very susceptible to hydrothermal alteration, during low-grade metamorphism and weathering, and readily reacts to form serpentine and a range of iron oxides, chlorites, and smectite clays. Serpentinisation is the most common and widespread alteration process to affect olivine-rich rocks such as dunite and peridotite. The serpentine polymorphs chrysotile, lizardite, and antigorite, along with brucite, talc, and carbonates (if CO_2 is also present in the fluid) are commonly the products of this retrograde metamorphism. The following chemical re-arrangement represents the serpentinisation process:

$$\underset{\text{forsterite}}{2Mg_2SiO_4} + 3H_2O = \underset{\text{serpentine}}{Mg_3Si_2O_5(OH)_4} + \underset{\text{brucite}}{Mg(OH)_2}$$

This hydrothermal alteration of forsterite takes place at about 400°C, and for an olivine of composition Fo_{90} this reaction takes place nearer 340°C. Any iron present in the original olivine can only be accommodated in the serpentine structure to a very limited extent and is, therefore, usually converted to iron oxide or oxyhydroxide (magnetite, hematite, or goethite) inclusions that give the otherwise green serpentine a dark brownish or reddish appearance.

Occurrence

Pure forsterite is rare, but is found in certain metamorphosed siliceous dolomitic limestones (marbles). Olivine with compositions between Fo_{96} and Fo_{82} are major constituents in ultramafic rocks such as dunite and peridotite, and in prograde metamorphosed serpentintes. Mantle-derived ultramafic nodules in basalts and kimberlites usually contain olivines in the range Fo_{91} to Fo_{86}. Olivines with compositions Fo_{80} to Fo_{50} are common in gabbros and basalts (Figure 2); the actual forsterite-fayalite composition provides a useful index of fractionation for these basic igneous rocks. A wide range of olivine compositions from Fo_{98} to Fo_{40} (often as zoned crystals) and occasional almost pure fayalite grains are important constituents in some carbonaceous chondrite meteorites. Fe-rich olivines occur in alkaline and acid igneous rocks; fayalite is present in small amounts in many volcanic rocks such as

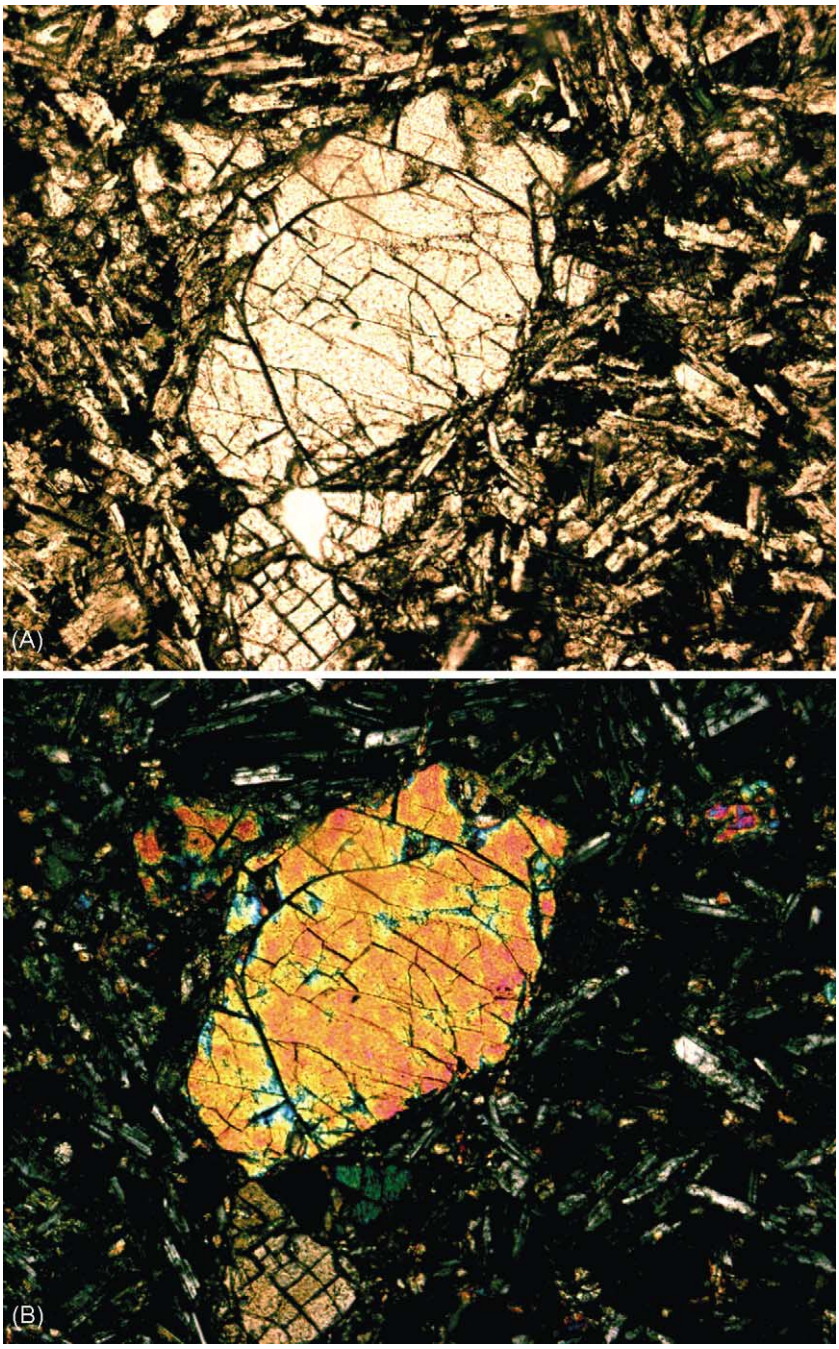


Figure 2 Basalt (from Melbourne, Victoria) seen in thin section through a polarizing microscope showing a euhedral phenocryst of olivine in plane polarized light (A), and with crossed polars (B).

obsidian, rhyolite, trachyte, and phonolite. Fayalitic and tephroitic olivines occur in rocks of metamorphic origin, such as thermally and regionally metamorphosed Fe-rich and Mn-rich sediments, respectively. Monticellite occurs in metamorphosed siliceous and magnesian limestones and in skarn zones developed by metasomatic alteration of limestones at intrusive igneous contacts. Monticellite also occurs in kimberlite and alnöite.

Meteoritic Olivine

Olivine is a major component of stony meteorites and some stony-irons (pallasites). In the former, it

occurs mainly in rounded chondrites. It is also found in some meteorites, including the almost monomineralic Chassigny stone and some basaltic shergottites. These are believed to come from Mars. Likewise, it is found in lunar-sourced meteorites and lunar surface breccias.

See Also

Earth: Mantle. **Gemstones**. **Igneous Rocks:** Komatiite. **Minerals:** Definition and Classification; Other Silicates. **Rocks and Their Classification**. **Solar System:** Moon; Mars.

Further Reading

Boyd FR and Nixon PH (1978) Ultramafic nodules from the Kimberly pipes, South Africa. *Geochemica et Cosmochemica Acta* 42: 1367–1382.

Deer WA, Howie RA, and Zussman J (1992) *An introduction to the Rock-Forming Minerals*. London: Longman.

Deer WA, Howie RA, and Zussman J (1997) Rock Forming Minerals, vol. 1A, p. 219, Orthosilicates. London: The Geological Society.

Donaldson CH (1976) An experimental investigation of olivine morphology. *Contributions to Mineral Petrology* 57: 187–213.

Ghose S, Wan C, and McCallum IS (1976) Fe^{2+}-Mg^{2+} order in an olivine from the lunar anorthosite 67075 and the significance of cation order in lunar and terrestrial olivines. *Industrial Journal of Earth Science* 3: 1–8.

Green DH (1964) The petrogenesis of the high-temperature peridotite intrusion in the Lizard area, Cornwall. *Journal of Petrology* 5: 134–188.

MacKenzie WS, Donaldson CH, and Guilford C (1982) *Atlas of Igneous Rocks and Their Textures*. Longman.

MacKenzie WS and Guilford C (1980) *Atlas of Rock-Forming Minerals in Thin Section*. London: Longman.

Wenk HR and Raymond KN (1973) Four new refinements of olivine. *Zeit. Krist.* 137: 86–105.

Other Silicates

R A Howie, Royal Holloway, University of London, London, UK

Introduction

Silicates not grouped in the major series separately dealt with, but which are of geological importance, include the garnet group, the aluminosilicates (the Al_2SiO_5 polymorphs), staurolite, the epidote group, cordierite, tourmaline, the chlorite group, and the serpentines.

The Garnet Group

The minerals (*see* **Minerals:** Definition and Classification, **Gemstones**) of the garnet group are particularly characteristic of metamorphic rocks but are also found in some igneous rocks, and as detrital grains in sediments. The group is subdivided into six principal species, which represent the end-members of an isomorphous series:

Pyrope	$Mg_3Al_2Si_3O_{12}$
Almandine	$Fe_3^{2+}Al_2Si_3O_{12}$
Spessartine	$Mn_3Al_2Si_3O_{12}$
Grossular	$Ca_3Al_2Si_3O_{12}$
Andradite	$Ca_3(Fe^{3+},Ti)_2Si_3O_{12}$
Uvarovite	$Ca_3Cr_2Si_3O_{12}$

A garnet corresponding in composition with any one end-member is rare, however, and the name is assigned according to the dominant 'molecular' type present. The garnets may be considered as two series: pyralspite (*py*rope, *al*mandine, *sp*essartine) and ugrandite (*u*varovite, *gr*ossular, *and*radite). Fairly complete and continuous variation in composition occurs within these two series, but there appears to be no continuous variation between them.

Garnets are cubic and are normally isotropic, but some calcic varieties show complex and sector twinning, which may be visible in birefringent varieties. Their colour ranges from red, orange, brown, and black for the pyralspites, to green, yellow, brown, and black for the ugrandites; they are colourless to pale pink, yellow, or brown in thin section. They are notable in having a strong power of crystallization, typically forming rhombic dodedcahedra or icositetrahedra.

The pyrope–almandine series occur in metamorphic rocks of a variety of types, of which the garnet-mica schists of regional metamorphism (*see* **Regional Metamorphism**) are typical; they are also found in certain ultrabasic rocks such as the mica peridotites and kimberlites. The pyralspites are also

known in peraluminous volcanic rocks, granites, and pegmatites. The distribution of Fe and Mg between coexisting equilibrium assemblages of garnet and biotite or garnet and clinopyroxene can be used in geothermometry. The equilibrium constant $K_{D\ Mg/Fe}^{Gt/Bi}$ for different metamorphic grades is shown in Figure 1. For the garnet–clinopyroxene pair, the pressure–temperature dependence of the Fe–Mg partition coefficient has been determined experimentally; the use of thermodynamic data for the garnet group is illustrated in Figure 2. Purplish chrome-pyrope is typical of diamondiferous kimberlites and peridotitic xenoliths.

Garnets with appreciable spessartine are usually found in granitic pegmatites and skarns.

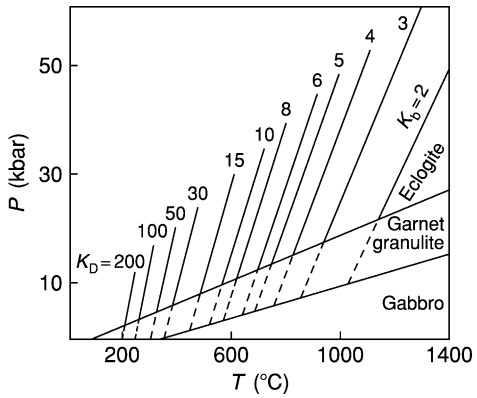


Figure 1 $K_{D\ Fe/Mg}^{Gt/Cpx}$ as a function of pressure and temperature (after Råheim and Green (1974)).

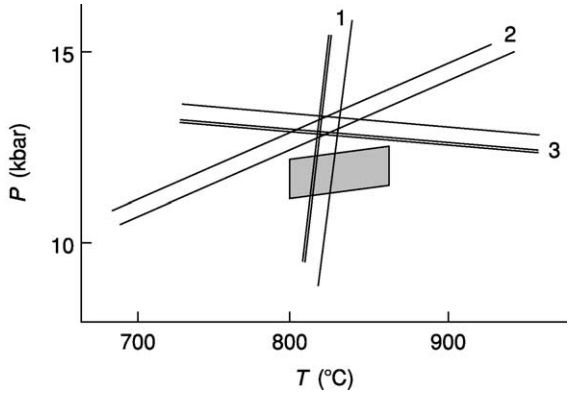


Figure 2 Thermodynamic data for natural equilibrium assemblages from pyroxene granulites of South Harris, Scotland, derived from: (1) Fe–Mg distribution between garnet and clinopyroxene; (2) the activity coefficient for the $Ca_3Al_2Si_3O_{12}$ component in garnet using the equilibrium 3 $CaAl_2Si_2O_8$ (plagioclase) = $Ca_3Al_2Si_3O_{12}$ + 2 Al_2SiO_5 + SiO_2; and (3) the activity coefficient for the equilibrium $CaAl_2Si_2O_8$ = $CaAl_2SiO_6$ (clinopyroxene) + SiO_2. The shaded area is the apparent region of 'bracketing' of metamorphic pressure and temperature derived from other equilibria (after Wood 1975 and Wood (1997)).

Grossular is characteristically found in regionally or thermally metamorphosed impure calcareous rocks; andradite also occurs in such rocks but is more typically found in metasomatic skarn deposits.

The black garnet melanite is a Ti-bearing andradite occurring in alkaline igneous rocks and some skarns. Uvarovite is rare and any appreciable Cr content in metasomatic garnets can be detected from their green colour.

The aluminosilicates (Al_2SiO_5)

This family comprises three polymorphs, all typically found in metamorphosed pelitic rocks formed in different metamorphic conditions. Their occurrence is, thus, of particular interest in metamorphic petrology.

Sillimanite is orthorhombic, typically colourless or white, and found as optically length-slow prismatic crystals or needles.

Andalusite is also orthorhombic, typically pink in hand specimen, and found as stout, length-fast prismatic crystals. In low-grades of metamorphism the anhedral grains frequently encloses a host of inclusions of foreign matter, but rapidly acquire a prismatic outline pushing aside the impurities which then form a characteristic chiastolite pattern.

Kyanite is triclinic, typically giving inclined exctinction and having a blue colour in hand specimen.

The experimentally determined phase diagram for the Al_2SiO_5 composition is given in Figure 3, indicating the position of the triple point between sillimanite, andalusite, and kyanite.

The application of these results to metamorphic petrology, thus, clearly shows that sillimanite is the high-temperature phase, to be found in the highest

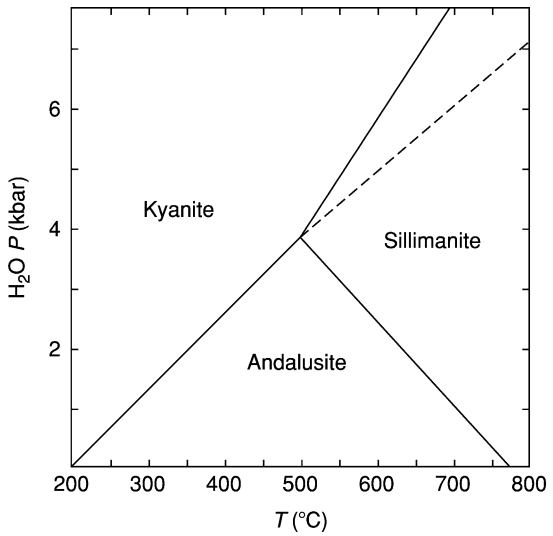


Figure 3 Phase diagram for kyanite–andalusite and kyanite–sillimanite equilibria.

grades of thermal metamorphism of pelitic rocks or in the highest grade of regional metamorphism of similar rocks. Andalusite is normally confined to thermally metamorphosed pelitic rocks, but may be found in some granites, probably as a result of contamination. Kyanite is typically a mineral of high-grade metamorphism of pelitic rocks, but has also been found in thermal aureoles in situations where appreciable shear-stress arose during emplacement of the igneous intrusion.

Staurolite

This yellow to brown ferromagnesian aluminous silicate characteristically occurs in medium grade metamorphosed pelitic schists and gives its name to a regional metamorphic zone. The staurolite–garnet–kyanite assemblage is typical of Barrovian-type metamorphic terrain, whereas the staurolite–cordierite–andalusite/sillimanite assemblage results from the lower pressure Buchan-type metamorphism. During the prograde metamorphism of pelitic rocks, staurolite develops earlier than kyanite, and then together with kyanite, before being replaced by kyanite and garnet:

$$\text{staurolite} + \text{muscovite} + \text{quartz} = \text{almandine} + Al_2SiO_5 + \text{biotite} + H_2O$$

Its variable Fe/Mg ratio allow it to be used in conjunction with garnet and other ferromagnesian silicates as a geoithermometer. Staurolite is affected by retrograde metamorphism, resulting in it being mantled and partially replaced by chlorite. Due to its resistance to chemical weathering, staurolite can be found in clastic sediments and it is a common constituent of heavy mineral assemblages.

The Epidote Group

All members of the epidote group contain both single $[SiO_4]$ tetrahedral and double $[Si_2O_7]$ tetrahedral groups, with large cavities within this framework typically occupied by large cations, typically Ca. The general formula can be given as $X_2Y_3Z_3(O,OH,F)_{13}$, in which:

X = Ca, Ce, La, Y, Th, Fe^{2+}, Mn^{2+}, Mn^{3+}

Y = Al, Fe^{3+}, Mn^{3+}, Mn^{2+}, Fe^{2+}, Ti,

Z = Si, Be

The major compositions within the group are:

Zoisite (clinozoisite)	$Ca_2(Al,Fe^{3+})Al_2O \cdot OH \cdot Si_2O_7 \cdot SiO_4$
Epidote	$Ca_2(Fe^{3+},Al)Al_2O \cdot OH \cdot Si_2O_7 \cdot SiO_4$
Piemontite	$Ca_2(Mn^{3+},Fe^{3+},Al)_3O \cdot OH \cdot Si_2O_7 \cdot SiO_4$
Allanite	$(Ca,Mn,Ce,La,Y)_2(Fe^{2+},Fe^{3+},Al)_3O \cdot OH \cdot Si_2O_7 \cdot SiO_4$

Epidote minerals occur in a wide variety of parageneses. They are typical products of greenschist and epidote–amphibolite facies of regional metamorphism, but also form under conditions of thermal metamorphism and metasomatism, during late-stage crystallisation of acid igneous rocks, and in peraluminous granites and volcanic rocks. Epidote also occurs in the hydrothermal alteration of plagioclase. Epidote itself often has a distinctive pistachio green colour and relatively high birefringence. Piemontite is reddish brown or black, but in thin section displays striking pleochroism in yellow–amethyst–crimson; it occurs as a product of low-grade regional metamorphism and as a hydrothermal product associated with manganese deposits. The rare-earth-bearing allanite is typically dark brown and occurs mainly as an accessory mineral in many granitic rocks.

Cordierite

Cordierite is a ring silicate, typically with low refractive indices and birefringence, and tends to be confused with quartz or feldspars (*see* **Minerals:** Feldspathoids; Feldspars). Twinning is known on {110} and {310}, forming simple, lamellar, or cyclic twins. It has the composition $(Mg,Fe)_2[Si_5Al_4O_{18}] \cdot nH_2O$, and is mainly found in a metamorphic environment. The Fe/Mg ratios in coexisting cordierite and garnet in divariant equilibrium are functions of pressure and temperature and may be of use in geothermometry or geobarometry. It also occurs in contaminated igneous rocks, for example, cordierite norites.

Tourmaline

Tourmaline is a boron-bearing ring silicate with a complex chemistry. Structural determinations have shown that its general formula can be taken as $XY_3Z_6B_3Si_6(O,OH)_{30}(OH,F)$. The principal recognised varieties can be distinguished in terms of their X, Y and Z cations;

	X	Y	Z
Elbaite	Na	Al,Li	Al
Olenite	Na	Al	Al
Dravite	Na	Mg	Al
Schorl	Na	Fe^{2+}	Al

Tsilaisite	Na	Al,Mn	Al
Buergerite	Na	Fe^{3+}	Al
Liddicoatite	Ca	Li,Mg	Al
Uvite	Ca	Mg	Al,Mg
Feruvite	Ca	Fe^{2+}	Al,Fe^{3+},Mg
Ferridravite	Na	Mg	Fe^{3+}
Chromdravite	Na	Mg	Cr
Foitite	□	Fe^{2+}	Al,Fe^{3+}

In the *X* position, Na may be partially replaced by K or Ca if valency conditions are satisfied; thus, in uvite the substitution NaAl = CaMg. For the commonest tourmaline varieties, *Y* can be predominantly Fe^{2+} as in schorl, Mg as in dravite, or (Al + Li) as in the elbaite series.

Optically, pleochroism is variable in intensity but is particularly strong for the iron-bearing tourmalines. The absorption is always $\omega > \varepsilon$, with the result that maximum absorption occurs when the *z*-axis is lying perpendicular to the vibration direction of the polariser. Thus schorl is black, but in thin section shows strong pleochroism from blue to yellow. Dravite is brown, while in the elbaites show a range of colours in hand specimen from red (rubellite), green (verdelite), blue (indicolite), to bi-coloured or colourless (achroite). Tourmaline crystals typically show striations along the length of prismatic faces.

Tourmaline is typically a mineral of granite pegmatites, pneumatolytic veins, and some granites; it is also commonly found in metamorphic rocks, as a product of boron metasomatism, or as a result of recrystallisation of dertrital grains from the original sediment. In granitic rocks, the tourmalines belong to the schorl–elbaite series and are generally iron-rich; typical tourmaline-bearing granites have black prismatic crystals visible in hand specimen and showing yellow or bluish yellow pleochroism in thin section. Both black and coloured tourmalines occur in the original pegmatitic phase, while rubellite and the zoned tourmalines are mainly restricted to the sodium replacement unit in the so-called pneumatogenic stage.

The Chlorite Group

The term chlorite is a portmanteau name for a mineral with a layered structure resembling the micas, but generally with a green colour. The chlorites principally occur in low-grade regionally metamorphosed rocks, as hydrothermal alteration products of ferromagnesian minerals in igneous rocks, and together with clay minerals in argillaceous sediments.Their general formula can be given as (Mg, $Fe^{2+}Fe^{3+}$,Mn,Al)$_{12}$[(Si,Al)$_8$O$_{20}$](OH)$_{16}$. The chlorites have a perfect basal cleavage and occur commonly as fine-grained scaly or massive aggregates.

The occurrence of very substantial, varied, and often continuous cation substitution in the chlorites has led to a proliferation of varietal names based mainly on chemical composition; commonly these cannot be specifically characterized optically.

The crystal structure (Figure 4) consists of regularly alternating negatively charged tetrahedral–octahedral–tetrahedral, 2:1, layers: talc-like layers with composition $Y_6Z_8O_{20}(OH)_4$ and positively charged interlayer brucite-like layers, $Y_6(OH)_{12}$, where *Y* and *Z* represent octahedral and tetrahedral sites, respectively. The chlorites are monoclinic

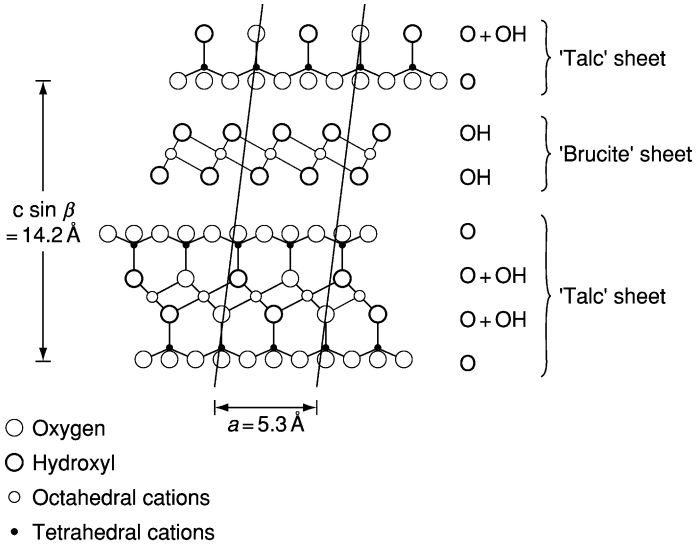


Figure 4 Idealized crystal structure of chlorite; projection on (010) (after Brown and Bailey (1962)).

(pseudohexagonal), like the micas, because of the offset of the layers.

In most common chlorites there are 12 octahedral cations per $O_{20}(OH)_{16}$ and approximately equivalent amounts of Al in tetrahedral and octahedral sites; such minerals are referred to as tri-octahedral chlorites. In a small number of chlorites, however, the number of octahedral cations is <10; these varieties are described as dioctahedal.

The principal factors influencing the optical properties of the chlorites are the iron/magnesium ratio (Figure 5). Most chlorites are optically positive, but the more iron-rich varieties are optically negative. Near the crossover point, where the birefringence passes through zero, abnormal birefringence colours (i.e., those not in the normal spectral range) may be seen.

In metamorphic rocks, chlorite is an early product of lithification and diagenesis, forming during the development of slaty cleavage in pelitic sediments. They are developed from mixed-layer clay minerals in such sediments, resulting from the expulsion of interlayer alkali ions, the fixation of (Fe,Mg) and increased tetrahedral substitution that occurs concurrently with increasing grain size. Chlorite is also a common constituent of rocks of the zeolite facies; in the lower part of the facies it is often associated with laumontite, stilbite, or heulandite, whereas in the upper part it occurs in assemblages with prehnite and pumpellyite. In the greenschist facies, chlorite–actinolite–epidote–albite assemblages are common and with increasing grade chlorite is involved in the production of amphiboles. In rocks of higher metamorphic grade, chlorite is a typical breakdown product associated with the retrogression of such ferromagnesian phases as garnet, staurolite, and biotite.

In igneous rocks, chlorite is a common product of the hydrothermal alteration of pyroxenes, amphibole, or biotite. Partial or complete chloritisation of biotite is particularly common in granites, and in most cases the transformation is pseudomorphous. The association of chlorite with albite and quartz is an essential characteristic of adinoles. Low-temperature hydrothermal Alpine-type veins in lowgrade metamorphosed sediments carry chlorite in addition to adularia and quartz.

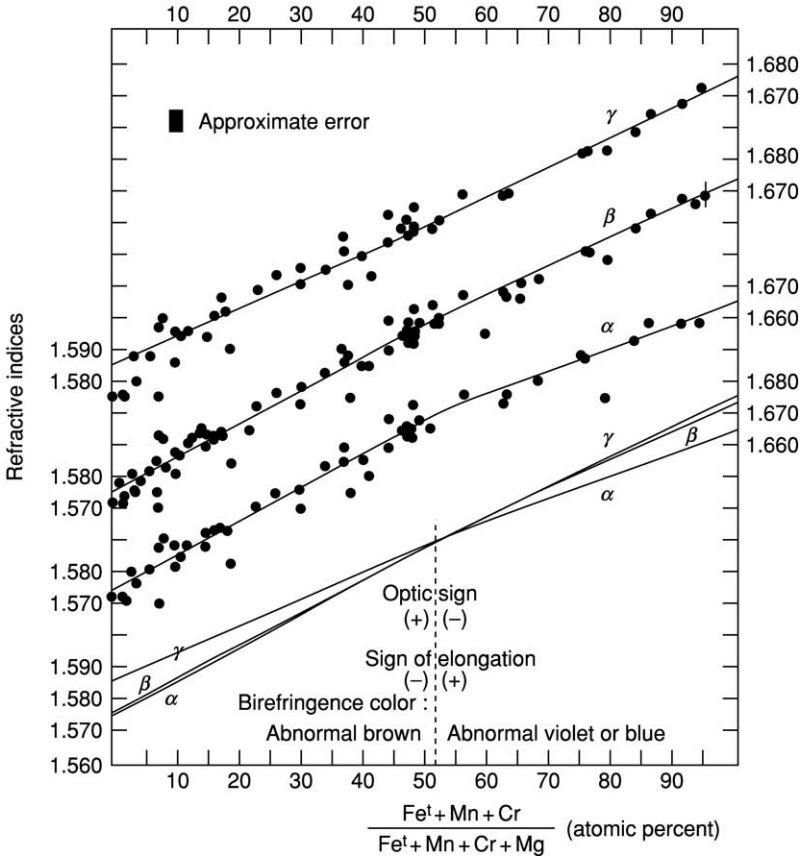


Figure 5 The relationship between optical properties and $(Fe^t + Mn + Cr)/(Fe^t + Mn + Cr + Mg)$ ratio (atomic %) for the principal chlorites (after Albee (1962)).

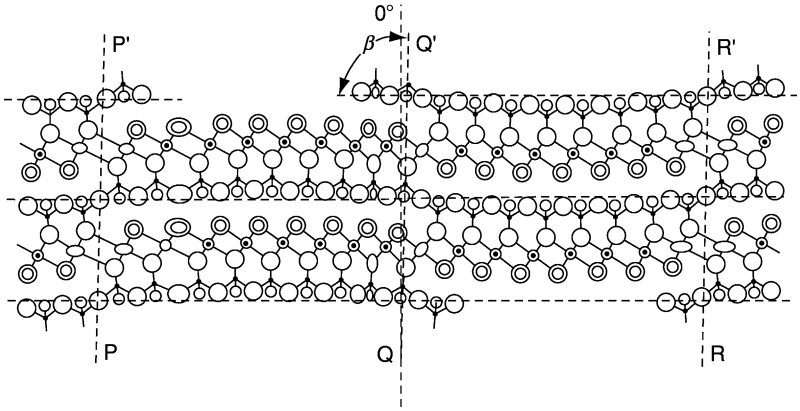


Figure 6 The crystal structure of antigorite as viewed along the *y*-axis. The curved layers (radius of curvature 75 Å) reverse polarity at PP′, RR′, and near QQ′ (after Kunze (1956)).

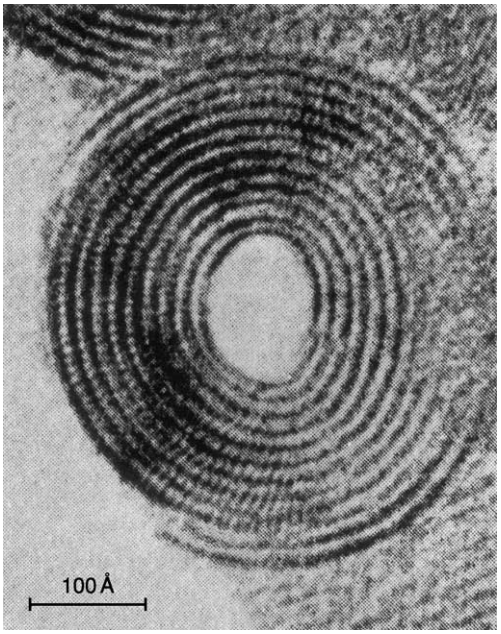


Figure 7 Transmission electron micrograph of the cross-section of a fibril of chrysotile. The structural layers can be seen to be in concentric cylindrical fashion.

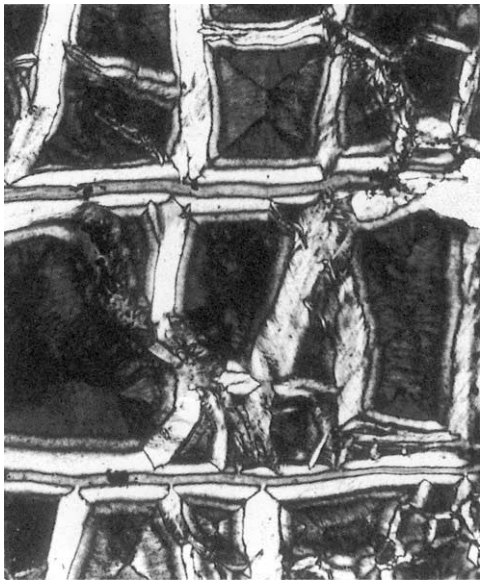


Figure 8 'Window' and 'hour-glass' textures of chrysotile and lizardite in serpentinite. Crossed polars ×120.

Serpentine

There are three principal forms of serpentine – lizardite, antigorite, and chrysotile – all with approximate composition $Mg_3Si_2O_5(OH)_4$. The most abundant is lizardite and the least common is chrysotile, though the latter is perhaps best known as it often occurs in veins of silky fibres, and was an important source of commercial asbestos. All three species have a layered structure: one part consists of a pseudo-hexagonal network of linked SiO_4 terahedra all pointing one way; joined to this sheet is a brucite layer in which, on one side only, two out of every three hydroxyls are replaced by apical oxygens of the SiO_4 tetrahedra. Although much of the matrix surrounding chrysotile veins is extremely fine-grained lizardite, electron microscopy shows that it has platy morphology. The fibrous nature of chrysotile is explained by it consisting of layers curved cylindrically or spirally, usually about the *x*-axis. The fundamental fibrils typically have average outer and inner diameters of about 250 and 75 Å, respectively (Figure 6). Antigorite is structurally distinct in having a large *a* cell parameter of ~33–51 Å, which can be related to the curvature of serpentine layers about *y* and their inversion to form a corrugated sheet structure with periodicity *a* (Figure 7).

The chemistry of most serpentines is relatively simple, with minor substitution of Al for Si and of Fe for Mg. The fine-grained nature of most serpentines makes a complete optical description impossible, but the refractive indices (1.54–1.55) and birefringes are low. They show a variety of textures, many of which ('mesh' and 'hour-glass') are pseudomorphous after olivine (**Figure 8**) or 'bastite' after pyroxene or amphibole.

The serpentine minerals typically form by the retrograde hydrothermal alteration of ultrabasic igneous rocks such as dunite and peridotite. The main mineral altered is olivine which frequently contains too much iron to be accommodated in the serpentine structure, resulting in its exsolution as iron oxide; this may occur as magnetite which alters to hematite to give the reddish mottled appearance reputedly resembling the skin of a serpent. Serpentines may also form, associated with thermally metamorphosed siliceous dolomitic limestones.

See Also

Gemstones. **Minerals:** Definition and Classification; Feldspars; Feldspathoids. **Regional Metamorphism**.

Further Reading

Albee AL (1962) Relationship between the mineral association, chemical composition and physical properties of the chlorite series. *American Mineralogy* 47: 850–870.

Bailey SW and Brown BE (1962) Chlorite polytypism: I. Regular and semi-random one layer structures. *American Mineralogist* 47: 819–850.

Deer WA, Howie RA, and Zussman J (1992) *An Introduction to the Rock-Forming Minerals,* 2nd ed. London: Longman.

Holdaway MJ (1971) Stability of andalusite and the aluminum silicate phase diagram. *American Journal of Science* 271: 97–131.

Kunze G (1956) Die gewellte Struktur des Antigorits. *Z. Krist.* 108: 82–107.

Råheim A and Green DH (1974) Talc–garnet–kyanite–quartz schist from an eclogite-bearing terrane, western Tasmania. *Contributions to Mineralogy and Petrology* 43: 223–231.

Wood BJ (1975) The influence of pressure, temperature and bulk composition on the appearance of garnet in orthogneiss – an example from South Harris, Scotland. *Earth and Planetary Science Letter* 26: 299–331.

Wood BJ (1977) The activities of components in clinopyroxene and garnet solid solutions and their applications to rocks. *Phil. Trans. Royal Society of London* 286: 331–342.

PHOSPHATES

See **SEDIMENTARY ROCKS: Phosphates**

Pyroxenes

R A Howie, Royal Holloway, University of London, London, UK

Pyroxenes are the most important group of rock-forming ferromagnesian silicates, and occur as stable phases in almost every type of igneous rock. They are also found in many rocks of widely different compositions formed under conditions of both regional and contact/thermal metamorphism.

The pyroxenes include both orthorhombic and monoclinic species: the orthopyroxenes and the clinopyroxenes. The orthopyroxenes consist essentially of a simple chemical series of (Mg, Fe)SiO_3 minerals (*see* **Minerals:** Definition and Classification) in contrast to the larger group of clinopyroxenes which have a very wide range of chemical compositions. There is a virtually complete series with a complete range of Fe/Mg ratios.

The clinopyroxenes can be considered, as a first approximation, as members of the four-component system $CaMgSi_2O_6$–$CaFeSi_2O_6$–$Mg_2Si_2O_6$–$Fe_2Si_2O_6$. The nomenclature used to describe these pyroxenes is illustrated in Figure 1 (*see* **Rocks and Their Classification**). The $Ca_2Si_2O_6$ end-member (wollastonite) is not a pyroxene (see below).

The crystal structure of the pyroxenes is based on infinite single chains of linked silicon—oxygen tetrahedra, the [Si_2O_6] groups, these chains being linked laterally by octahedral layers of (Ca, Mg, Fe, Na, etc.) cations (Figures 2 and 3). The formulae of the commoner pyroxenes are as follows:

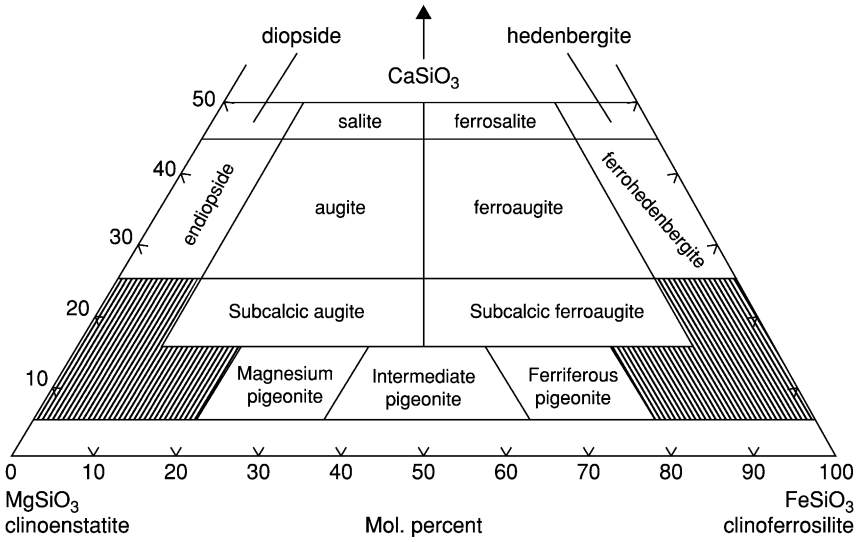


Figure 1 Nomenclature of clinopyroxenes in the system $CaMgSi_2O_6$–$CaFeSi_2O_6$–$Mg_2Si_2O_6$–$Fe_2Si_2O_6$.

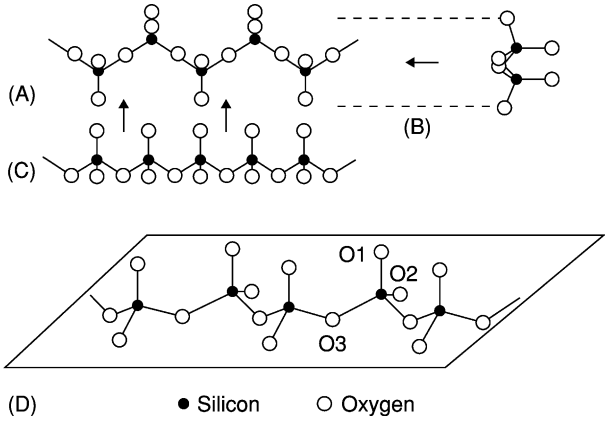


Figure 2 Idealized illustration of a single pyroxene chain as seen in projections on (A) (100), (B) along *z*, and (C) along *y*.

Orthopyroxenes	$Mg_2[Si_2O_6]$	enstatite
	$Fe_2^{2+}[Si_2O_6]$	ferrosilite
Clinopyroxenes	$CaMgSi_2O_6$	diopside
	$CaFe^{2+}[Si_2O_6]$	hedenbergite
	$(Mg, Fe^{2+}Ca)$	pigeonite
	$(Mg, Fe^{2+})[Si_2O_6]$	
	$(Ca, Mg, Fe^{2+}, Fe^{3+}, Ti, Al)_2$ $[(Si, Al)O_6]$	augite–ferroaugite
	$(Ca, Na)(Mg, Fe^{2+}, Fe^{3+}, Al)$ $[Si_2O_6]$	omphacite
	$NaAl[Si_2O_6]$	jadeite
	$NaFe^{3+}[Si_2O_6]$	aegirine
	$(Na, Ca)(Fe^{3+}, Fe^{2+}, Mg)$ $[Si_2O_6]$	aegirine-augite
	$LiAl[Si_2O_6]$	spodumene

Optically, the orthopyroxenes are colourless to greenish under the microscope and bronze to black in hand specimen. They have a good {210} cleavage and straight extinction. Their other properties vary with the Fe/Mg ratio.

The clinopyroxenes also are colourless to greenish under the microscope, have a good {210} cleavage, but with variable extinction angles from 0 to 45°. Within the $CaMgSi_2O_6$–$CaFeSi_2O_6$–$Fe_2Si_2O_6$–$Mg_2Si_2O_6$ quadrilateral, which covers 90% of the common rock-forming clinopyroxenes, their optical properties vary systematically, allowing their composition to be estimated.

The occurrence of members of the pyroxene group in general reflects the compositions of their parent rocks. Mg-rich enstatites are important constituents associated with olivine, diopside, and spinel in ultrabasic and ultramafic rocks. Other orthopyroxenes of intermediate composition are essential constituents of the granulite facies of metamorphic rocks. Diopside is formed by the metamorphism of siliceous dolomites, but is not uncommon as an early of basic igneous rocks. Pigeonite is a constituent of many andesites and dacites and can be distinguished by its small optic axial angle. In layered igneous intrusions, it may invert to lamellae of augite ± orthopyroxene. The most common clinopyroxenes are augites; these are essential constituents of basic igneous rocks, and also occur in ultramafic rocks, lherzolites and websterites, and in ultrabasic nodules and xenocrysts in basic rocks. They also occur in dolerites and in basic lavas.

The sodium-bearing clinopyroxene Jadeite is a diagnostic mineral of rocks generally considered to have formed in a high-pressure environment; it is particularly characteristic of eclogites and related

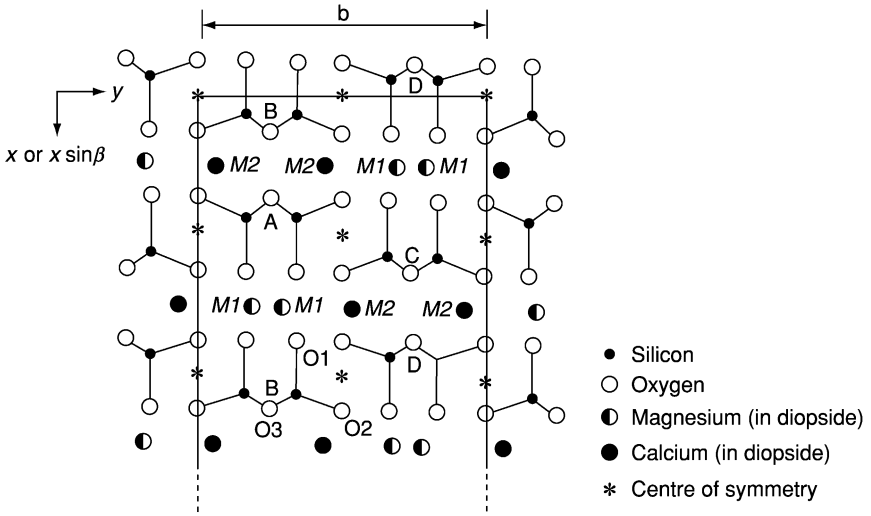


Figure 3 Projection along *z* of an idealized pyroxene structure, showing the pyroxene chains and the two distinct cation sites *M*1 and *M*2. The latter are in oxygen polyhedra which also form chains or bands parallel to *z*.

rocks. When near its end-member composition of $NaAlSi_2O_6$ Jadeite, is colourless or pale green (one variety of the gem material jade); it is typically found in metagreywackes and related rocks of regional metamorphism, particularly in the blueschist facies. Aegirine and aegirine-augite illustrate the replacement of ferric iron in the substitution $NaFe^{3+} = Ca(Mg, Fe^{2+})$; they are mainly found in alkaline or peralkaline igneous rocks. The lithium pyroxene spodumene is generally white and is restricted to Li-rich granitic pegmatites.

Wollastonite ($CaSiO_3$) and the three manganese-rich silicates rhodonite, bustamite, and pyroxmangite are all triclinic. Sometimes referred to as 'pyroxenoids', they differ from the pyroxenes in having, respectively, three, five, five, and seven silica tetrahedra in their structural chains. Wollastonite is a common mineral of metamorphosed siliceous limestones, whereas the three manganese-rich species typically occur in Mn-orebodies, usually associated with metasomatic activity.

See Also

Minerals: Definition and Classification; Amphiboles; Other Silicates. **Rocks and Their Classification**.

Further Reading

Deer WA, Howie RA, and Zussman J (1992) *An Introduction to the Rock-Forming Minerals,* 2nd edn. London: Longman.

Quartz

R A Howie, Royal Holloway, University of London, London, UK

Quartz is the second most abundant mineral in the Earth's crust being hard (7 on Mohs scale) and resistant to many forms of chemical weathering, with the result that it becomes concentrated in sedimentary rocks while the coexisting feldspars are rapidly weathered away. It has a compact framework structure and has trigonal symmetry. It is an essential mineral in many acid plutonic igneous rocks such as the granites (*see* **Igneous Rocks:** Granite) and granodiorites, and also in hypabyssal and volcanic rocks of equivalent composition. Metamorphism of such igneous or sedimentary rocks gives rise to quartzites and quartz-rich veins.

Although the chemical composition is comparatively simple, most quartz consisting of 99.9% SiO_2, it has α (low-temperature) and β (high-temperature) forms, with a transition at around 573°C, and several polymorphic forms, mostly at higher temperatures and pressures. This leads to the phase diagram for SiO_2 being extremely complex. There are the high-temperature polymorphs tridymite and cristobalite, both typical of volcanic rocks, and the more recently discovered high-pressure phases coesite and stishovite (Figure 1). Coesite occurs at a static pressure of

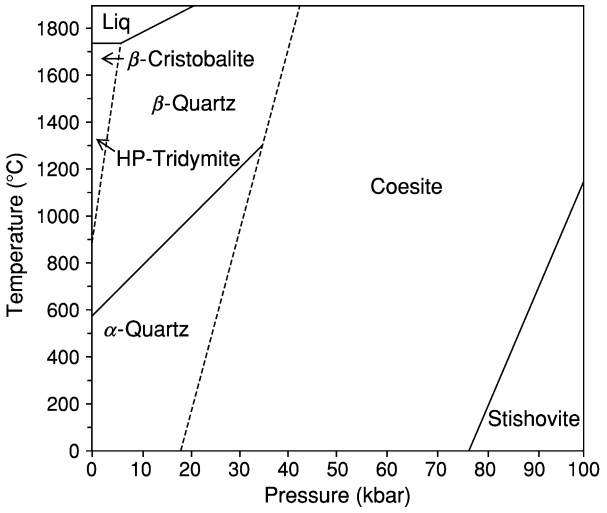


Figure 1 Phase diagram for the silica system (after Klein and Hurlbut, (1973) *Manual of Mineralogy*, 21st edn, p. 527. © *John Wiley and Sons* (Reprinted with permission).

$\geq$29 kbar; it is found as inclusions, generally partially replaced by quartz, in zircon, garnet, monazite, and other phases formed in low-temperature–high-pressure conditions and is used as an indicator of such ultrahigh-grade metamorphic conditions in rocks which have generally been exhumed relatively rapidly in orogenic belts (as in the Himalaya, the Alps and in SE China). Stishovite requires even higher pressures (of around 130 kbar) for its formation and is typical found in meteorites or meteorite impact craters, also in tektites.

Microscopic fluid inclusions often occur in quartz, and may contain a gas phase as well as a liquid and one or more microscopic crystals of such minerals as halite, sylvite, calcite, etc., believed to have been precipitated from the same source as the quartz. The development of sophisticated microanalytical techniques has allowed the composition of these liquids and solids to be determined and to the determination of the temperature at which the original precipitation and crystallization occurred. This has important applications in unravelling the petrogenetic history of both igneous and metamorphic rocks as well as in the geochemistry and conditions of deposition of ore deposits and the provenance of sedimentary rocks (*see* **Sedimentary Rocks:** Mineralogy and Classification).

Chalcedony is a group name for the compact varieties of silica composed of minute crystals of quartz with submicroscopic pores. The colour and texture vary considerably according to the impurities present, but in general such materials may be subdivided into chalcedony in which the colour is fairly uniform, and agate in which the colour may be arranged in bands or concentric zones. The terms 'chert (*see* **Sedimentary Rocks:** Chert)' and 'flint' are used for opaque dull-coloured or black chalcedony, and in common usage chert is taken as the name for this material when it occurs in stratified or massive form in rocks, while the term flint is generally restricted to dark chalcedony occurring in nodular form in a rock matrix, particularly in the chalk. Jasper is a red opaque massive form of chalcedony.

Natural hydrous silica may be divided into three well-defined structural groups (*see* **Minerals:** Other Silicates): opal-C (well-ordered γ-cristobalite), opal-CT (disordered α-cristobalite, α-tridymite), and opal-A (highly disordered, nearly amorphous). Opal may be colourless, milky white, yellow, red, green, blue, or black. In precious opal a play of delicate colours is seen, due not to chemical impurities but to optical diffraction from a randomly faulted close-packed structure of minute transparent silica spheres (around 0.25 μm). Colourless opal is a major constituent in hot-spring deposits and deep-sea precipitates.

Quartz is an extremely important in many industrial processes. At the simplest level it is raw material for the production of glasses, ceramics, and foundry moulds. Most quartz sands are a golden yellow due to the quartz particles being encased by a thin coating of iron oxides and/or hydroxides, and the iron would cause any glass to be dark green. In certain geological situations, however, the beach sands may be rolled to and fro by the tides for several million years, leading to the erosion of the iron coating and forming silver sands, which are much sought after for glass manufacture. For the more technical applications requiring an even higher degree of purity, colourless vein quartz is used, or nowadays reliance is placed on the manufacture of synthetic single crystals of quartz in an autoclave. One of the main technical aspects of quartz lies in its piezoelectric properties, whereby the crystal develops an electric polarization when mechanically stressed in an appropriate direction. In the construction of quartz oscillators, thin plates of quartz are cut in a particular direction specified to generate a desired resonant vibrational frequency.

Although most quartz is colourless or white, well-known coloured varieties include amethyst (violet), citrine (yellow to brown), and smoky quartz. Each of these contains either a substitutional or interstitial component other than SiO_2. For amethyst and citrine, iron is an important constituent (Fe 25–280 ppm). In smoky quartz the brownish or black colour is due to natural radiation acting upon Al^{3+}-bearing quartz.

Quartz is an essential constituent in many igneous, sedimentary, and metamorphic rocks, and also occurs as secondary material, often forming a cementing

medium in sediments. Quartz is also a common constituent of hydrothermal veins.

In granites, microgranites, tonalities, etc., quartz typically forms anhedral grains due to it being one of the final minerals to crystallize. Quartz phenocrysts in fine-grained rapidly cooled rhyolites, pitchstones, and quartz porphyries may show euhedral outlines, although in some of these rocks the quartz phenocrysts may suffer magmatic corrosion. In intermediate igneous rocks, the amounts of quartz are less than in those of granitic composition, and in basic rocks it usually amounts to less than 5% (though it is more abundant in some quartz dolerites).

Because of its chemical and physical resistance to erosion, quartz is an abundant mineral and becomes concentrated during sedimentary processes to give rise to sands and sandstones of various types. Secondary quartz is often deposited around pre-existing grains (of quartz or other minerals) and is a common cementing material in sediments.

Quartz is a common mineral in many metamorphic rocks, occurring in the metamorphosed equivalent of quartz-bearing sediments and igneous rocks. Although in low-grades of metamorphism quartz may survive unchanged, in the higher grades it undergoes recrystallization with concomitant increase in grain size. In addition, much is developed by the release of SiO_2 in reactions taking place during metamorphism. Intergrowths of quartz and other minerals are fairly common, particularly those of quartz and K-feldspar as in graphic or micrographic granite, and quartz and plagioclase as in myrmekite in association with K-feldspar.

The typical occurrence of tridymite is in acid volcanic rocks such as rhyolite, obsidian, trachyte, andesite, and dacite. In such rocks it is often found in cavities and may be associated with such minerals as sanidine, and less frequently augite or fayalite. Cristobalite is also a mineral of volcanic rocks, and may be found in (metastable) association with tridymite. It is often a late product of crystallization, sometimes replacing tridymite. The ability of cristobalite to occur as a metastable form, outside its equilibrium field, means that no definite conclusions can normally be drawn as to conditions at the time of its crystallization.

The SiO_2 polymorph coesite was first identified with quartz and fused silica glass in sheared porous sandstones at Meteor Crater, Arizona, and in granite and pumiceous tuff near the rim of the Ries crater, Bavaria, developed by the shock wave generated by the meteoritic impact. It is also being increasingly recognised as microscopic inclusions in such minerals as zircon, pyrope garnet, and other high-grade metamorphic minerals where it has often been partially inverted to quartz, resulting in radial cracking of the garnet around the inclusions. This coesite is considered to have crystallized in nearly static pressure conditions at $\geq$28 kbar, followed by fairly rapid exhumation.

See Also

Igneous Rocks: Granite. **Minerals:** Definition and Classification; Other Silicates. **Sedimentary Rocks:** Mineralogy and Classification; Chert; Sandstones, Diagenesis and Porosity Evolution. **Tektites**. **Ultra High Pressure Metamorphism**.

Further Reading

Audetát A and Gunther D (1999) Mobility and H_2O loss from fluid inclusions in natural quartz crystals. *Contributions to Mineral Petrology* 137: 1–14.

Chopin C (1984) Coesite and pure pyrope in high-grade blueschists of the Western Alps: a first record and some consequences. *Contributions to Mineral Petrology* 86: 107–118.

Heaney PJ, Prewitt CT, and Gibbs GV (eds.) (1994) Silica: Physical Behavior, Geochemistry and Materials Applications Reviewal Mineralogy 29: 606.

Klein C and Hurlbut CS (1993) *Manual of Mineralogy,* 21st edn. New York: John Wiley.

Liou JG and Zhang RY (1996) Occurrence of intergranular coesite in ultrahigh-*P* rocks from the Sulu region, eastern China: implications for lack of fluid during exhumation. *American Mineralogy* 81: 1217–1221.

O'Brien PJ, Zotov N, Law R, Khan MA, and Jan MQ (2001) Coesite in Himalayan eclogite and implications for models of India–Asia collision. *Geology* 29: 435–438.

Sulphates

G Cressey, The Natural History Museum, London, UK

Introduction

Gypsum ($CaSO_4 \cdot 2H_2O$), the most common of the sulphate minerals, is also known as alabaster (a fine-grained massive form), satin spar (a fibrous variety of gypsum), or selenite (colourless transparent gypsum crystals). Gypsum is often found in considerable thicknesses within evaporite sequences and in association with limestones and shales. Large deposits of alabaster gypsum are commonly observed at the sole of many large-scale tectonic dislocations and thrust faults. Gypsum is of great importance as a raw material in the production of cement and plaster for the building industry.

Crystal Structure of Gypsum

In the crystal structure of gypsum, calcium ions are coordinated by six oxygens from sulphate (SO_4) tetrahedral groups, and by two oxygens from water molecules (H_2O). Two layers of SO_4 tetrahedra are bound together by calcium ions forming a double-layer sheet. Water molecules are located at each side of these double-layer sheets. The water molecules are oriented such that one of the two hydrogens of each water molecule points almost perpendicular to the sheets; this facilitates weak hydrogen bonding to the next layer in the structure. These hydrogen bonds point approximately along the *y*-axis of the crystal, forming a layer of weak hydrogen bonding perpendicular to the *y*-axis, and explain the easy (010) cleavage of gypsum. The proton (H^+) positions in gypsum have been determined in deuterated gypsum by neutron diffraction, as weakly scattering proton positions cannot be determined accurately by X-ray diffraction methods (Figure 1).

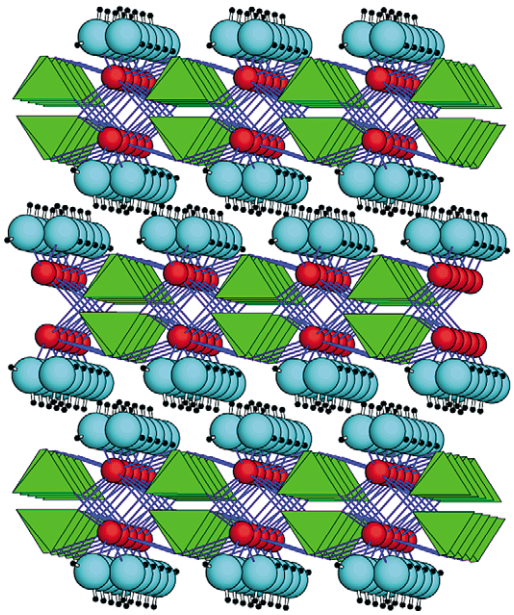


Figure 1 The structure of gypsum (view looking almost down the *z*-axis): SO_4 tetrahedral units (green) and calcium ions (red) form double-layer sheets perpendicular to the *y*-axis. At each side of these sheets, water molecules (with oxygen in blue and hydrogen shown as small black dots) form weak hydrogen bonds to the next layer in the structure.

Physical Properties of Gypsum

The weak hydrogen bonding in the structure of gypsum, occurring in layers perpendicular to the *y*-axis, accounts for the perfect (010) cleavage and is also responsible for the anisotropic thermal expansion behaviour of gypsum. The *b* dimension of the unit cell increases the most with temperature, and this is principally because the hydrogen bonds oriented across the sheets easily lengthen with increasing temperature. The static compressibility of gypsum is also highly anisotropic and, as is the case with the thermal expansion, is related to the hydrogen-bonding orientation across the sheet structure; these hydrogen bonds more easily shorten with pressure and therefore the *b* lattice parameter has the greatest compressibility.

Upon heating to about 65°C, gypsum dehydrates to form the hemihydrate phase, bassanite ($CaSO_4 \cdot 0.5H_2O$), with its *z*-axis parallel to that of the gypsum; this is known as plaster of Paris. Further heating dehydrates the sulphate completely, forming γ-$CaSO_4$ (a polymorph of anhydrite, $CaSO_4$) at about 95°C. γ-$CaSO_4$ is also known as 'soluble anhydrite', as unlike anhydrite it can be re-hydrated. Anhydrite has a different structure and cannot be easily hydrated. Commercial plaster of Paris consists mainly of calcium sulphate hemihydrate (bassanite) and is made by heating gypsum to about 200°C. Heating gypsum at higher temperatures produces anhydrite and this change is irreversible. At the stage where bassanite and/or γ-$CaSO_4$ are produced, re-hydration to gypsum ($CaSO_4 \cdot 2H_2O$) easily occurs; this is the process of the setting of plaster of Paris to form an interlocking mass of gypsum crystals.

Occurrence of Gypsum

Gypsum occurs mainly as sedimentary deposits associated with limestones and shales and in evaporite deposits. Gypsum can form by the direct evaporation of brine, by hydration of anhydrite and by oxidation of sulphides. Gypsum is also one of the products of the reaction of acid sulphate solutions with carbonate rocks. Seawater contains about 3.5% dissolved material, of which 80% is sodium chloride and about 4% is dissolved calcium sulphate. The experimental evaporation of seawater results in the crystal deposition sequence: calcium carbonate – calcium sulphate – sodium chloride – sulphates/chlorides of magnesium – sodium bromide/potassium chloride.

Gypsum also occurs in soils either as disseminated crystals or in horizons beneath calcrete layers. Percolating waters, which in dry seasons are drawn to the surface by capillary action, can evaporate and crystallize gypsum, sometimes in the form described as 'desert roses'. The calcium sulphate formed in evaporates is sometimes gypsum, sometimes anhydrite, and often both minerals occur together. Some geological evidence suggests that the original material of certain gypsum beds was anhydrite; this is supported by the fact that some gypsum beds grade into anhydrite at depth. Along with jarosite and goethite, gypsum is one of the products formed by the oxidation of iron sulphide (pyrite). Gypsum is also found in volcanic regions along with native sulphur where it forms by the reaction of sulphur-bearing fluids and gases with calcium-bearing rocks. Natural bassanite is rare, and nearly always forms as a product of alteration of gypsum. It has been found along with gypsum in volcanic craters, and in some soils and peat deposits.

Other Sulphates

The most abundant barium mineral in the Earth's crust is barite, $BaSO_4$, and this is also the least soluble sulphate mineral. It commonly occurs in hydrothermal metalliferous veins, but also occurs in cementations and nodules in sedimentary rocks. Because of its high density ($4.5\,g\,cm^{-3}$) barite is widely used as a component of drilling muds in the petroleum industry, or in other dense fluid media, and also as a filler or extender in the manufacture of paper, plaster, rubber and plastics. There is a complete solid solution series between $BaSO_4$ (barite) and $SrSO_4$ (celestine), although natural compositions are usually near to end-member compositions, reflecting either the Ba-rich or Sr-rich geochemical environments. $PbSO_4$ (anglesite) is isostructural with $BaSO_4$ (barite), so a complete solid solution between these two end members would also be expected, and some natural samples occur that are intermediate in composition. At room temperature, about 6% Ca^{2+} can be accommodated in the $BaSO_4$ structure. About 8% Ba^{2+} replacement of Ca^{2+} is the limit of solid solution tolerated by the $CaSO_4$ (anhydrite) structure.

The oxidation of pyrite (FeS_2), especially when mediated by the action of bacteria, frequently results in the formation of powdery assemblages of iron sulphates. This alteration commonly occurs during the storage of pyrite-bearing materials under humid conditions. Sulphates produced in this way include melanterite ($FeSO_4 \cdot 7H_2O$), rozenite ($FeSO_4 \cdot 4H_2O$), szomolnokite ($FeSO_4 \cdot H_2O$) and rhomboclase ($Fe^{3+}(SO_4)_2\,(H_3O^+) \cdot 3H_2O$).

Several hundred compounds that contain essential sulphate (SO_4) groups are known to occur in nature; these include sulphates, sulphate hydrates, sulphate hydroxy hydrates, sulphate/chlorides, and sulphate/arsenate/phosphate/vanadates of alkali- and transition-metals.

See Also

Minerals: Definition and Classification; Sulphides. **Sedimentary Rocks:** Evaporites.

Further Reading

Alpers CN, Jambor JL, and Nordstrom DK (eds.) (2000) *Sulfate Minerals–Crystallography, Geochemistry and Environmental Significance*. Reviews in Mineralogy and Geochemistry, vol. 40. Mineralogical Society of America.

Bragg WL (1937) *The Atomic Structure of Minerals*. New York: Cornell University Press, and London: Oxford University Press.

Chang LLY, Howie RA, and Zussman J (1996) *Rock Forming Minerals*, vol. 5B, *Sulphates, Carbonates, Phosphates and Halides*. London: Longman.

Jerz JK and Rimstidt JD (2003) Efflorescent iron sulfate minerals: paragenesis, relative stability, and environmental impact. *American Mineralogist* 88: 1919–1932.

Pedersen BF and Semmingsen D (1982) Neutron diffraction refinement of the structure of gypsum, $CaSO_4 \cdot 2H_2O$. *Acta Crystallographica* B38: 1074–1077.

Wooster WA (1936) On the crystal structure of gypsum $CaSO_4 \cdot 2H_2O$. *Zeitschrift fur Kristallographie* 94: 375–396.

Sulphides

D J Vaughan, University of Manchester, Manchester, UK

Introduction

The sulphide minerals are the major source of world supplies of a very wide range of metals and are the most important group of ore minerals. In addition to their concentrated occurrence in ore deposits and areas of mineralization, a limited number of sulphide minerals are found as accessory minerals in rocks. However, only pyrite, pyrrhotite, galena, sphalerite, chalcopyrite, and possibly the binary copper sulphides of the chalcocite family could be classed as rock-forming minerals. Pyrite (FeS_2) is by far the most abundant sulphide mineral. The very fine particle iron sulphides found in reducing environments beneath the surfaces of Recent sediments and soils, although transient species, are also volumetrically important. Formerly referred to as amorphous iron sulphide, these phases are now known to be largely composed of fine-particle mackinawite (tetragonal FeS) and, in some cases, the sulphospinel mineral greigite (Fe_3S_4). Both phases are metastable in relation to pyrite and pyrrhotite. As well as being important sources of metals, sulphide minerals are potential sources of pollution. In particular, the release of sulphur through the weathering of sulphides in natural rocks or in mine wastes generates sulphuric acid, resulting in acid rock drainage or acid mine drainage, and a significant proportion of the sulphurous fumes generated by combustion of coal (or by the smelting of sulphide ores) originates from sulphide minerals; these fumes react to form acid rain. A more positive role may have been played by sulphide minerals in the emergence of life on Earth; although speculative, such theories have been developed following the discoveries of ocean-floor hydrothermal systems generating large volumes of sulphide minerals and associated with novel life forms and ecosystems.

The mineralogy, mineral chemistry, and geological occurrences of sulphide minerals have been well researched, particularly because of their economic and environmental importance. The crystal structures, chemical compositions, physical properties, thermochemistry, and phase relations of the major sulphides were well established by the latter part of the twentieth century, as were the macroscopic and microscopic characteristics of sulphide minerals in natural occurrences. Recent decades have seen major advances in the understanding of the chemical bonding in sulphide minerals and in the surface chemistry of sulphides. An understanding of the genesis of sulphide-ore deposits has developed through systematic geological, petrographic, compositional, and isotopic studies of ores and host rocks.

In this article, the compositions, crystal structures, physical properties, and certain aspects of mineral chemistry and geological occurrence of the major sulphide minerals will be reviewed, with suggestions given for further reading.

Compositions and Structures

Several of the common sulphide minerals were among the first minerals to be studied by X-ray crystallography, and since that time the structures of nearly all the mineralogically significant sulphides have been determined. It is possible to categorize the mineral sulphides into a series of groups based on major structure types, or having key structural features in common, as shown in Table 1. In many cases, these structures are the classic structures of crystalline solids, such as the rock-salt structure of the galena group (Figure 1A), the sphalerite and wurtzite forms of zinc sulphide (Figures 1B and 1C), or the nickel arsenide structure (Figure 1E). The disulphides are characterized by the presence of dianion (S–S, S–As, As–As, etc.) units in the structure: as well as the pyrite structure, in which FeS_6 octahedral units share corners along the *c*-axis direction, there is the marcasite form of FeS_2, in which octahedra share edges to form chains along the *c*-axis (Figure 1D). The structures of loellingite ($FeAs_2$) and arsenopyrite (FeAsS) are variants of the marcasite structure that have shorter and alternate long and short, respectively, metal–metal distances across the shared octahedral edge (see Figure 1D). A few sulphides such as molybdenite, covellite (Figure 1F), and mackinawite have layer structures, and a small number exhibit structures that are best characterized as containing rings or chains of linked atoms (e.g. realgar, AsS). A diverse group of sulphides, the metal-excess group, is composed of an unusual set of structures that is well illustrated by the important example of the mineral pentlandite ($(Ni,Fe)_9S_8$; Figure 1G).

As can be seen from Table 1, in many of these groups a number of minerals share the actual structure type, but there are also, commonly, other minerals that have structures based on these 'parent' structures and that can be thought of as being derived from them. The relationship between a derivative structure and the parent structure may involve

Table 1 Sulphide structural groups

1. The disulphide group					
Pyrite structure	*Marcasite structure*	*Arsenopyrite structure*	*Loellingite structure*	*Derived by As/S ordered substitution*	
FeS_2 pyrite	FeS_2 marcasite	FeAsS arsenopyrite	$CoAs_2$ safflorite	(Co,Fe)AsS cobaltite	
CoS_2 cattierite		FeSbS gudmundite	$FeAs_2$ loellingite	(Ni,Co,Fe)AsS gersdorffite (l)	
			$NiAs_2$ rammelsbergite		
2. The galena group					
PbS galena					
α-MnS alabandite					
3. The sphalerite group					
Sphalerite structure	*Derived by ordered substitution*	*Stuffed derivatives*			
β-ZnS sphalerite	$CuFeS_2$ chalcopyrite	$Cu_9Fe_8S_{16}$ talnakhite			
CdS hawleyite	Cu_2FeSnS_4 stannite	$Cu_9Fe_9S_{16}$ mooihoekite			
Hg(S,Se) metacinnabar	Cu_2ZnSnS_4 kesterite	$Cu_4Fe_5S_8$ haycockite			
4. The wurtzite group					
Wurtzite structure	*Composite structure derivatives*	?*Further derivatives*	*Derived by ordered substitution*		
α-ZnS wurtzite	$CuFe_2S_3$ cubanite	$Cu_2Fe_2SnS_6$ hexastannite	Cu_3AsS_4 enargite		
CdS greenockite	?$AgFe_2S_3$ argentopyrite				
5. The nickel arsenide group					
NiAs structure	*Distorted derivatives*	*Ordered omission derivatives*			
NiAs niccolite	FeS troilite	Fe_7S_8 monoclinic pyrrhotite			
NiSb breithauptite	CoAs modderite	Fe_9S_{10}, $Fe_{11}S_{12}$ hexagonal pyrrhotite, etc?			
6. The thiospinel group					
Co_3S_4 linnaeite					
$FeNi_2S_4$ violarite					
$CuCo_2S_4$ carrollite					
7. The layer sulphides group					
Molybdenite structure	*Tetragonal PbO structure*	*Covellite structure*			
MoS_2 molybdenite	$(Fe,Co,Ni,Cr,Cu)_{1+x}S$ mackinawite	CuS covellite			
WS_2 tungstenite		$\sim Cu_3FeS_4$ idaite			
8. Metal excess group					
Pentlandite structure	*Argentite structure*	*Chalcocite structure*	*Digenite structure*	*Derived by ordered substitution*	*Nickel sulphide structures*
$(Ni,Fe)_9S_8$ pentlandite	Ag_2S argentite	Cu_2S chalcocite	Cu_9S_5 digenite	Cu_7S_4 anilite	NiS millerite
		?derivative			Ni_3S_2 heazlewoodite
		$Cu_{1.96}S$ djurleite			
9. Ring or chain structure group					
Stibnite structure	*Realgar structure*	*Cinnabar structure*			
Sb_2S_3 stibnite	As_4S_4 realgar	HgS cinnabar			
Bi_2S_3 bismuthinite					

(l) = one polymorph.

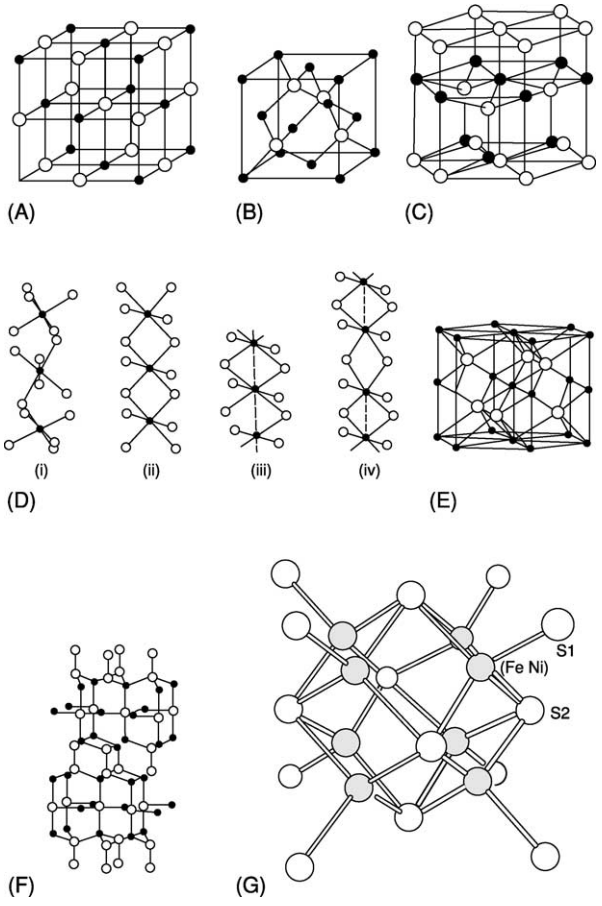


Figure 1 Crystal structures of the major sulphides: (A) galena (PbS) structure; (B) sphalerite (ZnS) structure; (C) wurtzite (ZnS) structure; (D) pyrite structure and the linkage of metal–sulphur octahedra along the *c*-axis direction in (i) pyrite (FeS_2), (ii) marcasite (FeS_2), (iii) loellingite ($FeAs_2$), and (iv) arsenopyrite (FeAsS); (E) niccolite (NiAs) structure; (F) covellite (CuS) structure; and (G) cube cluster of tetrahedrally coordinated metals in the pentlandite ($(Ni,Fe)_9S_8$) structure. In each case the metals are shown as smaller or shaded spheres. Adapted from Institute of Materials, Minerals and Mining, Sulphite Deposits – their origin and Processing, ed. P. Gray, 1990.

- ordered substitution (the structure of chalcopyrite ($CuFeS_2$) is derived from that of sphalerite (ZnS) by alternate replacement of zinc atoms with copper and iron, resulting in an enlarged (tetragonal) unit cell; the structure of stannite (Cu_2FeSnS_4) results from further ordered substitution of half of the iron atoms in chalcopyrite by tin (Figure 2A));
- a stuffed derivative (talnakhite ($Cu_9Fe_8S_{16}$) is derived from chalcopyrite by the occupation of additional normally empty cavities in the structure (Figure 2B));
- ordered omission (monoclinic pyrrhotite (Fe_7S_8) is derived from the nickel arsenide structured troilite by removal of iron atoms, leaving holes (vacancies) that are ordered (Figure 2C)); or
- distortion (the troilite form of FeS is simply a distortion of the parent nickel arsenide structure form (Figure 2C)).

In some cases, the relationships involved are more complex, as, for example, in certain of the sulphosalts (minerals with a general formula $A_mT_nX_p$ in which common elements are *A*: Ag, Cu, and Pb; *T*: As, Sb, and Bi; *X*: S, and which contain pyramidal TS_3 groups in the structure). Here, the resulting structure is composite and made up of slabs or units of a parent structure (or structures) arranged in an ordered fashion.

The compositional variations of the major sulphide minerals are reasonably well characterized, both as a result of numerous analyses of natural samples from a wide variety of deposits and as a result of systematic laboratory investigations of the phase equilibria. Many metal sulphides show evidence that the elements comprising them are not combined in simple whole-number ratios, i.e. they exhibit nonstoichiometry. In certain cases, the deviation from a simple ratio is considerable. For example, the pyrrhotites are sometimes given the general formula $Fe_{1-x}S$ where $0 < x < 0.125$ and the varying compositions correspond to varying concentrations of vacancies in sites that would otherwise be occupied by iron atoms. However, in systems such as these, ordering of the vacancies occurs at low temperatures, and the result may be a series of stoichiometric phases of slightly different compositions. Although Fe_7S_8 has a (monoclinic) superstructure resulting from vacancy ordering, the situation in the intermediate or hexagonal pyrrhotites is more complex. Some of these pyrrhotites may represent ordered phases with clearly defined compositions (Fe_9S_{10}, $Fe_{11}S_{12}$, etc.), but more complex and partial ordering in these systems may occur. One problem is certainly that free-energy differences between a series of phases resulting from vacancy ordering are very small, making a successful investigation of the relationships between synthetic products very difficult.

Other examples of nonstoichiometry involve relatively small deviations from the simple ratio. For example, galena (PbS) exhibits a range of nonstoichiometry of 0.1 atomic %. Galena is apparently stable over a wide range of values of aS_2 (activity of sulphur), and at high aS_2 it has lead vacancies, whereas at low aS_2 there are sulphur vacancies.

Electrical, Magnetic, and Optical Properties

The metal sulphides also show a tremendous range of electrical and magnetic behaviours. As Table 2

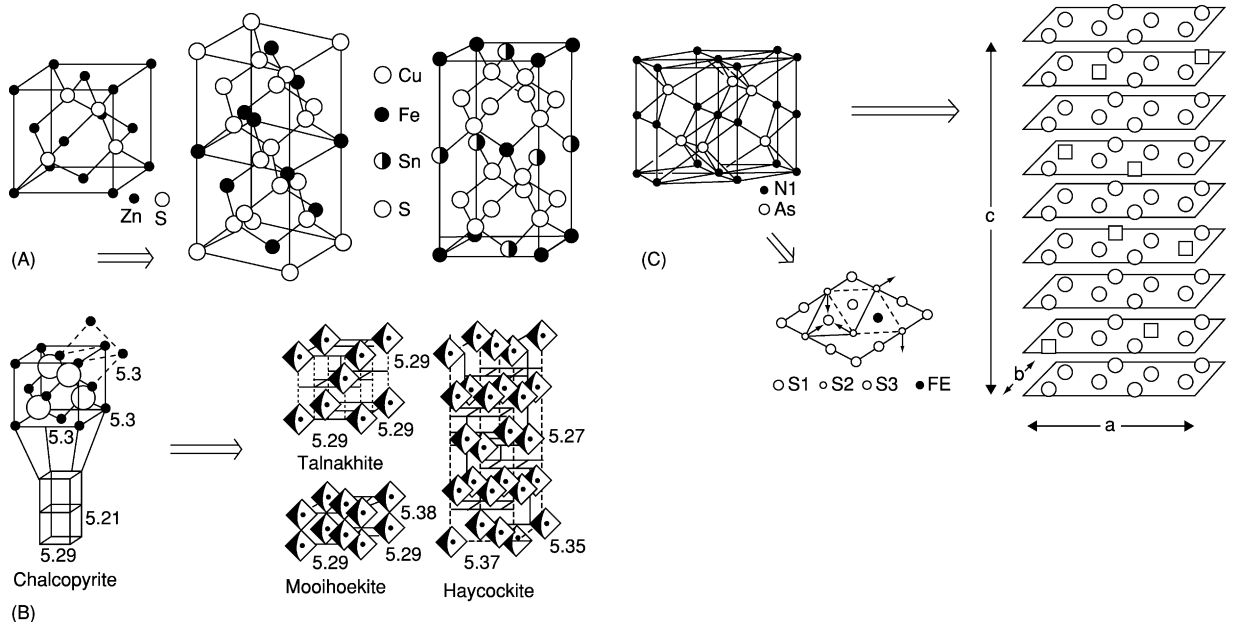


Figure 2 Parent and derivative crystal structures in the sulphide minerals. (A) The sphalerite (ZnS) structure with the chalcopyrite ($CuFeS_2$) and stannite (Cu_2FeSnS_4) structures. (B) The sphalerite and chalcopyrite unit cells with octahedra of metals outlined, within which may be an additional metal ion in the minerals talnakhite ($Cu_9Fe_8S_{16}$), mooihoekite ($Cu_9Fe_9S_{16}$), and haycockite ($Cu_4Fe_5S_8$); the arrangement of additional occupied metal sites is as shown (also shown are the dimensions of the parent sphalerite cell in (A)). (C) The niccolite unit cell of high-temperature FeS, which has vacancies in place of iron atoms in monoclinic pyrrhotite (Fe_7S_8) that are ordered as shown, with vacancies represented by squares (and only iron-atom layers shown); also shown in a projection onto the basal plane are the distortions that occur in the troilite modification of FeS. Adapted from Institute of Materials, Minerals and Mining, Sulphite Deposits – their origin and Processing, ed. P. Gray, 1990.

Table 2 Magnetic and electrical properties of some major sulphide minerals

Mineral species	*Magnetic properties*	*Electrical properties*
Sphalerite (ZnS)	diamagnetic	insulator (E_g~3.7 eV)
'Iron sphalerite' (Zn,Fe)S	paramagnetic	semiconductor (E_g~0.5 eV for *ca.* 12 at.% Fe)
Galena (PbS)	diamagnetic	semiconductor (n- and p-type, E_g~0.41 eV)
Pyrite (FeS_2)	diamagnetic	semiconductor (n- and p-type, E_g~0.9 eV)
Cattierite (CoS_2)	ferromagnetic ($T_c = 110$ K)	metallic
Chalcopyrite ($CuFeS_2$)	antiferromagnetic ($T_N = 823$ K)	semiconductor (n-type, E_g~0.5 eV)
Covellite (CuS)	diamagnetic(?)	metallic
(monoclinic) Pyrrhotite (Fe_7S_8)	ferrimagnetic ($T_c = 573$ K)	metallic
Carrollite ($CuCo_2S_4$)	Pauli paramagnetic	metallic
Pentlandite $(Ni,Fe)_9S_8$	Pauli paramagnetic	metallic

E_g, band or 'energy' gap; T_c, Curie temperature; T_N, Nèel temperature.

indicates, whereas such non-transition-metal sulphides as sphalerite and galena are diamagnetic; diamagnetism is also exhibited by pyrite and marcasite (the other FeS_2 polymorph). Substitution of iron for zinc in sphalerite leads to paramagnetic behaviour, and many transition-metal sulphides show various forms of magnetic ordering at lower temperatures, including antiferromagnetism (e.g. chalcopyrite), ferromagnetism (e.g. cattierite), and ferrimagnetism (e.g. monoclinic pyrrhotite, although hexagonal pyrrhotites are antiferromagnetic). Other transition-metal sulphides that are metallic conductors exhibit the weak temperature-independent paramagnetism known as Pauli paramagnetism. Metallic conductivity occurs in many sulphides of diverse magnetic character, and, in addition, numerous sulphides are semiconductors and a smaller number (e.g. pure sphalerite) are insulators. Among semiconducting sulphides, both intrinsic and impurity (or extrinsic) conduction mechanisms occur, and conduction both via electrons (n-type) and via holes (p-type) is common. Such variations are often a consequence of the presence of very minor impurities or of

slight nonstoichiometry (e.g. galena (PbS) shows p-type conductivity in lead-deficient samples and n-type conductivity in sulphur-deficient samples).

Given that the great majority of sulphide minerals are opaque, the optical properties of interest are those observed in reflected light from polished sections. The qualitative optical properties commonly serve for routine identification, but definitive characterization can be greatly aided by measuring reflectance at set wavelengths, or a whole series of wavelengths, across the visible-light region.

Sulphide Mineral Stability

Much work has been done to establish the stability relations of sulphide minerals in terms of variables that include temperature, pressure, composition, and the activities of various components.

The Fe–S system is the most important binary sulphide system because the iron sulphides are, by far, the dominant sulphide minerals in terms of geological abundance and variety of geological occurrences. Furthermore, the Fe–S system is an integral part of other important more complex systems such as the Fe–Zn–S, Cu–Fe–S, and Fe–Ni–S systems. The phase relations in the condensed Fe–S system above 400°C (Figure 3A) have been the subject of numerous studies and, in contrast to the low-temperature relationships, are well understood. The central part of the system is dominated by the large high-temperature pyrrhotite field of solid solution from stoichiometric FeS towards more sulphur-rich compositions. This high-temperature form with a hexagonal NiAs-type structure accommodates solid solution by random vacancies at the iron sites within the lattice. Hence, the compositions of high-temperature pyrrhotites, except for pure FeS, are best given as $Fe_{1-x}S$, where x is 0 to ~0.14. The maximum thermal stability of the pyrrhotite solid solution at pressures below about 1 atmosphere is 1192°C, above which it melts congruently. Pyrite has, at low pressures, a maximum thermal stability of 742 ± 1°C. The upper thermal stability of pyrite rises by approximately 14°C per kilobar of confining pressure. In spite of numerous studies, the phase relations in the Fe–S system at temperatures below 350°C are incompletely understood. A temperature–composition diagram for the system between FeS and FeS_2 from 0°C to 350°C is shown in Figure 3B. Phase relations in the compositional regions Fe–FeS and FeS_2–S remain straightforward, essentially as shown in Figure 3A, but the central part of the system is exceedingly complex. This complexity is caused by the crystal chemistry of the pyrrhotites, as discussed above, where iron atoms can be omitted, leaving holes or vacancies. At low temperatures, ordering of vacancies occurs, resulting in the development of various superstructures. The best-known of these superstructures is that of monoclinic pyrrhotite (Fe_7S_8). In this case, the vacancies occur in alternate layers of iron atoms parallel to the basal plane and in alternate rows in those layers (Figure 2C). This is known as the 4C structure, because the superstructure has a unit cell that is four times the c dimension of the parent nickel arsenide-type cell.

Although the structure of monoclinic pyrrhotite is well established, the structures of those compositions lying between FeS and Fe_7S_8 remain uncertain. Numerous superstructure types have been reported for these intermediate pyrrhotites, including examples with nonintegral multiples of the parent-cell c and a dimensions (so-called NC and NA types). Many reported superstructures have not been observed in natural samples, where there is evidence of compositions clustering around $Fe_{11}S_{12}$, $Fe_{10}S_{11}$, and Fe_9S_{10}, as well as Fe_7S_8. These compositions also correspond to superstructures with high degrees of order, the 5C (Fe_9S_{10}) and 6C ($Fe_{11}S_{12}$) superstructures. The thermodynamic stability of the various pyrrhotites, particularly the intermediate pyrrhotites, is problematic, with kinetic controls exerting the major influence over the compositions and structure types observed. The highly ordered monoclinic pyrrhotite (Fe_7S_8) and troilite (FeS, a 2C superstructure produced by distortion of the nickel arsenide high-temperature form) may be stable phases in the system, but reported or measured free-energy minima suggest that only pyrite and troilite are truly stable phases.

Much sulphide is formed in sedimentary or hydrothermal environments, so phase equilibria in aqueous systems are of particular interest. The stability of pyrite, for example, has been considered as a function of many variables. Perhaps the best-known relationships are those at 25°C, for which the stability of pyrite in aqueous solution has been examined as a function of the changing activities of two or more components in the system when the others are held constant. Thus, for the system Fe–S–O–H, the activity of hydrogen ions (pH) has been plotted against the 'activity of electrons' (Eh), and either one or both of these variables have been plotted against the activity of sulphur (expressed as pS^{2-} or pS_2). Figure 4 shows Eh–pH plots illustrating the stability of pyrite and other sulphides and oxides of iron in water at 25°C and 1 atmosphere total pressure. Figure 4A shows relationships at a total dissolved sulphur activity of 10^{-1} and illustrates the large stability field of pyrite under reducing conditions over a wide pH range. By contrast, the stability field of pyrrhotite is extremely small. Under more oxidizing conditions, haematite is the dominant iron mineral, with magnetite having

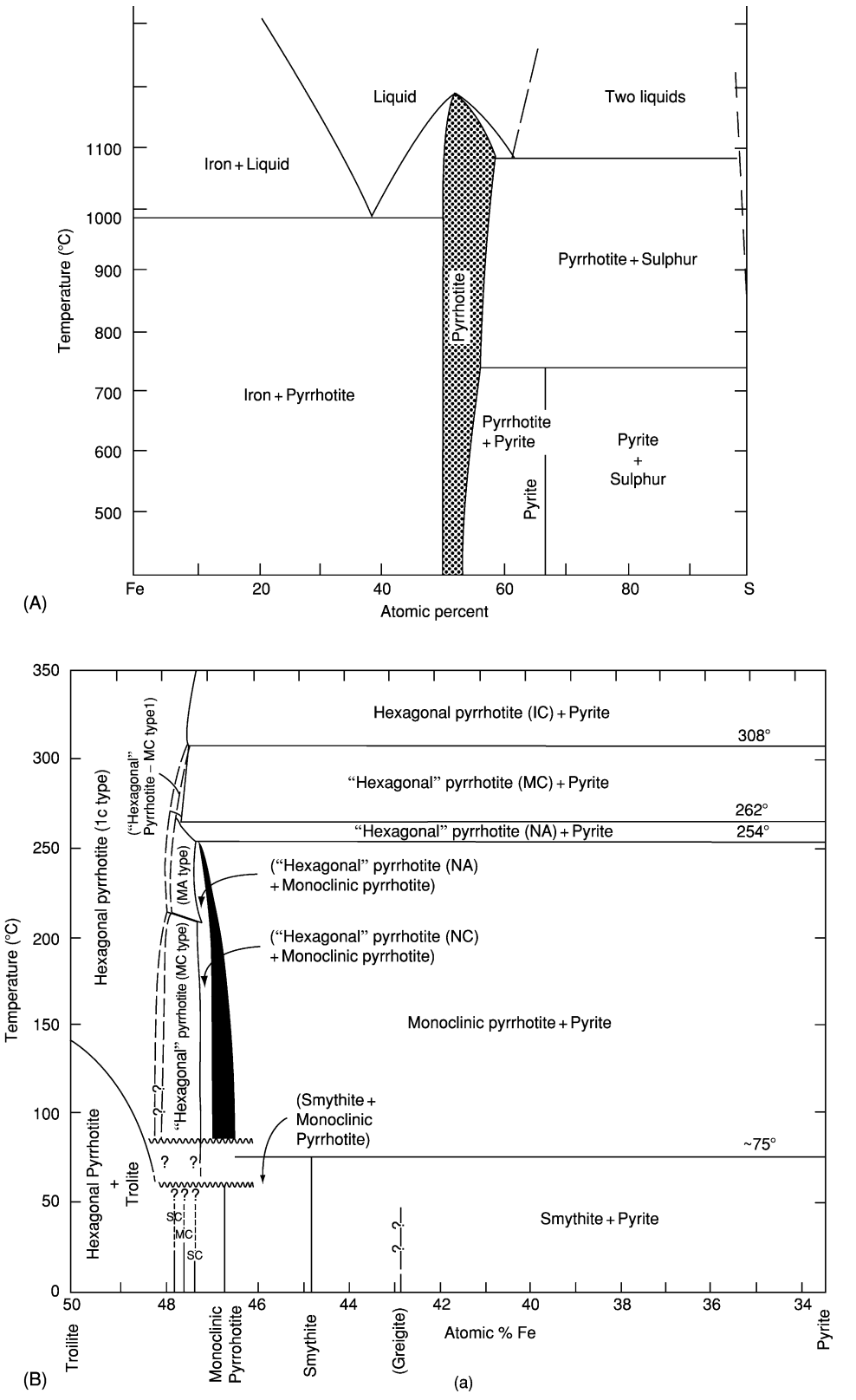


Figure 3 The iron–sulphur system: (A) phase relations in the Fe–S system above 400°C at low pressure; and (B) phase relations in the central portion of the Fe–S system below 350°C. Reproduced with permission from Geochemistry of Hydrothermal Ore Deposits, ed. H. L. Barnes, 1997, John Wiley and Sons.

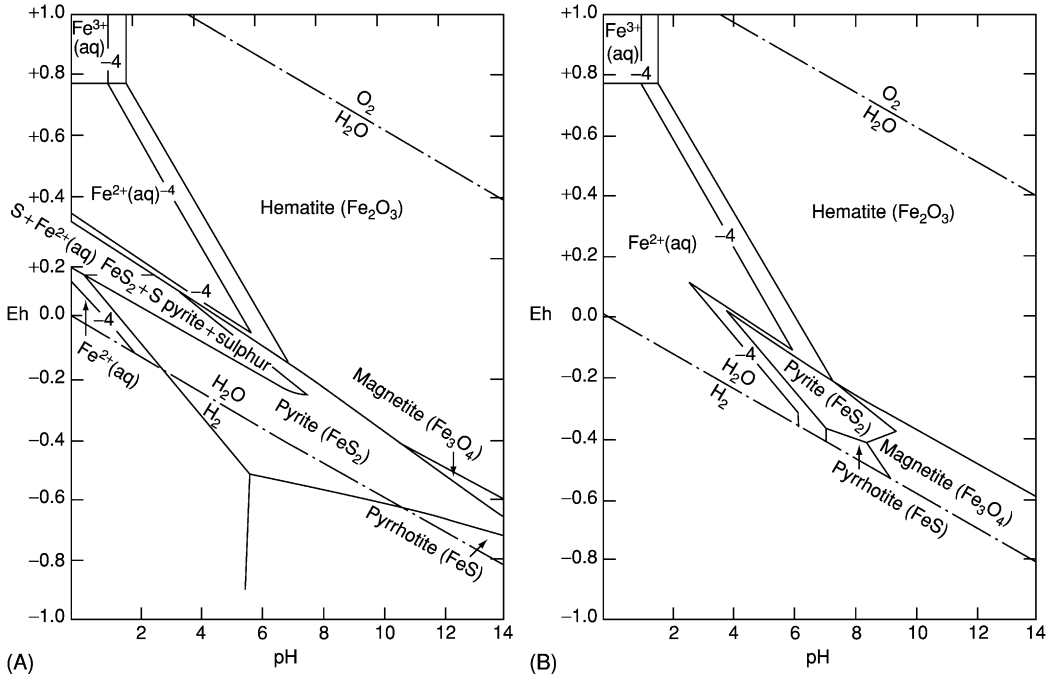


Figure 4 Stability relations of iron sulphides and oxides in water at 25°C and 1 atmosphere total pressure (A) at a total activity of dissolved sulphur of 10^{-1} and (B) at a total activity of dissolved sulphur of 10^{-6}.

only a very small stability field. The effect of lowering the total sulphur activity to 10^{-6} is shown in Figure 4B: the pyrite field is now much smaller and the magnetite field has increased in size. Notable from examination of Figure 4 is the very large stability field of pyrite relative to pyrrhotite in sediments. It can also be seen that quite low sulphide activities can stabilize pyrite over a considerable range of Eh conditions, and that iron oxides or carbonate will form only when the sulphide activity of the environment is very small.

The problem of metastability is implicit in the study of iron sulphides. For example, certain iron sulphides that are common in nature, such as marcasite (FeS_2) and mackinawite ($Fe_{1+x}S$), are not included in the phase diagrams of Figure 3. These sulphides cannot be synthesized by a direct reaction between iron and sulphur in dry systems but only by precipitation from solution. In the natural aqueous systems where such iron sulphides form, reduced sulphur is readily available from the bacterial reduction of sulphate. Here mackinawite is the dominant precursor phase, with the ultimate stable product being pyrite. Details of the reaction pathways involving mackinawite and other less common precursors are still being established, but Figure 5 makes an attempt to summarize known and more speculative phase transformations.

When interpreting the mineral assemblages and textures in order to understand the genesis of sulphide

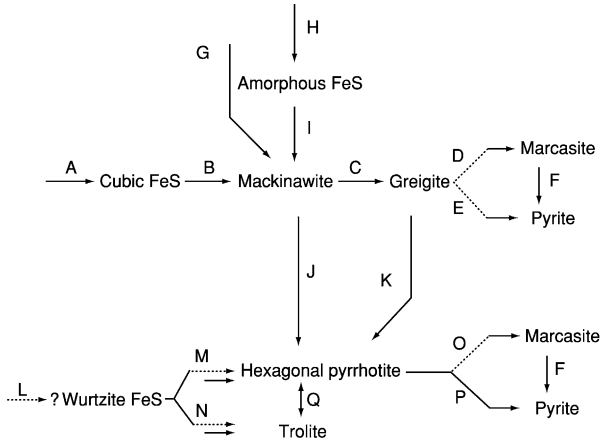


Figure 5 Summary of known and speculative phase transformations and reactions in the low-temperature iron–sulphur system (redrawn after Lennie and Vaughan, *Geochem. soc. spec pubh. No. 5*, p.128, 1996). The letters on the diagram are as follows: A, formation from Fe^{2+} or Fe and HS^- or H_2S at a pH of 4–5 ($T < 523$ K); E, inferred sulphidation of greigite at a pH of more than 5 ($T < 523$ K); F, solid state transformation; G, sulphidation of Fe by H_2S at a pH of 3–10 ($T < 523$ K); I, ageing of amorphous precipitate; J, solid state transformation, kinetics not established; L, precipitation from Fe^{2+} by HS at a pH of 4–6 (wide range of temperatures to above 573 K); M,N, reaction of Fe^{2+} and HS^- or H_2S at a pH of 4–5 (or speculative transformation of proposed wurtzite-FeS structure) to form either (M) hexagonal pyrrhotite ($T > 413$ K) or (N) troilite ($T < 413$ K); O, marcasite after pyrrhotite; P, pyrite formed by sulphidation of pyrrhotite (this reaction is rapid above 573 K); Q, troilite–hexagonal pyrrhotite reversible phase transition. Reproduced with permission from Geochemistry of Hydrothermal Ore Deposits, ed. H. L. Barnes, 1997, John Wiley and Sons.

ores, temperature–composition relations in the subsolidus region are particularly useful. This is because many sulphide minerals undergo changes in the solid state down to relatively low temperatures. These changes give rise to a wealth of intergrowth textures produced by exsolution or unmixing, as exemplified by the Ni–Fe–S system (Figure 6). Such changes occur because, at elevated temperatures in many sulphide systems, there are extensive fields of solid solution that shrink with falling temperature. In the case of the Ni–Fe–S system, it is the breakdown of the so-called monosulphide solid solution and the segregation of pentlandite according to the reaction

$$(Fe, Ni)S = Fe_{1-x}S + (Fe, Ni)_9S_8$$

monosulphide solid solution = pyrrhotite + pentlandite

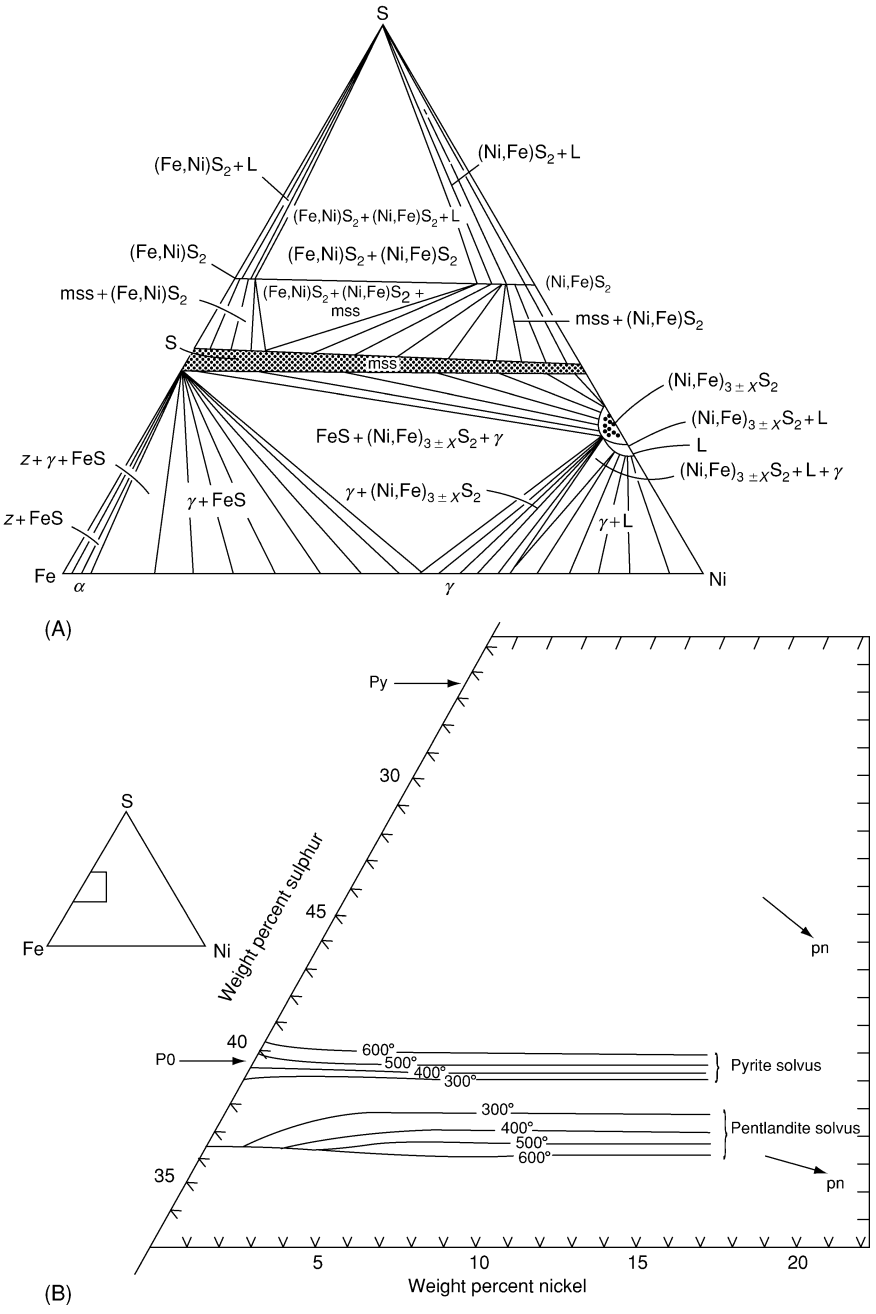


Figure 6 (A) The condensed phase relations in the Fe–Ni–S system at 650°C. Note the absence of pentlandite and the area of monosulphide solid solution (*mss*) that spans the system from $Fe_{1-x}S$ to $Ni_{1-x}S$. (B) A portion of the Fe–Ni–S system, showing the compositional limits of the monosulphide solid solution in (A) at 600°C, 500°C, 400°C, and 300°C. Pentlandite flames exsolve from the *mss* as the sulphur-poor boundary retreats during cooling. Reproduced with permission from Geochemistry of Hydrothermal Ore Deposits, ed. H. L. Barnes, 1997, John Wiley and Sons.

that is central to understanding the assemblages and textures in the sulphide nickel ores. Such processes may lead to completely separate grains of the two phases being formed on exsolution, or the two phases may be intergrown as laths, blebs, etc., often with a clearly defined crystallographic relationship. The crystal structures of the two phases will partly dictate the kind of texture that forms, but other crucial factors are the monosulphide solid solution bulk starting composition and the cooling history. The majority of binary and ternary and many quaternary sulphide systems have now been studied as regards temperature–composition relations, yielding evidence of the maximum stability temperatures of phases, solid solution limits, and coexistence of phases under the equilibrium conditions pertaining in such experiments.

The stable phase relations may also be defined in terms of the activities of components and, in regard to the sulphide minerals, it is the activity of sulphur (aS_2) that is most important. Thus, a 'petrogenetic grid' commonly used to present sulphide mineral relationships plots aS_2 against temperature (actually log aS_2 versus 1/temperature because sulphidation boundaries then plot as straight lines in many cases) as shown in Figure 7. Such diagrams show the progressive development of assemblages through sulphidation from metal to monosulphide to (where appropriate) disulphide.

A limited number of sulphide systems have proved valuable in geothermometry and geobarometry. Notably useful in geothermometry is the mineral arsenopyrite (FeAsS). Phase relations in the Fe–As–S system are shown schematically over the temperature range 688°C to 363°C in Figure 8. In the higher temperature range, arsenopyrite coexists with pyrrhotite, loellingite, arsenic, or an (As,S) liquid (represented on cooling by an arsenic sulphide) in the assemblages labelled 1 to 4. From 491°C to 363°C, arsenopyrite can coexist with pyrite, pyrrhotite, loellingite, arsenic, or (As,S) liquid in the assemblages labelled 5 to 8. Because arsenopyrite composition is a function of both temperature and the equilibrium with the coexisting phases in the system, it can be used in geothermometry only in buffered assemblages, i.e. those labelled 1 to 8 in Figure 8. For these assemblages, the composition of arsenopyrite has been calibrated as a function of temperature to provide a geothermometer. A method has also been devised of using the d_{131} X-ray spacing curve to determine arsenopyrite composition

$$(\text{As, at.}\%) = 866.67d_{131} - 1381.12.$$

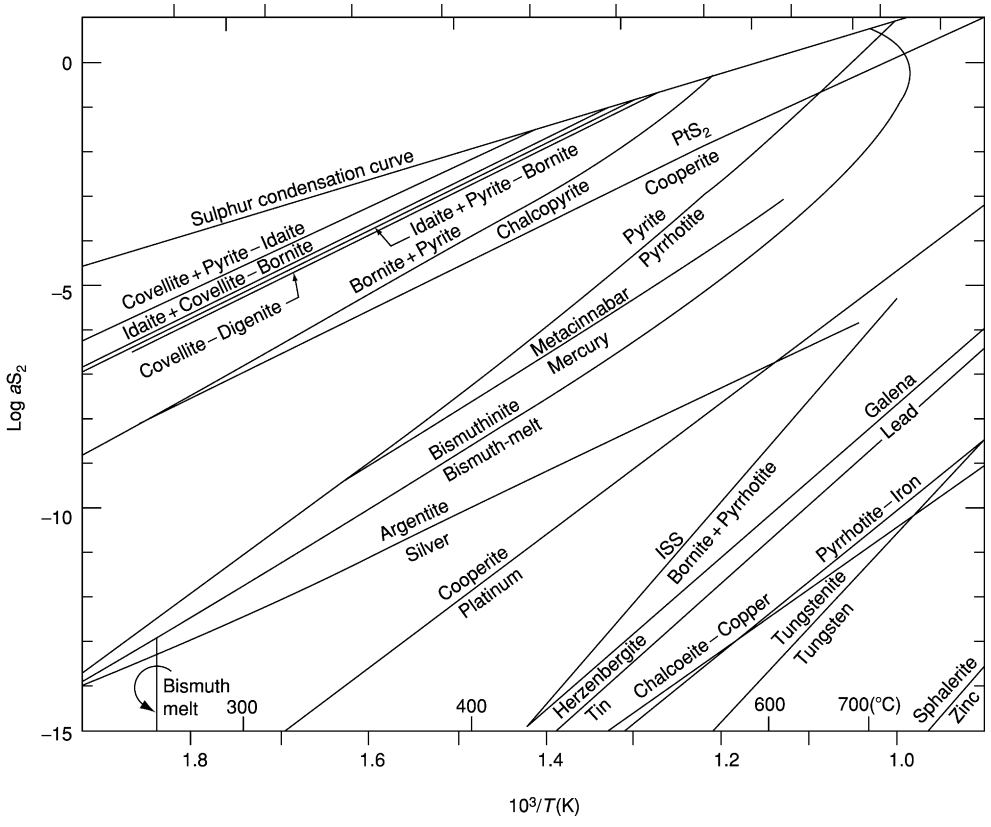


Figure 7 Sulphidation curves for various sulphides as a function of temperature. ISS = intermediate solid solution.

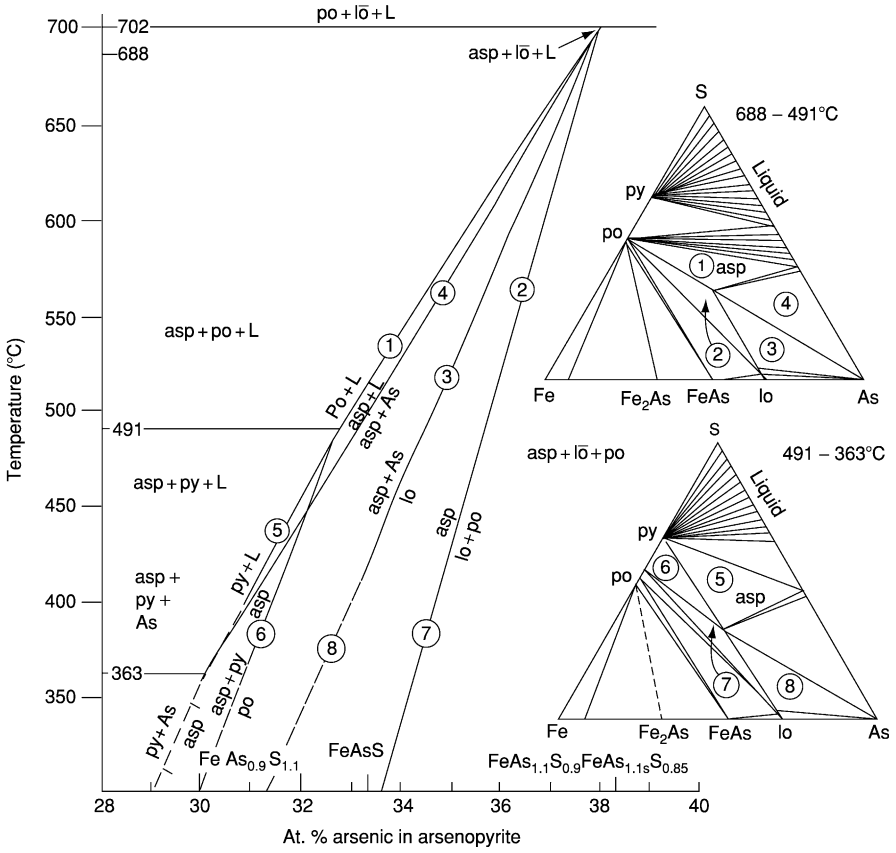


Figure 8 Pseudobinary temperature–composition plot, showing arsenopyrite composition as a function of temperature and equilibrium mineral assemblage. The assemblages numbered 1–8 in the Fe–As–S phase diagrams on the right correspond to the labelled curves in the left-hand diagram. Abbreviations: asp, arsenopyrite; py, pyrite; po, pyrrhotite; lo, lollingite; L, liquid.

Critical evaluation of the arsenopyrite geothermometer, using volume calculations, has revealed that pressure will shift the sulphidation buffer curves in Figure 8 for the system Fe–As–S to higher sulphur activities (e.g. by up to 10 times for arsenopyrite coexisting with pyrite). The effect of pressure on the composition of arsenopyrite buffered by two other phases is not large and can be ignored for low-pressure hydrothermal deposits but needs to be taken into account in metamorphosed deposits.

The sphalerite geobarometer is based on the observation that increasing pressure reduces the FeS content of a sphalerite equilibrated with pyrite and hexagonal pyrrhotite in a manner that is independent of temperature (Figure 9). Sphalerite is sufficiently refractory to preserve the original composition unless it remains in contact with pyrrhotite or chalcopyrite: low-temperature re-equilibration of sphalerite generally reduces the original FeS content and results in pressure estimations that are too high.

A very important aspect of sulphide geochemistry and mineral chemistry is concerned with the mechanisms and kinetics of sulphide oxidation, particularly in regard to pyrite, the overwhelmingly dominant sulphide in most natural systems. This is because of the applications of such studies in both mineral-processing technology and environmental chemistry (particularly in relation to the generation of acid waters). Despite a large amount of work, reflected by a substantial literature, many aspects of pyrite oxidation remain unresolved. This is because, in addition to the large number of variables (whether oxidation is in air, water, or other aqueous fluid; pH; Eh; temperature; presence of various oxidants; role of bacteria), the oxidation of pyrite to form sulphate requires the transfer of eight electrons per sulphur atom and must involve several steps, with sulphur species of intermediate oxidation states. The literature almost universally acknowledges that any oxidation reaction mechanism is surface-reaction controlled. It is suggested that, in aqueous solution, pyrite oxidizes by a combination of the half-reactions

$$FeS_2 + 8H_2O = Fe^{3+} + 2SO_4^{2-} + 16H^+ + 15e^-$$

and

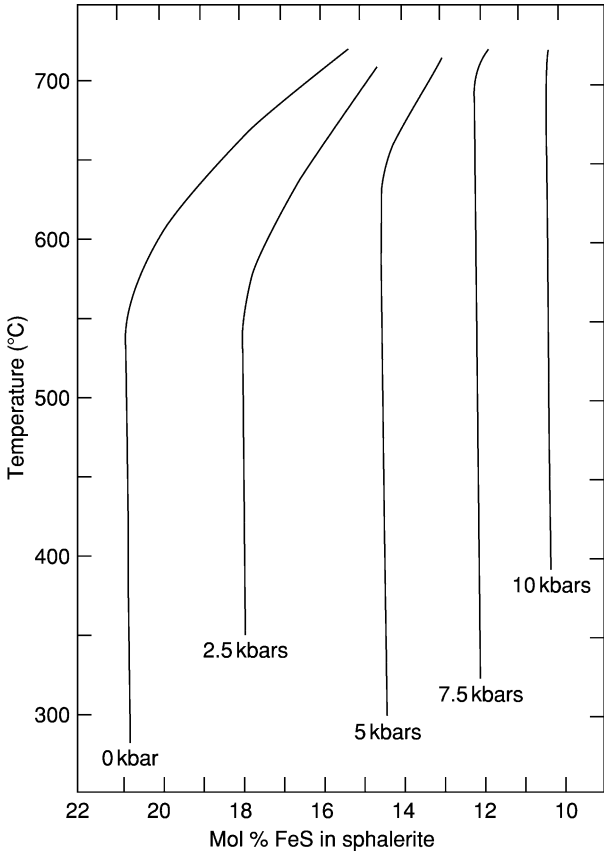


Figure 9 Plots of the FeS contents of sphalerite coexisting with pyrite and hexagonal pyrrhotite at 0, 2.5, 5, 7.5, and 10 kbar at temperatures from 300°C to 700°C. Reproduced with permission from Geochemistry of Hydrothermal Ore Deposits, ed. H. L. Barnes, 1997, John Wiley and Sons.

$$FeS_2 = Fe^{3+} + 2S + 3e^-,$$

for which the Eh and pH determine the dominant pathway: the sulphate route dominates under ambient conditions. At low pH, pyrite may be oxidized by O_2 and Fe^{3+}_{aq}:

$$FeS_2 + 7/2O_2 + H_2O = Fe^{2+} + 2SO_4^{2-} + 2H^+$$

and

$$FeS_2 + 14Fe^{3+} + 8H_2O = 15Fe^{2+} + 2SO_4^{2-} + 16H^+$$

The rates of pyrite oxidation are dramatically affected by the presence of bacteria; for example, at low pH, *Thiobacillus ferrooxidans* can accelerate the oxidation of Fe^{2+} by a factor of more than 10^6. Recent work on the mechanism of pyrite oxidation emphasizes the importance of the fact that pyrite is a semiconductor. Electrons can be transferred from sulphur atoms at an anodic site, where oxygen atoms from water molecules attach to sulphur atoms to form sulphoxy species, through the crystal to cathodic Fe(II) sites, where they are acquired by the oxidant species (usually O_2 or Fe^{3+}). Bulk measurements of oxidation rate show that the concentration of O_2 or Fe^{3+} is critical, so it appears that the 'cathodic' reaction is the rate-determining step.

The study of the surface structures and reactivities of sulphide minerals is being revolutionized by the application of the scanning probe microscopy techniques of scanning tunnelling microscopy and atomic force microscopy. These are leading to models of surface structures and reactions that have been developed from direct observations of the systems at the atomic scale.

Sulphide Paragenesis: Rocks, Sediments, and Ore Deposits

The major ore deposit types that contain substantial amounts of sulphide minerals are shown in Table 3 along with the major minerals they contain, the metals extracted from them, and examples of specific deposits.

Pyrite is exceptionally widespread in its occurrence, being known as an accessory mineral in many of the major rock types that make up the crust of the Earth and as a major or minor phase in many types of ore deposit. However, in ultramafic and mafic intrusive rocks, especially in the so-called 'sulphide nickel' deposits, it is pyrrhotite that dominates. The sulphides in these deposits (pyrrhotite, pentlandite, chalcopyrite, and, in some cases, pyrite) are regarded as having formed by the crystallization of an immiscible sulphide melt that separated from the main silicate melt. Studies of the sulphides in the best-known of all layered basic intrusions, the Bushveld Complex, also show a dominance of pyrrhotite, pentlandite, and chalcopyrite along with minor but economically vital platinum group minerals. In the classic Skaergaard intrusion (Greenland), pyrite is present chiefly in small amounts in the latest rock fraction, the acid granophyre. In the Stillwater Complex, the principal sulphides include pyrite, chalcopyrite, and pyrrhotite, which are fairly conspicuous near the floor of the complex and at one or two higher horizons.

In felsic igneous rocks, the predominant sulphide is pyrite. In addition to its occurrence as an accessory phase in a wide range of such rocks, pyrite is the dominant sulphide mineral in porphyry copper deposits, in which chalcopyrite is the most important ore mineral, along with various binary copper sulphides. Anhedral to euhedral grains of sulphides occur in these ores, where the mineralization is as disseminated grains or veinlets in host intrusions, which range in composition from quartz diorite to

Table 3 The major types of sulphide ore deposit

Type	*Major minerals*	*Metals extracted*	*Examples*
Ores related to mafic and ultramafic intrusions			
Sudbury nickel–copper	po, pn, py, cpy, v	Ni, Cu, Co, PGM	Sudbury, Ontario
Merensky reef platinum	po, pn, cpy	Ni, Cu, PGM	Merensky Reef, South Africa
Ores related to felsic intrusive rocks			
Tin and tungsten skarns	py, cass, sph, cpy, wolf	Sn, W	Pine Creek, California, USA
Zinc–lead skarns	py, sph, gn	Zn, Pb	Ban Ban, Australia
Copper skarns	py, cpy	Cu, Au	Carr Fork, Utah, USA
Porphyry copper/molybdenum	py, cpy, bn, mbd	Cu, Mo, Au	Bingham Canyon, Utah, USA Climax, Colorado, USA
Polymetallic veins	py, cpy, gn, sph, ttd	Cu, Pb, Zn, Ag	Carnsell River, NWT
Ores related to marine mafic extrusive rocks			
Cyprus-type massive sulphides	py, cpy	Cu	Cyprus
Besshi-type massive sulphides	py, cpy, sph, gn	Cu, Pb, Zn	Japan
Ores related to subaerial felsic to mafic extrusive rocks			
Creede-type epithermal veins	py, sph, gn, cpy, ttd, asp	Cu, Pb, Zn, Ag, Au	Creede, Colorado, USA
Almaden mercury type	py, cinn	Hg	Almaden, Spain
Ores related to marine felsic to mafic extrusive rocks			
Kuroko type	py, cpy, gn, sph, ttd, asp,	Cu, Pb, Zn, Ag, Au	Japan
Ores in clastic sedimentary rocks			
Quartz pebble conglomerate gold–uranium	py, uran, Au	Au, U	Witwatersrand, South Africa
Sandstone-hosted lead–zinc	py, sph, gn	Zn, Pb, Cd	Laisvall, Sweden
Sedimentary exhalative lead–zinc (Sedex)	py, sph, gn, cpy, asp ttd, po	Cu, Pb, Zn, Au, Ag	Sullivan, British Columbia, Canada Tynagh, Ireland
Ores in carbonate rocks			
Mississippi Valley type	py, gn, sph	Zn, Pb, Cd, Ga, Ge	South-eastern Missouri

Abbreviation: po, pyrrhotite; pn, pentlandite; py, pyrite; cpy, chalcopyrite; v, violarite; cass, cassiterite; sph, sphalerite; wolf, wolframite; gn, galena; bn, bornite; mbd, molybdenite; ttd, tetrahedrite; asp, arsenopyrite; cinn, cinnabar; uran, uraninite; PGM, platinum group metals.

quartz monzonite. Large masses of pyrite, sphalerite, galena, or chalcopyrite are also found in ore deposits formed by contact metamorphism (skarn deposits). Pyrite occurs, often as a major phase, in the majority of hydrothermal vein deposits and dominates the ore mineralogy of those deposits that might broadly be described as 'volcanogenic'. These include the ores that occur in thick volcanic sequences, such as the Kuroko deposits of Japan. In the so-called 'black ore' of Kuroko-type deposits, irregular masses of galena are intergrown on a fine scale with sphalerite, pyrite, and chalcopyrite. In the Besshi-type deposits of Japan, the sulphides occur in dominantly sedimentary sequences and are again dominated by pyrite, chalcopyrite, sphalerite, and galena.

In sulphide ores in volcanosedimentary sequences that are associated with ophiolite complexes, as in the Troodos Complex deposits (Cyprus), pyrite and chalcopyrite dominate. An understanding of the genesis of these deposits has rapidly developed in recent years through the observation of present-day volcanic and hydrothermal activity on the seafloor. The disseminated to massive stratiform sulphide ores that occur, often conformably, within sedimentary sequences grade into the volcanogenic deposits discussed above. Again, pyrite is the dominant ore mineral in these deposits, which include the Kupferschiefer-Marl Slate of northern Europe and the Copperbelt of Zambia and Zaire. The latter contains not only a range of copper and copper–iron sulphides (chalcopyrite, bornite, chalcocite, covellite) but also significant amounts of cobalt in pyrite and the sulphospinel mineral carrollite.

In the context of sedimentary rocks, sulphides (galena and sphalerite with some pyrite) are major phases in many of the lead–zinc–barite–fluorite ores, which occur chiefly in limestones, and pyrite, along with the copper sulphides, occurs in the uranium–vanadium–copper ores that are associated with sandstones in areas such as the Colorado Plateau, USA. In the gold–uranium ores that occur in conglomerates in the Witwatersrand, South Africa, and Elliot Lake, Ontario, Canada, pyrite is the major opaque phase. As well as occurring as common accessory minerals in black shales, sulphides, especially pyrite and marcasite, are the major opaque mineral in coals. The formation of sulphides in all of these environments is probably linked to the reaction of metals, particularly iron, released by the dissolution of detrital oxides and silicates, with sulphur produced by the bacterial reduction of sulphate present in interstitial waters. This is observed in Holocene sediments, with pyrite (preceded by mackinawite and possibly greigite) forming through a diagenetic process occurring in the reduced

zone beneath the sediment surface. The occurrence of marcasite is much more restricted than that of pyrite. It is never found as a primary phase in any magmatic association and is indicative of environments that have probably been low temperature and where an aqueous phase was present at the time of formation.

Many studies have been undertaken of sulphur isotope ratios ($^{32}S : {}^{34}S$) in sulphide minerals from rocks and ore deposits. Initially these studies were aimed at developing criteria to distinguish ore deposits of igneous–hydrothermal origin from those of sedimentary syngenetic origin. Since bacterial reduction of sulphate substantially enriches the involved sulphur in ^{32}S (by 50% or more in the case of H_2S), such enrichment is taken as an indication that the sulphur has passed through the sedimentary cycle. This key reaction and others that involve isotope exchange and occur during ore formation have been the basis for the application of sulphur isotope studies to genetic problems. However, studies of both synthetic systems and natural ores have shown that the isotopic composition of sulphur is also controlled by changes in pH and oxygen fugacity in the reacting system. When these can be determined, for example by using carbon isotope ratios, valuable information about the formation of a deposit may be obtained. Coexisting pairs of sulphide minerals, such as pyrite co-existing with galena, sphalerite, chalcopyrite, or pyrrhotite, have also been used for sulphur isotope geothermometry. The fractionation of sulphur isotopes between co-existing mineral pairs has been calibrated as a function of temperature for use in such geothermometric studies.

See Also

Analytical Methods: Geochemical Analysis (Including X-Ray); Mineral Analysis. **Mining Geology:** Hydrothermal Ores. **Sedimentary Environments:** Anoxic Environments. **Tectonics:** Hydrothermal Activity; Hydrothermal Vents At Mid-Ocean Ridges. **Thermal Metamorphism**.

Further Reading

Barnes HL (ed.) (1979) *Geochemistry of Hydrothermal Ore Deposits,* 2nd edn. New York: Wiley Interscience.

Barnes HL (ed.) (1997) *Geochemistry of Hydrothermal Ore Deposits,* 3rd edn. New York: Wiley Interscience.

Cabri LJ and Vaughan DJ (1998) *Modern Approaches to Ore and Environmental Mineralogy.* Short Course 27. Ottawa: Mineralogical Association of Canada.

Cotter-Howells J, Campbell L, Batchelder M, and Valsami-Jones E (eds.) (2000) *Environmental Mineralogy: Microbial Interactions, Anthropogenic Influences, Contaminated Land and Waste Management.* Monograph 9. London: Mineralogical Society of Great Britain and Ireland.

Craig JR and Vaughan DJ (1994) *Ore Microscopy and Ore Petrography,* 2nd edn. New York: Wiley Interscience.

Garrels RM and Christ CL (1965) *Solutions, Minerals and Equilibria.* New York: Harper and Row.

Jambor JL and Blowes DW (eds.) (1994) *The Environmental Geochemistry of Sulfide Minewastes.* Short Course Handbook 22. Ottawa: Mineralogical Association of Canada.

Kostov I and Micheeva-Stefanova J (1981) *Sulphide Minerals: Crystal Chemistry, Parageneses and Systematics.* Sofia: Bulgarian Academy of Sciences.

Mills KC (1974) *Thermodynamic Data for Inorganic Sulphides, Selenides and Tellurides.* London: Butterworth.

Ramdohr P (1980) *The Ore Minerals and their Intergrowths,* 2nd English edn. Oxford: Pergamon.

Ribbe PH (ed.) (1974) *Sulfide Mineralogy.* Short Course Notes Volume 1. Washington, DC: Mineralogical Society of America.

Russell MJ and Hall A (1997) The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *Journal of the Geological Society of London* 154: 377–402.

Shuey RT (1975) *Semiconducting Ore Minerals.* Developments in Economic Geology 4. Amsterdam: Elsevier.

Stanton RL (1972) *Ore Petrology.* New York: McGraw Hill.

Vaughan DJ and Craig JR (1978) *Mineral Chemistry of Metal Sulfides.* Cambridge: Cambridge University Press.

Vaughan DJ and Lennie AR (1991) The iron sulphide minerals: their chemistry and role in nature. *Science Progress* 75: 371–388.

Tungstates

P A Williams, University of Western Sydney, Penrith South DC, NSW, Australia

Introduction

Tungstates are compounds containing negatively charged oxytungstate ions. In the mineral kingdom, the simple tetraoxotungstate(VI), or tungstate ion, WO_4^{2-}, is present. As with the case of molybdate(VI), tungstate ions can polymerize under acid conditions to produce polytungstate and heteropolytungstate species, but minerals containing such complex anions are exceedingly rare and remain poorly understood and characterized. Only a few minerals containing essential tungstate are known, though some of these have great commercial significance. The standard potential at 25°C is −1.26 V for

eqn [1], and this explains the fact that tungstates with W in its highest oxidation state (+6) dominate the mineralogy of the element in both the primary and the secondary environments.

$$WO_4^{2-} + 2H_2O + 2e^- \rightarrow WO_2 + 4OH \qquad [1]$$

Tungstenite, WS_2, is an extremely rare species, isomorphous with its common congener molybdenite, MoS_2. Table 1 provides a list of known naturally occurring tungstates. The so-called tungstic acids of stoichiometry $WO_3 \cdot nH_2O$ are not listed because they are simple lattice compounds. However, due to possible confusion as to the attribution of certain species to the tungstate class, Table 1 does include certain species that are in fact related. For example, alumotungstite and ferritungstite are derivatives of WO_3 with the pyrochlore structure. Substitution of W by Al or Fe gives a positive charge discrepancy that is compensated by incorporation of other cations in a vacant lattice site. In related fashion, raspite, $PbWO_4$, the dimorph of stolzite, contains chains of edge-linked WO_6 octahedra. Pinalite, russellite, and tungstibite are complex layer structure oxides; cerotungstite-(Ce) and yttrotungstite-(Y) are complex secondary oxyhydroxide species containing chains of WO_6 octahedra. The same situation may obtain for anthoinite and mpororoite, two poorly characterized minerals that require further investigation.

Table 1 Tungstate(VI) minerals

Mineral	*Chemical composition*
Simple tungstates	
Ferberite	$MnWO_4$
Hübnerite	$FeWO_4$
Paraniite-(Y)	$Ca_2Y(AsO_4)(WO_4)_2$
Raspite	$PbWO_4$
Sanmartinite	$ZnWO_4$
Scheelite	$CaWO_4$
Stolzite	$PbWO_4$
Basic double salts	
Anthoinite	$WAl(O,OH)_3(?)$
Cuprotungstite	$Cu_3(WO_4)_2(OH)_2$
Mpororoite	$WAlO_3(OH)_3 \cdot 2H_2O(?)$
Pinalite	$Pb_3WO_5Cl_2$
Russellite	Bi_2WO_6
Tungstibite	Sb_2WO_6
Complex uranium salts	
Uranotungstite	$(Fe,Ba,Pb)(UO_2)_2(WO_4)(OH)_4 \cdot 12H_2O$
Polytungstates	
Phyllotungstite	$(Ca,Pb)Fe_3H(WO_4)_6 \cdot 10H_2O$
Rankachite	$CaFeV_4W_8O_{36} \cdot 12H_2O$
Other complex species	
Cerotungstite-(Ce)	$CeW_2O_6(OH)_3$
Ferritungstite	$(W,Fe)(O,OH)_3$[a]
Alumotungstite	$(W,Al)(O,OH)_3$[a]
Yttrotungstite-(Y)	$YW_2O_6(OH)_3$

[a]Other cations, including Ca, Na, and Pb, may substitute in a site vacancy to compensate for charge imbalance (see text).

Primary Tungstates

All of the primary tungstates, including ferberite, hübnerite, paraniite-(Y), sanmartinite, and scheelite, contain the simple WO_4^{2-} ion. Paraniite-(Y) and sanmartinite are both extremely rare minerals, but other members of the group have great economic importance, constituting the only minerals of tungsten to have value commercially. Scheelite takes up molybdenum in the lattice and a complete solid solution extends to the isomorphous mineral powellite, $CaMoO_4$. Primary scheelite, however, does not usually contain much molybdenum. The term 'wolframite' was formerly applied to members of the ferberite–hübnerite series and the nomenclature is still commonly in use. Solid solution in the series is in fact limited to about 20 mol% in each end-member and the separate end-member terms are now applied to individual specimens, depending on the exact composition.

Economically valuable deposits of the Alpine Cleft or sedimentary types of scheelite are well documented, but the main settings of scheelite, ferberite, and hübnerite mineralization are either with acid-intrusive rocks or associated skarns. Frequent associates of the tungstates are cassiterite, SnO_2, molybdenite, MoS_2, base metals such as copper, and minor amounts of gold. Important mines were once located in all continents, but at present deposits in China, Indochina, and Brazil are the main sources of tungsten.

Secondary Tungstates

Scheelite is also well known as a secondary mineral, especially associated with the oxidation of ferberite and hübnerite. In this setting, it also may incorporate molybdate in its lattice. Other secondary tungstates are rare, with perhaps the most frequently encountered species being stolzite. Notable locations of stolzite include the Clara mine and other mines in the Black Forest in Germany, and at Broken Hill and the Cordillera mine in New South Wales, Australia. The Clara, Broken Hill, and Cordillera mines have produced raspite as a mineralogical curiosity and the latter is renowned for the association of the dimorphs with cuprotungstite.

As with molybdate (*see* **Rocks and Their Classification**), polymerization of tungstate in acid solutions yields polymeric species. These are represented in the mineral kingdom by phyllotungstite and rankachite. Both are much rarer than are their molybdenum-bearing congeners and are of academic interest

alone. None of the secondary tungstates represents an economically important source of tungsten.

See Also

Minerals: Definition and Classification; Chromates; Molybdates. **Rocks and Their Classification**.

Further Reading

Anthony JA, Bideaux RA, Bladh KW, and Nichols MC (2003) *Handbook of Mineralogy. Volume 5. Borates, Carbonates, Sulfates, Chromates, Germanates, Iodates, Molybdates, Tungstates, etc., and Organic Materials*. Tucson, AZ: Mineral Data Publishing.

Baes CF Jr and Mesmer RE (1986) *The Hydrolysis of Cations*. Malabar, FL: Krieger Publishing Company.

Bard AJ, Parsons R, and Jordan J (1985) *Standard Potentials in Aqueous Solution*. New York: Marcel Dekker.

Birch WD (ed.) (1999) *Minerals of Broken Hill*. Broken Hill, NSW: Broken Hill City Council.

Cox DP and Singer DA (1986) *Mineral Deposit Models. United States Geological Survey Bulletin 1693*. Washington, DC: United States Geological Survey.

Gaines RV, Skinner HCW, Foord EE, Mason B, and Rosenzweig A (eds.) (1997) *Dana's New Mineralogy: The System of Mineralogy of James Dwight Dana and Edward Salisbury Dana*, 8th edn. London: Wiley Europe.

Guilbert JM and Park CF Jr (1986) *The Geology of Ore Deposits*. New York: WH Freeman.

Mandarino JA (1999) *Fleischer's Glossary of Mineral Species 1999*, 8th edn. Tucson, AZ: Mineralogical Record Inc.

Roberts WL, Campbell TJ, and Rapp GR Jr (1990) *Encyclopedia of Minerals*, 2nd edn. New York: Van Nostrand Reinhold.

Sahama TG (1981) *The Secondary Tungsten Minerals, a Review. The Mineralogical Record*, vol. 11. pp. 81–87. Tucson, AZ: Mineralogical Record Inc.

Williams PA (1990) *Oxide Zone Geochemistry*. Chichester: Ellis Horwood.

Vanadates

P A Williams, University of Western Sydney, Sydney, Australia

Naturally occurring vanadate(V) minerals are described here, together with the environments in which they are found. Attention is drawn to the existence of polyvanadates as well as to normal vanadates. In the latter, extensive solid solution involving arsenate and phosphate is evident. The chemistry of polymerisation of vanadate in acidic solution is well understood and is reflected in many examples in the mineral kingdom. Depending upon the prevailing redox potential in the solutions from which such species crystallise, highly coloured, mixed V(IV, V) species are formed, the so-called vanadium bronzes. Significant amounts of vanadium are recovered from secondary vanadate minerals, but much is derived from the treatment of vanadium-rich fly ash and the mining of primary oxide minerals that carry minor amounts of the element.

Vanadates are compounds containing negatively charged oxyvanadium ions and are extremely widespread in the natural environment in small amounts. A list of them is given in Table 1. Nearly all of the vanadates are found in the oxidised zones of base metal ore bodies or in other supergene environments, such as drainage channels and oxidised sedimentary rocks. A few, including wakefieldite-(Y), clinobisvanite, and pucherite occur as accessory minerals in granite pegmatites, while others such as leningradite, averierite, and stoiberite are volcanic sublimates. Many vanadate minerals contain the simple tetraoxovanadate(V) or vanadate ion, VO_4^{3-}. The arrangement of such minerals under the first heading in Table 1, however, obscures the fact that in many cases complex solid solution phenomena attend naturally occurring examples. Some of the formulae indicate solid solution series where these are found to be significant for all known examples of particular phases. Nevertheless, all of the simple vanadates incorporate other ions in their lattices. With cations, this phenomenon is common and several series are known, including the descloizite-mottramite and mounanite-krettnichite pairs, for example. Solid solution involving anionic substitution is ubiquitous. Phosphate and arsenate frequently substitute for vanadate in these minerals, albeit to different extents. The vanadate ion, VO_4^{3-}, is somewhat larger than phosphate and arsenate ions, whose thermochemical radii are 230 and 237 picometres, respectively. Other

Table 1 Vanadate minerals

Normal vanadate, oxyvanadate and chlorovanadate(V) minerals	
Averievite	$Cu_5O_2(VO_4)_2 \cdot n(Cu,Cs,K)Cl$ ($n \approx 1.2$)
Brackebuschite	$Pb_2(Mn,Fe,Zn)(VO_4)_2(OH,H_2O)$
Čechite	$Pb(Fe,Mn)VO_4(OH)$
Cheremnykhite	$Pb_3Zn_3O_2(TeO_4)(VO_4)_2$
Clinobisvanite	$BiVO_4$
Descloizite	$Pb(Zn,Cu)VO_4(OH)$
Dreyerite	$BiVO_4$
Duhamelite	$Pb_2Cu_4Bi(VO_4)_4(OH)_3 \cdot 8H_2O$
Fingerite	$Cu_{11}O_2(VO_4)_6$
Gamagarite	$Ba_2(Fe,Mn)(VO_4)_2(OH)$
Hechtsbergite	$Bi_2(VO_4)(OH)$
Heyite	$Pb_5Fe_2O_4(VO_4)_2$
Howardevansite	$NaCu(Fe,Al,Mn)_2(VO_4)_3$
Kolovratite	$Ni_xZn_y(VO_4)_z \cdot nH_2O$
Kombatite	$Pb_{14}O_9(VO_4)_2Cl_4$
Krettnichite	$PbMn_2(VO_4)_2(OH)_2$
Leningradite	$PbCu_3(VO_4)_2Cl_2$
Lyonsite	$Cu_3Fe_4(VO_4)_6$
Mcbirneyite	$Cu_3(VO_4)_2$
Mottramite	$Pb(Cu,Zn)VO_4(OH)$
Mounanaite	$PbFe_2(VO_4)_2(OH)_2$
Namibite	$Cu(BiO)_2(VO_4)(OH)$
Palenzonaite	$NaCa_2Mn_2[(V,As,Si)O_4]_3$
Pottsite	$HPbBi(VO_4)_2 \cdot 2H_2O$
Pucherite	$BiVO_4$
Pyrobelonite	$PbMnVO_4(OH)$
Reppiaite	$Mn_5(VO_4)_2(OH)_4$
Rusakovite	$(Fe,Al)_5[(V,P)O_4]_2(OH)_9 \cdot 3H_2O$
Santafeite	$(Na,Ca,Sr)_3(Mn,Fe)_2Mn_2(VO_4)_4(OH,O)_5 \cdot 2H_2O$
Schumacherite	$Bi_3[(V,As,P)O_4]_2(OH)$
Stoiberite	$Cu_5O_2(VO_4)_2$
Tangeite	$CaCuVO_4(OH)$
Turanite	$Cu_5(VO_4)_2(OH)_4$
Vanadinite	$Pb_5(VO_4)_3Cl$
Vésigniéite	$BaCu_3(VO_4)_2(OH)_2$
Wakefieldite-(Ce)	$(Ce,La,Nd,Y,Pr,Sm)[(V,As)O_4]$
Wakefieldite-(Y)	YVO_4
Layered uranyl vanadate(V) minerals	
Carnotite	$K_2(UO_2)_2(VO_4)_2 \cdot 3H_2O$
Curienite	$Pb(UO_2)_2(VO_4)_2 \cdot 5H_2O$
Francevillite	$(Ba,Pb)(UO_2)(VO_4)_2 \cdot 5H_2O$
Margaritasite	$(Cs,K,H)_2(UO_2)_2(VO_4)_2 \cdot H_2O$
Metatyuyamunite	$Ca(UO_2)_2(VO_4)_2 \cdot 4H_2O$
Metavanuralite	$Al(UO_2)_2(VO_4)_2(OH) \cdot 8H_2O$
Sengierite	$Cu(UO_2)(VO_4)(OH) \cdot 3H_2O$
Strelkinite	$Na_2(UO_2)_2(VO_4)_2 \cdot 6H_2O$
Tyuyamunite	$Ca(UO_2)_2(VO_4)_2 \cdot 8H_2O$
Vanuralite	$Al(UO_2)_2(VO_4)_2(OH) \cdot 11H_2O$
Vanuranylite	$(H_3O,Ba,Ca \cdot K)_2(UO_2)_2(VO_4)_2 \cdot 4H_2O(?)$
Pyrovanadate(V) minerals	
Blossite	$Cu_2(V_2O_7)$
Chervetite	$Pb_2(V_2O_7)$
Fianelite	$Mn_2(V_2O_7) \cdot 2H_2O$
Volborthite	$Cu_3(V_2O_7)(OH)_2 \cdot 2H_2O$
Ziesite	$Cu_2(V_2O_7)$
Minerals containing the decavanadate(V) isopolyanion	
Hummerite	$K_2Mg_2(V_{10}O_{28}) \cdot 16H_2O$
Huemulite	$Na_4Mg(V_{10}O_{28}) \cdot 24H_2O$
Pascoite	$Ca_3(V_{10}O_{28}) \cdot 17H_2O$
Rauvite	$Ca(UO_2)_2(V_{10}O_{28}) \cdot 16H_2O$
Metavanadate(V) minerals[a]	
Alvanite	$(Zn,Ni)Al_4(VO_3)_2(OH)_{12} \cdot 2H_2O$
Delrioite	$CaSr(V_2O_6)(OH)_2 \cdot 3H_2O$
Metadelrioite	$CaSrV_2O_6(OH)_2$
Metamunirite	$NaVO_3$
Metarossite	$Ca(V_2O_6) \cdot 2H_2O$
Munirite	$NaVO_3 \cdot (2\text{-}x)H_2O$, chains
Rossite	$Ca(V_2O_6) \cdot 4H_2O$
Layered $V_6O_{16}^{n-}$ vanadate(IV,V) minerals of the hewittite group[a]	
Barnesite	$(Na,Ca)(V_6O_{16}) \cdot 3H_2O$
Grantsite	$(Na,Ca)(V_6O_{16}) \cdot 4H_2O$
Hewettite	$Ca(V_6O_{16}) \cdot 9H_2O$
Hendersonite	$Ca_{1.3}(V_6O_{16}) \cdot 6H_2O$
Metahewettite	$Ca(V_6O_{16}) \cdot 3H_2O$
Layered $V_8O_{20}^{n-}$ vanadate(IV,V) minerals of the straczekite group[a]	
Bariandite	$Al_{0.6}(V_8O_{20}) \cdot 18H_2O$
Bokite	$(Al,Fe)_{1.4}[(V,Fe)_8O_{20} \cdot 7 \cdot 4H_2O$
Corvusite	$(Na,KCa,Mg)_2(V_8O_{20}) \cdot 6\text{-}104H_2O$
Fernandinite	$Ca_{0.6}(V_8O_{20}) \cdot 10H_2O$
Straczekite	$(Ca,K,Ba)_2(V_8O_{20}) \cdot 6H_2O$
Other vanadium bronze-type minerals[a]	
Bannermanite	$Na_{0.7}(V_6O_{15})$
Fervanite	$Fe_4(V_4O_{16}) \cdot 5H_2O$
Kazakhstanite[b]	$Fe_5(V_{15}O_{39})(OH)_9 \cdot 9H_2O$
Melanovanadite	$Ca(V_4O_{10}) \cdot 5H_2O$
Schubnelite	$Fe_{2-x}(V_2O_4)(OH)_8$ (x is small)
Miscellaneous vanadate minerals	
Phosphovanadylite[c]	$(Ba,Ca,K,Na)_x[(V,Al)_4P_2(O,OH)_{16}] \cdot 12H_2O$
Rankachite[d]	$CaFeV_4W_8O_{36} \cdot 12H_2O$
Satpaevite[e]	$Al_{12}V_8O_{37} \cdot 30H_2O$
Sherwoodite[f]	$Ca_9[AlV_{14}O_{40}]_2 \cdot 56H_2O$
Simplotite[g]	$CaV_4O_9 \cdot 5H_2O$
Uvanite	$(UO_2)_2(V_6O_{17}) \cdot 15H_2O(?)$
Vanalite[h]	*ca* $NaAl_9(V_{12}O_{44})(OH)_4 \cdot 33H_2O$

[a] Vanadium bronzes that contain variable amounts of V(IV); stoichiometries are indicative.
[b] Related to minerals of the straczekite group.
[c] Contains V(IV) in a V_4O_{16} cluster linked by phosphate groups.
[d] A heteropolytungstate containing V(V) and W(VI).
[e] A hydrated, mixed Al(III)-V(IV,V) oxide cluster species.
[f] Sherwoodite is the calcium salt of the 14-vanadoaluminate heteropolyanion $(AlV_{14}O_{40})^{n-}$, in which a portion of the vanadium is in the V(IV) state.
[g] Contains a V(IV) oxyanion cluster.
[h] Contains a mixed V(IV,V) isopoly- or hereropolyanion.

comparable ions, such as silicate, may also substitute with charge balance compensation. Thus, many of the formulae conceal more complex compositions for natural materials, although substitution for vanadate is more limited in many cases than is found for mutually soluble phosphate and arsenate analogues and related minerals.

Separate mention is made in Table 1 of the layered uranyl vanadates, which represent important ore minerals both of uranium and vanadium. Carnotite and tyuyamunite are perhaps the most significant in this respect. All of these species are members of the group of minerals known as the uranium micas. Their structures are based on infinite sheets of linked uranyl

(UO_2^{2+}) and vanadate ions (phosphate and arsenate in the cases of the autunite, meta-autunite, and related groups), with other cations, hydroxide ions, and water molecules lying between the sheets. Limited substitution of phosphate and arsenate is reported for the vanadate minerals, which are extremely insoluble species.

In aqueous V(V) solutions at low concentrations, only mononuclear species are present, these comprising VO_2^+, $VO(OH)_3$, $VO_2(OH)_2^-$, VO_3OH^{2-}, and VO_4^{3-}. In contrast to this, at higher total V(V) concentrations under acid conditions, extensive polymerisation of vanadate occurs according to the following equations:

$$2VO_4^{3-} + 2H^+ \leftrightarrow V_2O_7^{4-} + H_2O \quad \text{(pyrovanadate)}$$

$$3VO_4^{3-} + 6H^+ \leftrightarrow V_3O_9^{3-} + 3H_2O \quad \text{(metavanadate)}$$

$$2V_2O_7^{4-} + 4H^+ \leftrightarrow V_4O_{12}^{4-} + 2H_2O \quad \text{(metavanadate)}$$

$$5V_4O_{12}^{4-} + 8H^+ \leftrightarrow 2V_{10}O_{28}^{6-} + 4H_2O \quad \text{(decavanadate)}$$

Protonated analogues are also present in solution. The pyrovanadates, metavanadates, and decavanadates are all expressed in a number of mineral structures, as listed in Table 1. Metavanadates, in the solid state, consist of infinite chains of VO_3^- composition, $(VO_3^-)_n$.

Divanadate and tetravanadate chains are the basic structural units of many of the vanadium bronzes, also listed in Table 1. These minerals assume a variety of colours, owing to the fact that a variable though usually small proportion of the V(V) present is reduced to V(IV). Consequent variations of stoichiometry are noted as a result of charge compensation and this is an additional feature of other members of the mineral groups containing both metavanadate and decavanadate, and their protonated derivatives. Other vanadium bronzes are based on different structural motifs, but all share the above characteristics in terms of the presence of V(IV) and attendant compositional variation. The natural expression of the vanadium bronzes, exhaustively studied as artificial compounds, is dependent upon subtle fluctuations of redox potential in the solutions from which they form. They, and other vanadium species, are important constituents of sandstone vanadium deposits and their geochemical relationships have been thoroughly explored, based on pioneering studies dating back nearly five decades. The mineral phosphvanadylite contains V(IV) in a V_4O_{16} cluster linked by phosphate groups, but its chemistry of formation remains to be elucidated.

Like the tetaoxomolybdate(VI) and tetaoxotungstate(VI) ions, vanadate forms heteropolyanions in acid solutions. Rankachite and sherwoodite are two examples, the former carrying W(VI) in the cluster and the latter being a naturally occurring salt of the well-known synthetic 14-vanadoaluminate heteropolyanion $(AlV_{14}O_{40}]^{n-}$. No doubt, other congeners exist in nature in line with observations concerning the geochemistry of molybdenum and tungsten.

As mentioned above, vanadate minerals are an important source of the element vanadium. In addition to species already singled out in this connection, vanadinite, mottramite, and descloizite have been exploited as vanadium ores. However, the bulk of present vanadium production comes from vanadium-bearing oxides (principally magnetite) and by-product refining.

See Also

Minerals: Arsenates. **Sedimentary Rocks:** Phosphates.

Further Reading

Anthony JA, Bideaux RA, Bladh KW, and Nichols MC (1997) *Handbook of Mineralogy. Volume 3. Halides, Hydroxides, Oxides*. Tucson: Mineral Data Publishing.

Anthony JA, Bideaux RA, Bladh KW, and Nichols MC (2000) *Handbook of Mineralogy. Volume 4. Arsenates, Phosphates, Vanadates*. Tucson: Mineral Data Publishing.

Anthony JA, Bideaux RA, Bladh KW, and Nichols MC (2003) *Handbook of Mineralogy. Volume 5. Borates, Carbonates, Sulfates, Chromates, Germanates, Iodates, Molybdates, Tungstates, etc., and Organic Materials*. Tucson: Mineral Data Publishing.

Baes CF, Jr and Mesmer RE (1986) *The Hydrolysis of Cations*. Reprint Edition. Malabar, Florida: Krieger Publishing Company.

Bard AJ, Parsons R, and Jordan J (1985) *Standard Potentials in Aqueous Solution*. New York: Marcel Dekker.

Evans HT, Jr and Hughes JM (1990) Crystal chemistry of the natural vanadium bronzes. *The American Mineralogist* 75: 508–521.

Evans HT, Jr and Garrels RM (1958) Thermodynamic equilibria of vanadium in aqueous systems as applied to the interpretation of the Colorado Plateau ore deposits. *Geochimica et Cosmochimica Acta* 15: 131–149.

Gaines RV, Skinner HCW, Foord EE, Mason B, and Rosenzweig A (eds.) (1997) *Dana's New Mineralogy: The System of Mineralogy of James Dwight Dana and Edward Salisbury Dana*, 8th edn. London: Wiley Europe.

Mandarino JA (1999) *Fleischer's Glossary of Mineral Species 1999*, 8th edn. Tucson: Mineralogical Record Inc.

Roberts WL, Campbell TJ, and Rapp GR, Jr (1990) *Encyclopedia of Minerals*, Second Edition. New York: Van Nostrand Reinhold.

Schindler M, Hawthorne FC, and Baur WH (2000) A crystal-chemical approach to the composition and occurrence of vanadium minerals. *The Canadian Mineralogist* 38: 1443–1456.

Williams PA (1990) *Oxide Zone Geochemistry*. Chichester: Ellis Horwood.

Zeolites

W S Wise, University of California–Santa Barbara, Santa Barbara, CA, USA

Introduction

Zeolites are aluminosilicate minerals that occur as low-temperature (generally less than 200°C) alteration products of volcanic and feldspathic rocks. They are well known in cavities of basalt, having crystallized as a result of diagenetic or hydrothermal alteration (Figure 1). Some zeolites completely replace rhyolitic tuff in saline alkaline lacustrine environments or through groundwater percolation. Thick sequences of sediment from arc-source terranes contain several different kinds of zeolite, which formed through diagenetic alteration and very low-grade metamorphism. These occurrences of zeolites have economic significance, because they either produce useful rock or affect the porosity of reservoir rocks.

A zeolite is a hydrated aluminosilicate mineral with a structure characterized by a framework of linked tetrahedra, each consisting of four oxygen atoms surrounding a silicon or aluminium cation. This three-dimensional network has open cavities in the forms of channels and cages, which are occupied by water molecules and non-framework cations. The channels are large enough to allow the passage of guest ionic and molecular species. This feature allows certain zeolites to exchange non-framework cations with the surrounding solution or to act as molecular sieves. For example, clinoptilolite (Table 1) can selectively remove radiogenic strontium from nuclear wastewater, and chabazite can remove carbon dioxide from methane in the gas generated in landfill sites.

Figure 1 Crystals of laumontite and calcite in a cavity in an altered basalt near Kailua, northeast Oahu, Hawaii. This area is within the former conduit system of the 2.6 Ma Koolau volcano. It is likely that the pervasive rock alteration and zeolite formation is a result of hydrothermal activity when the volcano was active. The long dimension of the cavity is 8 cm.

Each zeolite mineral or, in many cases, zeolite series has a different framework arrangement. Table 1 lists the currently recognized zeolite mineral species and series, their generalized formulae, and structure-type codes. Many more zeolites have been synthesized in the laboratory. The Structure Commission of the International Zeolite Association assigns a three-letter code to each different arrangement, including the synthetic ones. At the time of writing there are 145 structure types, of which 42 occur as minerals. For updates see the International Zeolite Association website http://www.iza-structure.org/.

Structure of Zeolites

All silicates have silicon (or other elements such as aluminium) located in the small space between four oxygen anions. This site is called a tetrahedral site because the cation is strongly bonded to the four oxygen anions. In framework silicates each oxygen anion is shared between two tetrahedral sites. Therefore, the resulting framework has two oxygen anions for each silicon cation, giving the composition $Si^{4+}O_2^{2-}$, as in the mineral quartz (*see* **Minerals:** Quartz). Most other framework silicate minerals have aluminium in some of the tetrahedral sites, which results in a charge imbalance. For example, in some feldspars (*see* **Minerals:** Feldspars) the framework has the composition $[AlSi_3O_8]^-$, and the charge is balanced by the addition of alkali cations in sites near the oxygen anions bonded to the aluminium tetrahedral site.

The zeolite minerals all have framework arrangements with channels or cages (large pockets). In order to illustrate some of the variety among the zeolites, we will examine three different framework structures: natrolite (NAT), chabazite (CHA), and heulandite (HEU).

Natrolite forms long prismatic crystals, in which the basic unit of the structure is the $[Al_2Si_3O_{10}]^{2-}$ chain (Figure 2). Continuous repetition of this unit forms long chains parallel to the prism length. Cross-linking of these chains forms channels parallel to the chains. To balance the charge in the framework, caused by the inclusion of aluminium in two of the

Table 1 Names, general formulae, and framework type of accepted zeolite species

Name	*Generalized formula*	*Z (Z = number of formula units per unit cell.)*	*Structure-type code*
Amicite	$K_4Na_4[Al_8Si_8O_{32}].10H_2O$	1	GIS
Ammonioleucite	$(NH_4)[AlSi_2O_6]$	16	ANA
Analcime	$Na[AlSi_2O_6].H_2O$	16	ANA
Barrerite	$Na_8[Al_8Si_{28}O_{72}].26H_2O$	2	STI
Bellbergite	$(K,Ba,Sr)_2Sr_2Ca_2(Ca,Na)_4[Al_{18}Si_{18}O_{72}].30H_2O$	1	EAB
Bikitaite	$Li[AlSi_2O_6].H_2O$	2	BIK
Boggsite	$(Ca,Na_{0.5},K_{0.5})_9[Al_{18}Si_{78}O_{192}].70H_2O$	1	BOG
Brewsterite series	$(Sr,Ba)_2[Al_4Si_{12}O_{32}].10H_2O$	1	BRE
Brewsterite-Ba	$(Ba,Sr)_2[Al_4Si_{12}O_{32}].10H_2O$		
Brewsterite-Sr	$(Sr,Ba)_2[Al_4Si_{12}O_{32}].10H_2O$		
Chabazite series	$(Ca_{0.5},Na,K)_x[Al_xSi_{12-x}O_{24}].12H_2O$, $x = 2.4–5.0$	1	CHA
Chabazite-Ca	$(Ca_{0.5},K,Na)_x[Al_xSi_{12-x}O_{24}].12H_2O$, $x = 2.4–5.0$		
Chabazite-K	$(K,Na,Ca_{0.5})_x[Al_xSi_{12-x}O_{24}].12H_2O$, $x = 3.0–4.5$		
Chabazite-Na	$(Na,K,Ca_{0.5})_x[Al_xSi_{12-x}O_{24}].12H_2O$, $x = 2.5–4.8$		
Chabazite-Sr	$(Sr_{0.5},Ca_{0.5},K)_4[Al_4Si_8O_{24}].12H_2O$		
Chiavennite	$CaMn[Be_2Si_5O_{13}(OH)_2].2H_2O$	4	-CHI
Clinoptilolite series	$(Na,K,Ca_{0.5},Mg_{0.5})_6[Al_6Si_{30}O_{72}].20H_2O$	1	HEU
Clinoptilolite-Ca	$(Ca_{0.5},Na,K,Sr_{0.5},Ba_{0.5},Mg_{0.5})_6[Al_6Si_{30}O_{72}].20H_2O$		
Clinoptilolite-K	$(K,Na,Ca_{0.5},Sr_{0.5},Ba_{0.5},Mg_{0.5})_6[Al_6Si_{30}O_{72}].20H_2O$		
Clinoptilolite-Na	$(Na,K,Ca_{0.5},Sr_{0.5},Ba_{0.5},Mg_{0.5})_6[Al_6Si_{30}O_{72}].20H_2O$		
Cowlesite	$Ca[Al_2Si_3O_{10}].5.3H_2O$	52 unknown	
Dachiardite series	$(Ca_{0.5},Na,K)_{4–5}[Al_{4–5}Si_{20–19}O_{48}].13H_2O$	1	DAC
Dachiardite-Ca	$(Ca_{0.5},Na,K)_5[Al_5Si_{19}O_{48}].13H_2O$		
Dachiardite-Na	$(Na,K,Ca_{0.5})_4[Al_4Si_{20}O_{48}].13H_2O$		
Edingtonite	$Ba[Al_2Si_3O_{10}].4H_2O$	2	EDI
Epistilbite	$(Ca,Na_2)_3[Al_6Si_{18}O_{48}].16H_2O$	1	EPI
Erionite series	$K_2(Na,Ca_{0.5})_8[Al_{10}Si_{26}O_{72}].30H_2O$	1	ERI
Erionite-Ca	$K_2(Ca_{0.5},Na)_8[Al_{10}Si_{26}O_{72}].30H_2O$		
Erionite-K	$K_2(K,Na,Ca_{0.5})_7[Al_9Si_{27}O_{72}].30H_2O$		
Erionite-Na	$K_2(Na,Ca_{0.5})_7[Al_9Si_{27}O_{72}].30H_2O$		
Faujasite series	$(Na,Ca_{0.5},Mg_{0.5},K)_x[Al_xSi_{12-x}O_{24}].16H_2O$, $x = 3.2–3.8$	16	FAU
Faujasite-Ca	$(Ca_{0.5},Na,Mg_{0.5},K)_x[Al_xSi_{12-x}O_{24}].16H_2O$, $x = 3.3–3.9$		
Faujasite-Mg	$(Mg_{0.5},Ca_{0.5},Na,K)_{3.5}[Al_{3.5}Si_{8.5}O_{24}].16H_2O$		
Faujasite-Na	$(Na,Ca_{0.5},Mg_{0.5},K)_x[Al_xSi_{12-x}O_{24}].16H_2O$, $x = 3.2–4.3$		
Ferrierite series	$(K,Na,Mg_{0.5},Ca_{0.5})_6[Al_6Si_{30}O_{72}].20H_2O$	1	FER
Ferrierite-K	$(K,Na,Mg_{0.5},Ca_{0.5})_6[Al_6Si_{30}O_{72}].20H_2O$		
Ferrierite-Mg	$(Mg_{0.5},K,Na,Ca_{0.5})_6[Al_6Si_{30}O_{72}].20H_2O$		
Ferrierite-Na	$(Na,K,Mg_{0.5},Ca_{0.5})_6[Al_6Si_{30}O_{72}].20H_2O$		
Garronite	$(Ca_{0.5},Na)_6[Al_6Si_{10}O_{32}].14H_2O$	1	GIS
Gaultite	$Na_4[Zn_2Si_7O_{18}].5H_2O$	8	VSV
Gismondine	$Ca_4[Al_8Si_8O_{32}].18H_2O$	1	GIS
Gmelinite series	$(Na,Ca_{0.5},K)_8[Al_8Si_{16}O_{48}].22H_2O$	1	GME
Gmelinite-Ca	$(Ca_{0.5},Sr,K,Na)_8[Al_8Si_{16}O_{48}].22H_2O$		
Gmelinite-K	$(K,Ca_{0.5},Na)_8[Al_8Si_{16}O_{48}].22H_2O$		
Gmelinite-Na	$(Na,K,Ca_{0.5})_8[Al_8Si_{16}O_{48}].22H_2O$		
Gobbinsite	$Na_5[Al_5Si_{11}O_{32}].12H_2O$	1	GIS
Gonnardite	$(Na,Ca_{0.5})_{8–10}[Al_{8+x}Si_{12-x}O_{40}].12H_2O$, $x = 0–2$	1	NAT
Goosecreekite	$Ca[Al_2Si_6O_{16}].5H_2O$	2	GOO
Gottardiite	$(Na,K)_3Mg_3Ca_5[Al_{19}Si_{117}O_{272}].93H_2O$	1	NES
Harmotome	$(Ba_{0.5},Ca_{0.5},K,Na)_5[Al_5Si_{11}O_{32}].12H_2O$	1	PHI
Heulandite series	$(Ca_{0.5},Sr_{0.5},Ba_{0.5},Mg_{0.5},Na,K)_9[Al_9Si_{27}O_{72}].24H_2O$	1	HEU
Heulandite-Ca	$(Ca_{0.5},Sr_{0.5},Ba_{0.5},Mg_{0.5},Na,K)_9[Al_9Si_{27}O_{72}].24H_2O$		
Heulandite-Na	$(Na,Ca_{0.5},Sr_{0.5},Ba_{0.5},Mg_{0.5},K)_9[Al_9Si_{27}O_{72}].24H_2O$		
Heulandite-K	$(K,Ca_{0.5},Sr_{0.5},Ba_{0.5},Mg_{0.5},Na)_9[Al_9Si_{27}O_{72}].24H_2O$		
Heulandite-Sr	$(Sr_{0.5},Ca_{0.5},Ba_{0.5},Mg_{0.5},Na,K)_9[Al_9Si_{27}O_{72}].24H_2O$		
Hsianghualite	$Li_2Ca_3[Be_3Si_3O_{12}]F_2$	8	ANA
Kalborsite	$K_6[Al_4Si_6O_{20}]B(OH)_4Cl$	2	?EDI
Laumontite	$Ca_4[Al_8Si_{16}O_{48}].18H_2O$	1	LAU
Leucite	$K[AlSi_2O_6]$	16	ANA

Continued

Table 1 Continued

Name	*Generalized formula*	*Z (Z = number of formula units per unit cell.)*	*Structure-type code*
Levyne series	$(Ca_{0.5},Na,K)_6[Al_6Si_{12}O_{36}].17H_2O$	1	LEV
Levyne-Ca	$(Ca_{0.5},Sr,K,Na)_6[Al_6Si_{12}O_{36}].17H_2O$		
Levyne-Na	$(Na,K,Ca_{0.5})_6[Al_6Si_{12}O_{36}].17H_2O$		
Lovdarite	$K_4Na_{12}[Be_8Si_{28}O_{72}].18H_2O$	1	LOV
Maricopaite	$(Pb_7Ca_2)[Al_{12}Si_{36}(O,OH)_{100}].nH_2O$	1	?MOR
Mazzite	$(Mg_{2.5}K_2Ca_{1.5})[Al_{10}Si_{26}O_{72}].30H_2O$	1	MAZ
Merlinoite	$(K,Ca_{0.5},Ba_{0.5},Na)_{10}[Al_{10}Si_{22}O_{64}].22H_2O$	1	MER
Mesolite	$Na_2Ca_2[Al_6Si_9O_{30}].8H_2O$	8	NAT
Montesommaite	$K_9[Al_9Si_{23}O_{64}].10H_2O$	1	MON
Mordenite	$(Na_2,Ca,K_2)[Al_8Si_{40}O_{96}].28H_2O$	1	MOR
Mutinaite	$Na_3Ca_4[Al_{11}Si_{85}O_{192}].60H_2O$	1	MFI
Nabesite	$Na_2BeSi_4O_{10}.4H_2O$	4	NAB
Natrolite	$Na_2[Al_2Si_3O_{10}].2H_2O$	8	NAT
Offretite	$CaKMg[Al_5Si_{13}O_{36}].16H_2O$	1	OFF
Pahasapaite	$(Ca_{5.5}Li_{3.6}K_{1.2}Na_{0.2}\square_{13.5})Li_8[Be_{24}P_{24}O_{96}].38H_2O$	1	RHO
Parthéite	$Ca_2[Al_4Si_4O_{15}(OH)_2].4H_2O$	4	-PAR
Paulingite series	$(K,Ca_{0.5},Na,Ba_{0.5})_{10}[Al_{10}Si_{32}O_{84}].27–44H_2O$	16	PAU
Paulingite-K	$(K,Ca_{0.5},Na)_{10}[Al_{10}Si_{32}O_{84}].44H_2O$		
Paulingite-Ca	$(Ca_{0.5},K,Na,Ba_{0.5})_{10}[Al_{10}Si_{32}O_{84}].27–34H_2O$		
Perlialite	$K_9Na(Ca,Sr)[Al_{12}Si_{24}O_{72}].15H_2O$	1	LTL
Phillipsite series	$(K,Na,Ca_{0.5})_x[Al_xSi_{16-x}O_{32}].12H_2O,\ x = 3.8–6.4$	1	PHI
Phillipsite-Na	$(Na,K,Ca_{0.5})_x[Al_xSi_{16-x}O_{32}].12H_2O,\ x = 3.7–6.7$		
Phillipsite-K	$(K,Na,Ca_{0.5})_x[Al_xSi_{16-x}O_{32}].12H_2O,\ x = 3.8–6.4$		
Phillipsite-Ca	$(Ca_{0.5},K,Na)_x[Al_xSi_{16-x}O_{32}],12H_2O,\ x = 4.1–6.8$		
Pollucite	$(Cs,Na)[AlSi_2O_6].nH_2O$ $(Cs + n = 1)$	16	ANA
Roggianite	$Ca_2[BeAl_2Si_4O_{13}(OH)_2].2–2.5H_2O$	8	-ROG
Scolecite	$Ca[Al_2Si_3O_{10}].3H_2O$	8	NAT
Stellerite	$Ca_4[Al_8Si_{28}O_{72}].28H_2O$	2	STI
Stilbite series	$(Ca_{0.5},Na,K)_9[Al_9Si_{27}O_{72}].30H_2O$	1	STI
Stilbite-Ca	$(Ca_{0.5},Na,K)_9[Al_9Si_{27}O_{72}].30H_2O$		
Stilbite-Na	$(Na,K,Ca_{0.5})_9[Al_9Si_{27}O_{48}].28H_2O$		
Terranovaite	$NaCa[Al_3Si_{17}O_{40}].13H_2O$	1	TER
Thomsonite series	$(Ca,Sr)_2Na[Al_5Si_5O_{20}].6–7H_2O$	4	THO
Thomsonite-Ca	$Ca_2Na[Al_5Si_5O_{20}].6H_2O$		
Thomsonite-Sr	$(Sr,Ca)_2Na[Al_5Si_5O_{20}].7H_2O$		
Tschernichite	$(Ca,Mg,Na_{0.5})[Al_2Si_6O_{16}].8H_2O$	8	BEA
Tschörtnerite	$Ca_4(K_2,Ca,Sr,Ba)_3Cu_3(OH)_8[Al_{12}Si_{12}O_{48}].nH_2O$	16	TSC
Wairakite	$Ca[Al_2Si_4O_{12}].2H_2O$	8	ANA
Weinebeneite	$Ca[Be_3PO_8(OH)_2].4H_2O$	4	WEI
Willhendersonite	$Ca_{3-x}K_x[Al_6Si_6O_{24}].10H_2O,\ x = 0.0–2.0$	1	CHA
Yugawaralite	$Ca[Al_2Si_6O_{16}].4H_2O$	2	YUG

Reproduced with permission from Deer WA, Howie RA, Wise WS, and Zussman J (2004) *Rock Forming Minerals*, *Volume 4B Framework Silicates: Silica Minerals*, *Feldspathoids and the Zeolites*. London: The Geological Society.

five tetrahedral sites, sodium cations are bonded within the channels (Figure 3). The bonding of these sodium cations to opposite sides of the channel partially collapses the channel from a possible square cross-section to an elliptical one. Each sodium cation is bonded by two water molecules, which are in turn each bonded to two sodium cations.

The crystal form of chabazite is a rhombohedron that is almost cubic. The framework consists of six-member rings of tetrahedra, which are arranged in the sequence AABBCC... (Figure 4). Each different letter represents a different position of the six-member ring around the stacking direction parallel to the *c*-axis. The distribution of aluminium in the tetrahedral sites is nearly random. Non-framework cations partially fill four different sites: (i) outside the double six-member ring and bonded with three oxygen atoms of the ring; (ii) near the eight-member ring port of the chabazite cage and bonded with three framework oxygen atoms; (iii) near the centre of the cage and not bonded with the framework; and (iv) at the centre of the eight-member ring. Positions (i) and (iii) generally contain most of the cations.

Crystals of heulandite and clinoptilolite are tabular with a perfect cleavage. The framework consists of sheets of tetrahedra with no openings large enough

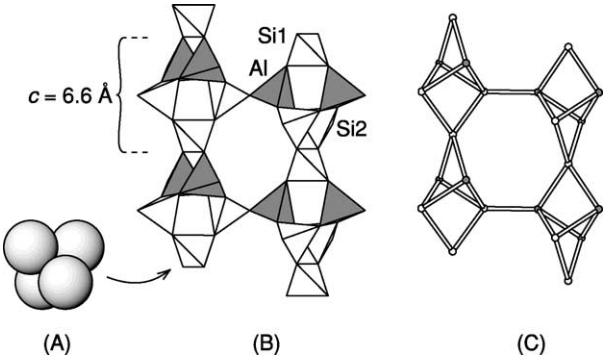


Figure 2 The framework arrangement in natrolite. (A) A representation of four oxygen anions surrounding a silicon cation (unseen). This site can be represented by a tetrahedron by reducing the oxygen spheres to points and connecting the four points with lines, as in (B). (B) Two chains of the natrolite framework. A unit of the natrolite chain consists of a Si1 tetrahedron, two Si2 tetrahedra (all three shown as white), and two Al tetrahedra (grey). The length of this unit, which is repeated continuously along the length of the chain, is 6.6 Å. Si2 tetrahedra connect with Al tetrahedra to link chains together. Because an Al–O–Al bond sequence is energetically less favourable than an Al–O–Si sequence, they tend not to occur. (C) Another way to represent the framework arrangement (topology) is to eliminate the oxygen positions and connect the tetrahedral sites with double lines. The method makes visualization of more complex arrangements easier.

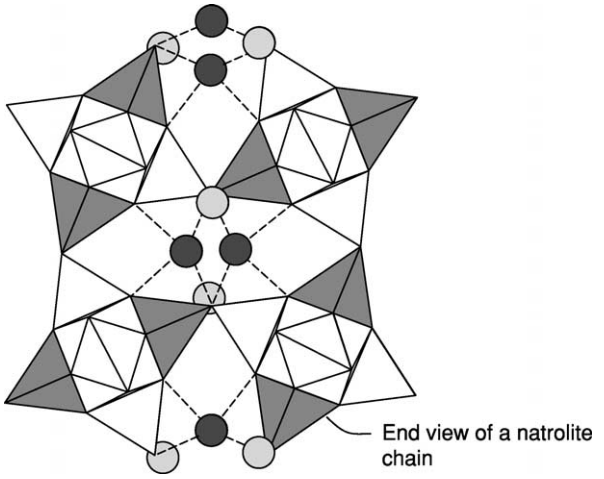


Figure 3 The bonding of sodium cations (dark grey circles) and water molecules (light grey circles) in the channels of natrolite. Sodium cations are bonded to four of the framework oxygen anions between Al and Si tetrahedra (dashed lines show bonds) and to two water molecules, which are located in spaces between linked chains (see **Figure 2**).

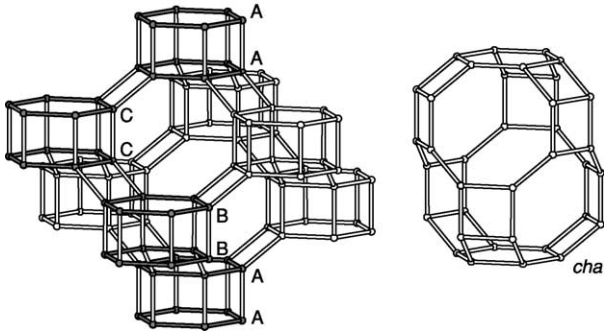


Figure 4 The arrangement of the framework of chabazite, in which only tetrahedral sites and linkages are plotted. A set of six-member rings that defines the framework is darkened, to illustrate the stacking sequence AABBCC.... Within the spiral-stair-stepped linked double six-member rings is a large space, called a *cha* cage. For clarity this *cha* cage has also been drawn outside the structure. Reproduced with permission from Deer WA, Howie RA, Wise WS, and Zussman J (2004) *Rock Forming Minerals, Volume 4B Framework Silicates: Silica Minerals, Feldspathoids and the Zeolites*. London: The Geological Society.

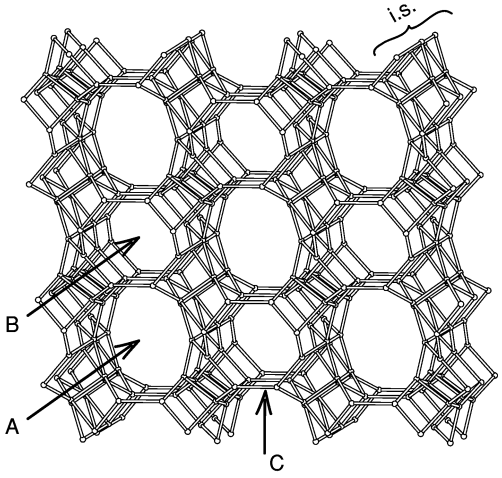


Figure 5 The arrangement of the framework of heulandite and clinoptilolite, in which only the tetrahedral sites are plotted and connected without the bridging oxygen ions. The framework consists of impermeable sheets (i.s.) that are cross-linked, forming three different types of channel, A, B, and C; A and B are almost perpendicular to the page, while C is parallel to it. Non-framework cations and water molecules occupy various sites within these channels. Reproduced with permission from Deer WA, Howie RA, Wise WS, and Zussman J (2004) *Rock Forming Minerals, Volume 4B Framework Silicates: Silica Minerals, Feldspathoids and the Zeolites*. London: The Geological Society.

to allow the passage of ions or molecules. In effect the sheets are impermeable, as shown in Figure 5. The sheets are bonded together with widely spaced oxygen bridges, between which are three different channels. There are relatively few aluminium cations in this structure, leaving the framework silicon-rich, and the distribution of the aluminium is random. Non-framework cations are located along the centres of the channels, weakly bonded to the sides and to surrounding water molecules.

Chemical Composition of Zeolites

A perusal of Table 1 shows that zeolite minerals are for the most part aluminosilicates of sodium, potassium, or calcium, with magnesium, strontium, and barium being less common. There is significant variation in composition between mineral species, but also within several of the species. Compositional variation may occur within the framework, i.e. the amount of aluminium in tetrahedral sites, and in the charge-balancing non-framework cations. One way to illustrate the range of zeolite compositions is with a plot like that shown in Figure 6. This diagram shows the wide range of silicon (or aluminium) content observed among zeolite species, as well as the variability in the non-framework cations.

Some zeolite mineral series exhibit a considerable variation in the framework composition. For example, within the chabazite structure group the silicon content ranges from high, Si/Al = 3.0, giving a framework composition of $[Al_{2.4}Si_{9.6}O_{24}]$, to low, Si/Al = 1.0, yielding $[Al_6Si_6O_{24}]$ in willhendersonite (Table 1). Other mineral series that have wide ranges of framework composition are heulandite, clinoptilolite, and phillipsite. The ratio Si/Al = 1.0 appears to be the lower limit for the amount of aluminium in any framework. More aluminium would require some Al–O–Al bond sequences, which are energetically less stable than Al–O–Si bond sequences.

Detailed structural studies of heulandite, clinoptilolite, and phillipsite indicate that the Si/Al distribution throughout the framework is random. The framework arrangements of other zeolites seem to require a fixed Si/Al ratio, with the aluminium and silicon in specific sites, i.e. an ordered arrangement. A good example is the arrangement and composition of natrolite (Figure 2). Interestingly, this is true only for well-formed crystals. Crystals that are forced to grow rapidly form small crystals with disordered structures, in which the distribution of aluminium and silicon tends to be random and the ratio of Si : Al varies from the 2 : 3 of ordered natrolite.

The ordered frameworks of some species are commonly preferential to one of these cations. For example, the sodium in analcime tightens the framework such that inclusion of calcium in the sodium sites is strongly restricted. Because only one calcium cation is required to balance two aluminium cations in the framework, they tend to distort the structure.

An important chemical property of any zeolite is its ion-exchange capacity, i.e. the extent to which external cations can be exchanged with those naturally present in the structure. High-silica zeolites (Figure 6), for example clinoptilolite, have few aluminium cations in tetrahedral sites, and therefore have few non-framework cations available for exchange. Conversely, the low-silica zeolites, such as phillipsite, have high ion-exchange capacities.

Included in the original definition of a zeolite by Cronstedt in 1756 was the release of bubbles of water vapour upon heating a crystal in a blowpipe flame. All presently recognized zeolite minerals (Table 1) contain water molecules in the channels and cages. In some, such as natrolite, the molecules are tightly bonded, whereas in others, such as clinoptilolite, the water is loosely held. Heating of the zeolite generally releases the water at temperatures of less than about 500°C. Most structures will rehydrate with exposure to a humid environment. The number of molecules listed for each zeolite is one of the least-restrictive features of their composition, because the amount held in the structure may vary with humidity.

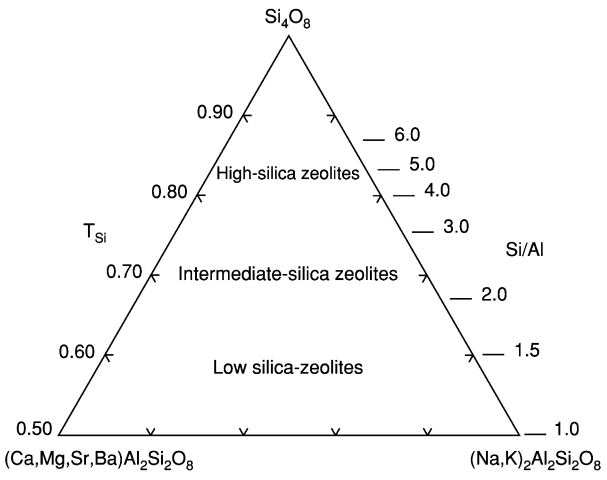


Figure 6 Compositional diagram illustrating major variations of zeolite composition. Components for the plot are: Si_4O_8 (molecular proportion [Si − 2(Ca + Mg + Sr + Ba) − 2(Na + K)]/4); $(Ca,Mg,Sr,Ba)Al_2Si_2O_8$ (molecular proportion Ca + Mg + Sr + Ba); and $(Na,K)_2Al_2Si_2O_8$ (molecular proportion (Na + K)/2). T_{Si} = Si/(Si + Al). Reproduced with permission from Deer WA, Howie RA, Wise WS, and Zussman J (2004) *Rock Forming Minerals, Volume 4B Framework Silicates: Silica Minerals, Feldspathoids and the Zeolites*. London: The Geological Society.

Occurrences of Zeolites

Zeolite minerals occur in sedimentary, metamorphic, and igneous rocks, almost exclusively as alteration products. Most form at temperatures less than 250°C and at depths less than 10 km in water-saturated environments (Figure 7). The alteration process must produce an abundance of silicon and aluminium along with alkali or alkaline earth elements. This commonly occurs where volcanic materials – lava or tuff – react with surface water or groundwater, or where plagioclase-bearing rock undergoes a reaction with pore fluids.

Zeolites in Sedimentary Rocks

The most widespread and economically important of all zeolite occurrences are in sediments and sedimentary rocks composed partially or wholly of volcanic clasts. During diagenesis (*see* **Diagenesis, Overview**) the volcanic component of these rocks reacts with pore fluids to form clay and zeolite minerals. This alteration can take place in various environments, including shallow alkaline lakes, the deep seafloor, thick terrestrial pyroclastic deposits, and sandstone beds in shallow to deep marine basins. Diagenetic reactions merge with those of burial metamorphism (Figure 7), and the distinction is a matter of the terminology used by researchers in this field. In this review, diagenesis refers to all changes of the initial sediment between deposition and the completion of lithification or cementation, and metamorphism refers to changes occurring after diagenesis, generally involving the breakdown of the first diagenetic minerals in response to increases in temperature and pressure.

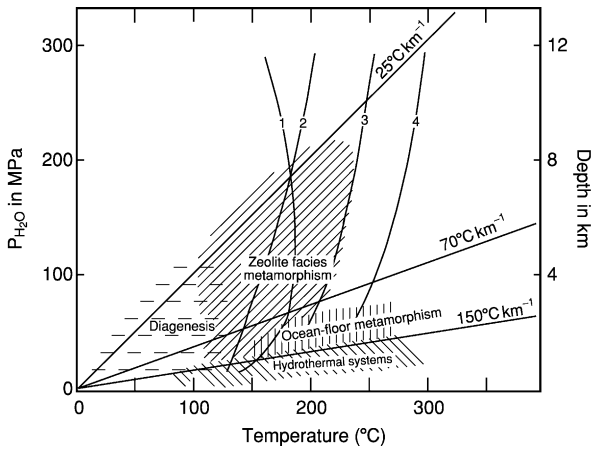


Figure 7 Pressure–temperature ranges of environments of zeolite formation. The boundary between diagenesis and metamorphism overlaps considerably more than shown here. Numbered curves are experimentally determined or calculated stability limits of selected zeolites: 1, analcime + quartz = albite + water; 2, heulandite = laumontite + quartz + water; 3, laumontite + prehnite = epidote + quartz + water at low-pressure end and laumontite + pumpellyite = epidote + chlorite + quartz + water at high-pressure end; and 4, laumontite = wairakite + water. The depth scale is based on a lithostatic gradient of 25 MPa km^{-1}. Reproduced with permission from Deer WA, Howie RA, Wise WS, and Zussman J (2004) *Rock Forming Minerals, Volume 4B Framework Silicates: Silica Minerals, Feldspathoids and the Zeolites*. London: The Geological Society.

Shallow alkaline lakes Beds of zeolite-rich rock a few centimetres to several metres thick occur in several areas of the world, such as the western USA, Serbia, and western Turkey. Clinoptilolite is the most abundant zeolite in these beds, but it may be associated with chabazite, phillipsite, erionite, and analcime. The zeolitic beds are interlayered with silt and were deposited in shallow playa lakes in closed basins. With no outlets, these lakes become saline and alkaline through extended periods of evaporation. Nearby explosive rhyolitic volcanism produced vitric ash that reacted with the alkaline water of the lake to produce the beds of zeolite-rich rock. Figure 8 shows the relationships between mineral zoning and hydrology in this type of environment.

Glass persists unaltered where pore solutions have a pH of less than about 8.5. The clinoptilolite zone may contain other zeolites, such as erionite, chabazite, and phillipsite, all of which pseudomorph glass shards. Analcime, which is abundant in the next zone inwards, replaces both pre-existing zeolites of the clinoptilolite zone and, less commonly, glass shards. Potassium feldspar, dominant in the innermost zone, replaces earlier-formed zeolites. This horizontal zonation of the zeolites and potassium feldspar along a single tuff bed is characteristic of this type of alteration, and puzzling to understand.

A kinetic model has been proposed to explain the formation of zeolites in tuffaceous rocks. The model

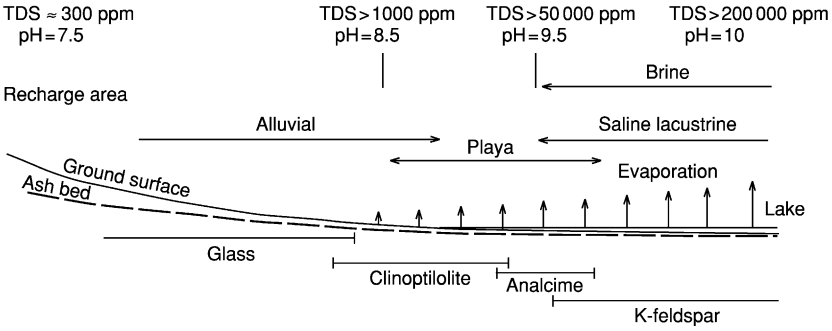


Figure 8 Relationships among hydrology, geology, and mineralogy in a playa saline lake. TDS, total dissolved solids. Reproduced with permission from Deer WA, Howie RA, Wise WS, and Zussman J (2004) *Rock Forming Minerals, Volume 4B Framework Silicates: Silica Minerals, Feldspathoids and the Zeolites*. London: The Geological Society.

incorporates the idea that glass dissolution proceeds at much greater rates than stable authigenic minerals, such as analcime and feldspar, can crystallize. The less-stable disordered alkali zeolites (chabazite, clinoptilolite, phillipsite, erionite) nucleate and crystallize readily, lowering the free energy of the system. With time, these metastable phases are replaced by stable ones, but changes in intensive parameters, such as temperature or pH, can greatly accelerate the reaction.

Deep sea sediment Zeolites, predominantly phillipsite and clinoptilolite, occur as a moderate component (5–15%) of many deep-marine sediments that include calcareous ooze and chalk, radiolarian ooze, hemipelagic mud, hyaloclastite, ash layers, or volcaniclastic sand. Zeolite-bearing sediment has been recovered from all parts of all oceans. Phillipsite tends to be common at shallow core depths, while clinoptilolite is more abundant in deeper (and older) sediment. Because phillipsite has been recovered from very near the sediment–water interface, it is believed that initial reactions of seawater with clay and/or volcanic debris occur at or just below the seafloor.

Thick terrestrial pyroclastic deposits Most terrestrial accumulations of volcanic material – lavas and pyroclastic deposits (*see* **Pyroclastics; Lava**) – are subject to alteration by meteoric water and groundwater. Initial alteration occurs as meteoric water percolates through permeable pyroclastic deposits and reacts with very-fine-grained glass, forming smectitic clay and alkaline vadose water (Figure 9).

Above the water table, fresh glass shards persist, even though the outer surface has been altered to smectite, and the very fine particles may be replaced by smectite and opal. Because the water-to-rock ratio is high (between 5 and 10), through-flowing vadose water carries away alkali and alkaline earth cations as they are released from the altering glass surface. The remaining alumina, iron, and silica readily form clay, but zeolite growth is minor. Below the water table, where fluid flow rates are much slower, the water-to-rock ratio is less than one, and zeolite crystallization occurs. At this depth clinoptilolite is an abundant replacement of glass shards and may be associated with chabazite, phillipsite, erionite, and mordenite, along with smectite and a silica phase. At somewhat greater depths, analcime generally replaces shards and pumice lapilli, and fills voids, but in some areas textural evidence suggests that analcime replaces clinoptilolite. At the deepest levels, authigenic albite and potassium feldspar replace pre-existing zeolites.

Compositions of the pyroclastic rock other than rhyolite lead to different zeolitic alteration products.

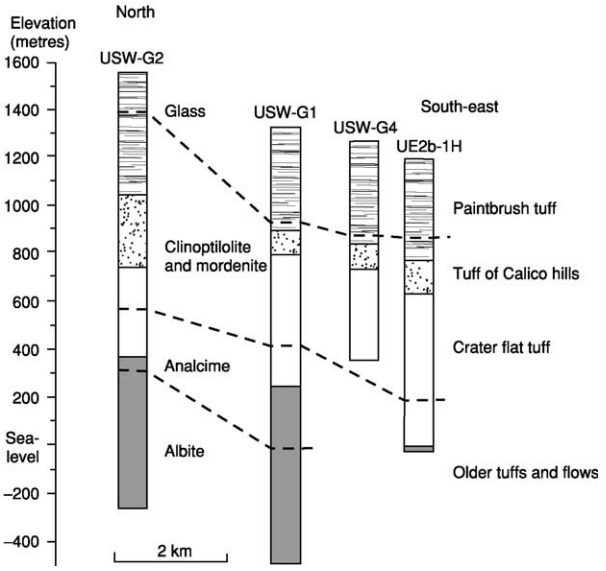


Figure 9 The relationships between diagenetic mineral facies (heavy dashed lines) and ash-flow tuff stratigraphy (patterns in columns) at the Nevada Test Site, Yucca Mountain, Nevada, USA. The listed mineral is the dominant diagenetic mineral for that facies. The boundaries are higher to the north (left side of diagram) because that is the direction of the source of eruptions and presumably the temperatures were higher. Reproduced with permission from Deer WA, Howie RA, Wise WS, and Zussman J (2004) *Rock Forming Minerals, Volume 4B Framework Silicates: Silica Minerals, Feldspathoids and the Zeolites*. London: The Geological Society.

For example, phonolitic and tephritic tuffs in Italy alter mainly to phillipsite.

Sandstone beds in shallow to deep basins The largest volumes of zeolite-bearing sedimentary rock are adjacent to volcanic arcs, where volcaniclastic debris is shed into fore-arc or back-arc basins. Depending somewhat on the composition and amount of the volcanic detritus, common diagenetic minerals are clinoptilolite, mordenite, heulandite, analcime, and laumontite. The amount of sediment accumulated in some basins approaches 10 km, such as that now exposed in the Southland Syncline, New Zealand.

In Japan, diagenetic minerals in Tertiary pyroclastic rocks are distributed in the following zones: zone I, unaltered glass; zone II, alkali-clinoptilolite and mordenite; zone III, analcime and heulandite; zone IV, laumontite; and zone V, albite. The boundaries between zones are approximately parallel to stratification. Drilling through these zones in oilfields shows that the zones are temperature-dependent, with the boundary between zones I and II at about 45°C, that between zones II and III at about 85°C, and that between zones III and IV at about 123°C. A similar distribution of zeolite phases is observed in the Southland Syncline, New Zealand, in a general sense,

but is complicated in detail, suggesting that fluid composition, as well as temperature, may play a role.

Because these sequences are so thick, the alteration of the original volcanic material extends beyond diagenesis into burial metamorphism, especially where early-formed heulandite and analcime are replaced by laumontite and albite. In fact, the zeolite and feldspar minerals of zones III, IV, and V in Japan show replacement textures, indicating early metamorphism.

Metamorphosed Rocks

In volcanic rock sequences, laumontite and analcime are common products following diagenesis and are developed through burial, contact, or ocean-floor metamorphism. Burial of sediment derived from arc-terrains, containing abundant volcanic detritus and plagioclase sand grains, produces such zeolite minerals as heulandite during diagenesis. Early minerals are replaced by higher temperature (or more stable) zeolites with deeper burial. The type example of this style of metamorphism, called burial metamorphism, is the Southland Syncline, New Zealand. At higher temperature and pressure, the zeolite phases ultimately give way to prehnite, pumpellyite, and feldspar.

The lowest-temperature zones of thermal metamorphic aureoles in volcanic terrains, lavas, and pyroclastic rocks may contain zeolite assemblages. Laumontite-bearing assemblages are common in the zeolite zone of the thermally metamorphosed tholeiitic pillow lavas and pyroclastic rocks of the Karmutsen Group, British Columbia. Here, too, the zeolite facies assemblages are replaced by the prehnite–pumpellyite facies. Similar sequences of mineral assemblages have been described in the Tanzawa Mountains, central Japan, although the geological history suggests a combination of thermal and burial metamorphism.

Metamorphism of pillow lavas and dykes near mid-ocean spreading ridges, now observed in cores from deep-sea drilling sites and in ophiolite sequences, commonly produces zeolite-bearing assemblages. This style of metamorphism is known as ocean-floor metamorphism; analcime and natrolite occur at the shallowest depths, and laumontite, scolecite, and heulandite occur at somewhat deeper levels.

Zeolites in Lava Flows

Many zeolites occur in cavities in basaltic rocks. Some notable localities are: the Tertiary lavas of eastern Iceland; County Antrim, Northern Ireland; the Faeroe Islands; the Deccan Plateau, India; the Jurassic basalt units of New South Wales, Australia; and the Triassic basalt belts in and near Paterson, New Jersey, and in Nova Scotia, Canada. This type of zeolite occurrence has been widely referred to as 'hydrothermal' in the literature. However, the zeolites in these occurrences may have originated in any of a number of different ways, including: burial diagenesis; burial metamorphism; hydrothermal alteration and metamorphism; reaction with groundwater in areas of high heat and fluid flow (high geothermal gradient); and reaction with surface water during the initial cooling of a lava flow or plug.

Studies of thick accumulations of basalt and andesite in extensional basins have documented regional low-grade metamorphism. In basins along the Andes of Peru and Chile the rocks are little deformed and the metamorphic grade increases with depth from zeolite to greenschist facies.

The effect of depth on zeolite crystallization in subaerial flood-basalt flows is demonstrated by the distribution of amygdale minerals in the basalt lavas of eastern Iceland. Zeolites or assemblages of zeolites define nearly flat-lying zones that cut across the lava stratigraphy (Figure 10). Similar regionally zoned occurrences of zeolite have been described in sequences in the Faeroe Islands, eastern Greenland, western Greenland, the Deccan Traps in western India, the North Shore Volcanic Group of Minnesota, the Parana lavas of Brazil, and Northern Ireland.

In Iceland and Greenland zeolite crystallization did not occur until after the accumulation and weak

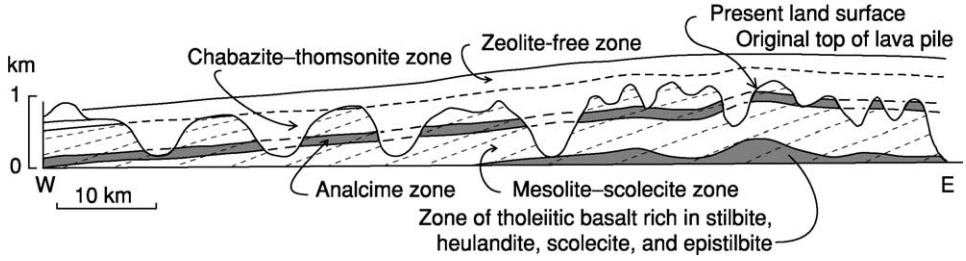


Figure 10 Diagrammatic section through the Tertiary lava pile of eastern Iceland, showing the inferred relationship between the lava stratigraphy (sloping dashed lines) and the distribution of amygdale minerals (subhorizontal bands of grey and white). Reproduced with permission from Deer WA, Howie RA, Wise WS, and Zussman J (2004) *Rock Forming Minerals, Volume 4B Framework Silicates: Silica Minerals, Feldspathoids and the Zeolites*. London: The Geological Society.

deformation of the lava pile, because the zones are parallel to the inferred original surface. The reaction of groundmass glass with interstitial groundwater results in the replacement of the glass by smectite and the crystallization of smectite, celadonite, and zeolites in open spaces. Because the water-to-rock ratio is small, the chemical composition of the glass in the host rock exerts important controls on the Si/Al ratio of the framework and on the non-framework cation content of the resulting zeolites. In the Icelandic sequences, tholeiitic basalt vesicles contain a somewhat different assemblage of zeolites from those in olivine basalt. For example, analcime occurs in olivine basalt, whereas in the same depth zone mordenite is found in tholeiitic basalt. Along with evolving zeolite assemblages, the associated phyllosilicate minerals also change with depth. In the shallowest affected rocks smectite, commonly ferroan saponite, is abundant, and this gives way to mixed-layer smectite–chlorite and ultimately to chlorite with increasing temperature. However, the range of permeability within a single flow can cause variations in the degree of crystallization and in mineral assemblages.

Hydrothermal Alteration

Hydrothermal alteration is a chemical replacement of the original minerals (or glass) in a rock by new minerals where hot water delivers reactants and removes the aqueous reaction products. Hydrothermal systems may develop alteration aureoles within sequences of volcanic rocks (*see* **Tectonics:** Hydrothermal Activity). The best-known examples are in active geothermal areas, where the alteration process is presently occurring. However, the term hydrothermal is also used to refer to the alteration caused by a fluid phase following magmatic crystallization of pegmatitic bodies.

Active geothermal systems Zeolites have been found in drill cores from several active geothermal areas, for example in Japan, Iceland, the Wairakei geothermal field (New Zealand), and Yellowstone National Park (Wyoming, USA). These systems involve warm to hot (70–150°C) water circulating through relatively large volumes of rock. Common zeolites found in this environment are mordenite, analcime, heulandite, clinoptilolite, laumontite, yugawaralite, and wairakite.

Late-stage crystallization and alteration of pegmatitic bodies Zeolites are common phases in the late stages of pegmatite crystallization. Early in the formation of a pegmatite dyke, magmatic crystallization produces the bulk of the minerals, but as the magma is depleted and the temperature decreases excess water forms a fluid phase. This fluid produces minerals at lower temperatures, many as alteration products of earlier phases. Excellent examples of hydrothermal zeolite formation in pegmatite bodies are found in the Lovozero Massif, Kola Peninsula, Russia. Here, all of the pegmatite pods, lenses, pipes, and dykes are nepheline syenite. Some analcime may be magmatic in origin but much of it, along with the very abundant natrolite, is of hydrothermal origin.

Zeolites also occur in vapour pockets late in the crystallization of granitic pegmatite dykes, mostly as a result of the alteration of feldspar either by residual magmatic fluid or by introduced water. Laumontite and stilbite are commonly precipitated along fractures or in miarolitic cavities. A well-known example is the granodiorite pegmatite dykes near San Piero in Campo on the island of Elba, Italy, where mordenite, dachiardite, stilbite, epistilbite, heulandite, and chabazite occur with pegmatite pocket minerals quartz, tourmaline, beryl, pollucite, and lepidolite.

Fractures and cavities in granitic gneiss Gash fractures formed by the semi-brittle deformation of granitic gneiss may fill with water before the rock has completely cooled, and new minerals may crystallize in the open space. These are the well-known alpine clefts, which contain chlorite, adularia, and quartz crystals that grow at the expense of wall-rock minerals. In some cases reactions continue, producing zeolites such as stilbite, heulandite, scolecite, and chabazite.

Alteration along faults An early discovery in the drill hole into the San Andreas fault zone at Cajon Pass, southern California, was the pervasive occurrence of laumontite and stilbite in shattered granodiorite and gneiss. Both minerals replace plagioclase and fill microfractures throughout the fault zone. This mineralization is attributed to groundwater circulation in fractured and frictionally heated rocks within the fault zone.

Uses of Zeolites

In those areas of the world where beds of vitric tuff or other volcanic-rich sedimentary rock have been altered to nearly pure zeolite, the deposits are suitable for economic exploitation. Over the past 40 years considerable effort has been directed towards finding uses for those zeolites that occur in relative abundance. Applications have been developed in agriculture, in wastewater treatment, in gas separation or enrichment, in building construction, as catalysts, and in energy storage from solar collectors. The most common zeolite

occurring in large quantities is clinoptilolite, so most applications involve this mineral.

The cation-exchange capacity, cation selectivity, and molecular-sieve property of bulk zeolites are being applied to various aspects of agriculture. Some of the beneficial effects of using clinoptilolite are improved growth and feed utilization, improved eggshell quality in chickens, reduced free ammonia in the rumens of sheep and cattle, and reduced exposures to ammonia and odours in feed lots. Zeolites are used in soil conditioning to improve plant productivity, increase fertilizer efficiency, and aid remediation of soils. For example, clinoptilolite-bearing tuff is used to increase the yield of wheat in Azerbaijan. Zeoponics is a new method of plant growth, using artificial soil with zeolite as a major component. As in other agricultural applications, the zeolite helps to control soil solutions.

The selectivity of clinoptilolite for certain heavy metals is the basis for many applications in water and wastewater treatment. Examples are the removal of ^{137}Cs from contaminated waters resulting from the Chernobyl accident and the removal of copper from electroplating effluents.

The molecular-sieve property of many zeolites is used to separate gases. Phillipsite has been shown to have adsorption selectivity of ammonia over methane in gases from coal gasification. In Japan activated mordenite is used to increase the oxygen in air for steel making, and in the USA chabazite is used to separate carbon dioxide from methane in gas from a landfill site in southern California.

Where zeolitic tuff is abundant it has been used as a dimension stone at least since Roman times to construct castles and cathedrals in Italy and other countries of Europe. These tuffs are also used as a constituent of lightweight aggregate, as in the production of pozzolanic cement.

Although most zeolites used as catalysts are synthetic and made for specific applications, a few natural zeolites have also been tried. Natural ferrierite has been shown to be an effective material as a Claus tail-gas catalyst, to remove elemental sulphur from gas streams, and natural mordenite is effective in the hydroisomerization of n-hexane, with results similar to those of synthetic mordenite.

Glossary

Hydrothermal alteration Alteration of a host rock by hot water percolating through fractures, generally where the water-to-rock ratio is very large.

Metastable zeolites Those zeolites that crystallize quickly but are later replaced by more stable phases (zeolites or feldspars).

Oxygen bridge The oxygen anion between two tetrahedral sites.

Tetrahedral sites The space between four tightly held oxygen anions.

See Also

Analytical Methods: Geochemical Analysis (Including X-Ray); Mineral Analysis. **Diagenesis, Overview**. **Lava**. **Metamorphic Rocks:** Facies and Zones. **Minerals:** Definition and Classification; Feldspars; Quartz. **Pyroclastics**. **Tectonics:** Hydrothermal Activity.

Further Reading

Bargar KE and Beeson MH (1985) Hydrothermal alteration in Research Drill Hole Y-3, Lower Geyser Basin, Yellowstone National Park, Wyoming. Professional Paper 1054-C. Washington, DC: US Geological Survey.

Bish DL and Ming DW (2001) *Natural Zeolites: Occurrence, Properties, Applications*. Reviews in Mineralogy and Geochemistry 45. Washington, DC: Mineralogical Society of America.

Coombs DS, Alberti A, Armbuster T, *et al.* (1997) Recommended nomenclature for zeolite minerals: report of the subcommittee on zeolites of the International Mineralogical Association, Commission on New Minerals and Mineral Names. *Canadian Mineralogist* 35: 1571–1606.

Deer WA, Howie RA, Wise WS, and Zussman J (2004) *Rock Forming Minerals, Volume 4B Framework Silicates: Silica Minerals, Feldspathoids and the Zeolites*. London: The Geological Society.

Dibble WE Jr and Tiller WA (1981) Kinetic model of zeolite paragenesis in tuffaceous sediments. *Clays and Clay Mineralogy* 29: 323–330.

Gottardi G (1989) The genesis of zeolites. *European Journal of Mineralogy* 1: 479–487.

Gottardi G and Galli E (1985) *Natural Zeolites*. Berlin: Springer-Verlag.

Hay RL (1966) *Zeolites and Zeolitic Reactions in Sedimentary Rocks*. Special Paper 85. Boulder: Geological Society of America.

Neuhoff PS, Fridriksson T, and Bird DK (2000) Zeolite parageneses in the North Atlantic Igneous Province: implications for geotectonics and groundwater quality of basaltic crust. *International Geological Review* 42: 15–44.

Sand LB and Mumpton FA (1978) *Natural Zeolites: Occurrence, Properties, Use*. Elmsford, New York State: Pergamon Press.

Sheppard RA and Gude AJ III (1973) *Zeolites and Associated Authigenic Silicate Minerals in the Tuffaceous Rocks of the Big Sandy Formation, Mohave County, Arizona*. Professional Paper 830. Washington, DC: US Geological Survey.

Sheppard RA, Gude AJ III, and Fitzpatrick JJ (1988) *Distribution, Characterization, and Genesis of Mordenite in Miocene Silicic Tuffs at Yucca Mountain, Nye County, Nevada*. Bulletin 1777. Washington, DC: US Geological Survey.

Zircons

G J H McCall, Cirencester, Gloucester, UK

Introduction

The word 'zircon' is very old; the word in Arabic, *zarqun*, is said to derive from the original form of the word in Persian *zar* ('gold') and *gun* ('colour'). Zircon is an orthosilicate mineral with the formula $ZrSiO_4$. Its hardness on Moh's scale is 7.5 and it is slowly attacked by hot concentrated sulphuric acid. Its density is 4.6–4.7. Its fracture is uneven and may be conchoidal in metamict varieties. Its streak is white. Some of the properties of zircon are given in Table 1.

Structure

Zircon is tetragonal (Figure 1), with a unit-cell structure where $a = 6.61$ Å and $c = 5.99$ Å. X-Ray diffraction analysis shows that each zircon atom is surrounded by a tetrahedral group, with four oxygen atoms surrounding a central silicon atom (Figure 2).

Chemistry

Zircon always contains an amount of hafnium, with the HfO_2/ZrO_2 ratio varying. The amount is usually minute, but certain zircons from Norway contain up to 24% of HfO_2. Hafnian zircon occurs mainly in pegmatites. Zoned crystals of members of the zircon–hafnon series from Mozambique were described by Correia Neves and colleagues in 1974; these crystals ranged from hafnian zircon to zirconian hafnon to hafnon ($HfSiO_4$). Zircon is thus a source of hafnium and well as zirconium. Phosphorous may also replace silicon. Zircons often contain rare earth elements and may be metamict (amorphous). Yttria, which is commonly present, is related to the isostructural nature of xenotime (YPO_4) to zircon. Some metamict zircons are mixtures of silica, cubic ZrO_2, and baddeleyite (monoclinic ZrO_2). Uranium and thorium are present in zircons from granites, a property that makes zircons valuable for radiometric age determinations. Chemical analyses of zircons from two geographical locations are given in Table 2.

Table 1 Properties of zircon

Property	*Description*
Refractive index	Optically positive $\omega = 1.923–1.960$ $\varepsilon = 1.961–2.015$
Birefringence	$\delta = 0.042–0.065$
Dispersion	Very strong D
Cleavage	{110} imperfect, {111} poor
Twinning	Rare, on {111}; some zoning may occur
Colour	Reddish brown, yellow, grey, green, colourless; in thin section, colourless to pale brown; transparent to translucent, vitreous to adamantine
Pleochroism	Very weak; in thick sections, may show absorption, $\omega > \varepsilon$
Unit cell	$a = 6.61$, $c = 5.99$; A type; $Z = 4$; space group $I4_1/amd$

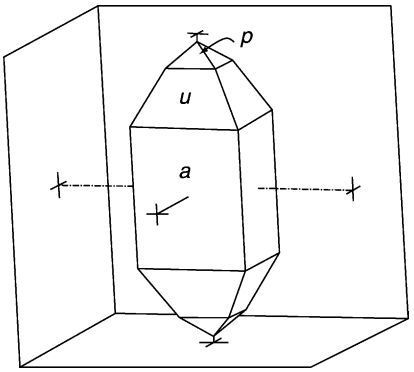


Figure 1 Sketch of a typical zircon crystal, showing crystal faces *a* (100), *p* (101), and *u* (301). Reproduced from Berry LG and Mason B (1959) *Mineralogy – Concepts, Descriptions and Determinations*. San Francisco: WH Freeman.

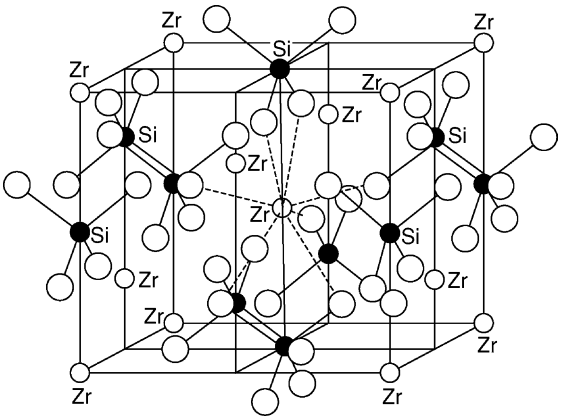


Figure 2 The structure of zircon. The eight bonds from each zircon atom to neighbouring oxygen atoms are shown only at the centre of the figure. Reproduced from Deer WA, Howie RA, and Zussman J (1962) *Rock Forming Minerals*, 2nd edn, vol. 1A. London: Longmans.

Occurrence

Zircon, a commonly occurring accessory mineral in igneous rocks, is fairly common as a detrital mineral in sediments and is also found in metamorphic rocks. It is always present in ancient shoreline concentrations of heavy minerals ('black sands'); on the coasts of Queensland–New South Wales and Western Australia (Capel, Yoganup, and Eneabba), it is typically subordinate to ilmenite and rutile.

Gemstone Zircon

Mined for a very long time for gemstone purposes in Ceylon (Sri Lanka), zircon did not really become popular until the 1920s. It is perhaps the most variable of the gemstone minerals, and its legendary mystical qualities are now known to be due to the contained radioactive elements. Grains in alluvia and heavy mineral sand deposits are commonly rounded by erosion and do not possess the full crystalline shape, short to long columnar prisms with bipyramidal terminations. Zircon has the second highest index of refraction after diamonds; the dispersion gives it an intensive lustre and cut stones have a vivid fire. Though zircon is quite hard, it is brittle and it is susceptible to damage by scratch marks. The radioactive elements in zircon do with time destroy the internal structure, producing the metamict forms, and such crystals are of lower than gemstone quality. Zircons of this type are classified as 'low quality'. Those not metamict are classified as 'high quality', and there may be varieties classified as 'intermediate'. Table 3 shows how the optical properties vary systematically with the amount of radioactivity. Heating up to 800–1000°C transforms minerals of intermediate and low quality to high quality, with colour changes, and the colour varies according to whether an oxidizing or reducing atmosphere is present during the heating. Pale blue 'starlite' and straw-yellow 'jargoons' are popular. Colourless, lucid zircon gemstones have been incorrectly termed 'Matura diamonds'. Artificially produced zircons may change colour over time, either growing pale or returning to their original hue. Gem-quality zircons never occur as very large crystals and are mined from placer deposits in Sri Lanka, Cambodia, and Thailand. Some natural gem-quality zircons are shown in Figure 3. Zirconia is an artificial imitation of diamond, a double oxide of zirconium and yttrium.

Table 2 Two Chemical analyses of zircon

Mineral	*Sample 1 (%)*[a]	*Sample 2 (%)*[b]
SiO_2	32.51	31.52
ZrO_2	67.02	64.69
HfO_2	—	1.31
TiO_2	—	0.27
Al_2O_3	0.21	0.27
Fe_2O_3	0.08	0.43
RE_2O_3[c]	0.04	0.20
MgO	0.01	0.57[d]
CaO	0.22	0.04
ThO_2	—	0/04
Nb_2O_5	—	—
UO_3	—	0.05[e]
P_2O_5	—	0.19
H_2O+	0.03	0.03
H_2O-	—	0.11
Total	100.12	100.23

[a]From North Burgess, Ontario, Canada.
[b]From South Africa.
[c]RE_2, rare earth elements.
[d]Y_2O_3.
[e]U_3O_8.

Zircon as an Economic Heavy Mineral

Zircon occurs typically in placer deposits (Table 4). The placers in which zircon occurs are commonly known as 'black sands'.

Beach Placers

The most important minerals in beach placers are cassiterite, diamond, gold, ilmenite, magnetite, monazite, rutile, xenotime, and zircon. Black sand deposits commonly contain ilmenite, monazite, rutile, xenotime, and zircon, all valuable minerals. Ilmenite may be dominant, as in Travencore, India, and Capel/Yoganup, Western Australia, but elsewhere rutile is dominant, as in the deposits of Queensland and Eneabba, Western Australia. Recent marine placers occur at different levels, owing to Pleistocene sea-level changes, and they may be many kilometres inland, as at Eneabba. The tidal zone and wave-cut terraces are optimum situations for such deposition.

Table 3 Variation in optical properties with amount of radioactivity

Classification	*D*	ω	ε	δ	*Radioactivity*
Normal zircon	4·6–4·71	1·924–1·934	1·970–1·977	0·036–0·053	Low
Intermediate	4·2–4·6	1·903–1·927	1·921–1·970	0·017–0·043	Medium
Metamict zircon	3·9–4·2	1·782–1·864	1·827–1·872	0–0·008	High

Figure 3 Gem-quality zircons. (A) Zircon crystal in host rock. (B) Hyacinth zircon from Sri Lanka. (C) Cut zircons from Thailand. (D) Cut zirconia. Reproduced with permission from Bauer J and Bouska V (1983). *A Guide in Colour to Precious and Semi-precious Stones*, pp. 108–109. London: Octopus Books.

Table 4 Classification of placers

Mode of origin	*Class*	*Alternative classification*
Accumulation *in situ* during weathering	Residual placers	Eluvial
Concentration in a moving solid medium	Eluvial placers	Colluvial
Concentration in a moving liquid medium (water)	Stream or alluvial placers, beach placers, offshore placers	Fluvial, strandline, marine placers
Concentration in a moving gaseous medium (air)	Aeolian placers	Desert aeolian, coastal aeolian

In Eastern Australia, these deposits of rutile and zircon occur over a stretch of 900 km in a strip 13 km wide, the thickness of the deposits averaging 30-40 m. Typical sites of beach placer deposition are shown in Figure 4. Deposition is by wave and shore current action. Preservation of beach deposits of heavy minerals would seem to require either progradation or sea-level fall, to shield the deposits from further erosion. The two shorelines in south-west Western Australia are shown in Figure 5, and cross-sections of the deposits there are shown in Figure 6 (Eneabba and Yoganup).

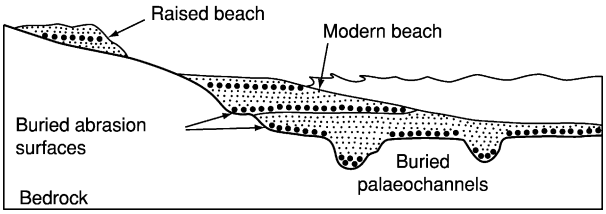


Figure 4 Typical sites of deposition of beach placer deposits. Reproduced with permission from Evans AM (1993) *Ore Geology and Industrial Minerals*, 3rd edn. Oxford: Blackwells Science Publications.

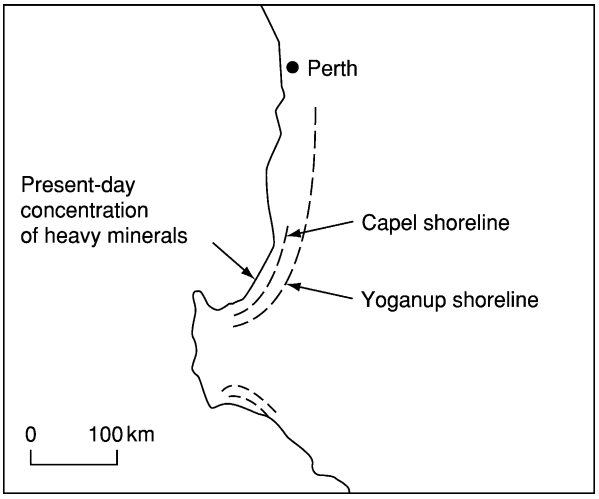


Figure 5 Sketch of the coastline in south-west Western Australia, showing the two shorelines of the younger (outer) Capel deposit and older (outer) Yoganup deposit of heavy minerals. The Eneabba deposits are about 300 km north of Perth and include an older beach deposit, about 25 km inland from the present coast, as in the case of the Yoganup deposit, and aeolian dune concentrations in the Jennings part of the Eneabba deposits just to the east of the beach placers. The Eneabba deposits are dominated by rutile whereas ilmenite is dominant in the south. Reproduced with permission from Evans AM (1993) *Ore Geology and Industrial Minerals*, 3rd edn. Oxford: Blackwells Science Publications.

Aeolian Placers

Aeolian placers are formed by reworking of the beach placers by wind. The coastal dunes that are created in this way may have a great volume and may be an important economic source for extraction of heavy minerals, even if of low grade. Three types of dunes may contain mineral deposits: foredunes, transgressive dunes that have moved inland from the beach area, and cliff-top dunes. Aeolian placers may be minor and subsidiary to large beach placer deposits, but they may be large and contain deposits in their own right. The Jennings deposit at Eneabba appears to have formed as a cliff-top dune, whereas the deposits just to the west of it are beach placer deposits. High-grade deposits of rutile and zircon occur in dunes at Richards Bay, Natal.

World Production

Australia is by far the biggest producer of zircon, with 70% of the world's supply; zircon deposits are found in other regions of the world (Figure 7), primarily in South Africa. World production of zirconium mineral concentrates is shown in Table 5.

Mining and Processing

Heavy mineral sands are extracted by dredge (Figure 8), bucket-wheel excavators or draglines, and bulldozers. The raw material is then processed, first by gravity concentrators, to remove light minerals; the concentrate is scrubbed and dried, ilmenite, leucoxene, and rutile being removed by magnetic and high-tension techniques (Figure 9). The remaining minerals are then separated by a system of spirals, gravity tables, high-tension separators, magnets, and air tables to produce zircon. The complex process requires large plants on site: the different minerals of economic interest are separated at the plant location because transport of the raw mined material would be prohibitively costly. Mining these deposits is usually done under a requirement to restore the land, and at Capel in Western Australia, the land is restored to agricultural use.

Uses

The production of zirconia (ZrO_2) for ceramic refractory purposes is the main usage of zircon, but about 10% has a variety of other uses, in ferro alloys, paints, pharmaceuticals, and abrasives. Zircon is also used in leather tanning, chemical manufacture, and in nuclear industries, the latter requiring high-purity zirconium metal and extraction of the ~2% of hafnium contained in the zircon. There are foundry uses and alloy uses, including use in welding rods and sandblasting. The hafnium contained in the zircon is used in unalloyed form in nuclear reactors. Hafnium carbide is also one of the most refractory materials known.

Zircon Analyses

Zircon has a unique status as a key mineral for placing the evolution of Earth within a clear framework of precisely measured time.

Radiometric Geochronology

Before 1896, when Henri Becquerel discovered radioactivity, estimates of the ages of minerals, rocks, and Earth were dependent on theological dogma, outright

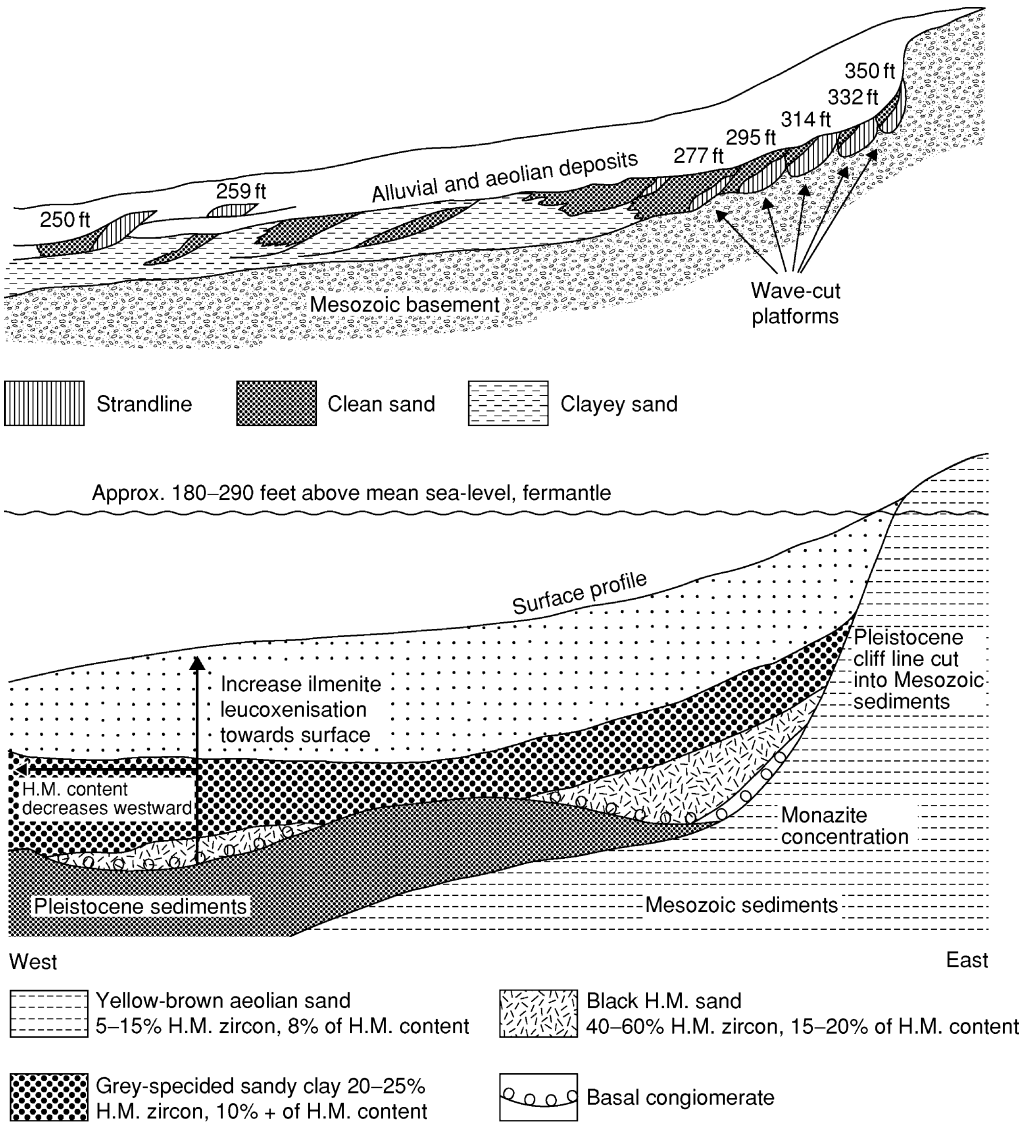


Figure 6 Cross-section of the Eneabba (top) and Yoganup (bottom) beach placer deposits. H.M., Heavy mineral. Horizontal scale, 1:120; vertical scale, 1:24. Reproduced with permission from Evans AM (1993) *Ore Geology and Industrial Minerals*, 3rd edn. Oxford: Blackwells Science Publications.

speculation, or crude physical measures (e.g., estimates based on ocean salinity). By 1903, Marie and Pierre Curie had been awarded the Nobel Prize in physics for showing that a radioactive parent element decays to various daughter elements at a constant rate, and this laid the foundations for precise physical measurement of mineral and rock ages. Despite Wilhelm Rutherford realizing in 1905 that the decay of U to He and Pb provided a powerful technique for this purpose, precise dating methods were slow to develop, awaiting the discovery of isotopes by JJ Thomson and the development of mass spectrometric techniques in the 1920s by Francis Aston and Alfred Nier. It was not until the 1940s that that a wide variety of decay schemes were tested in succession (U–Pb, K–Ar, K–Ca, Rb–Sr, Lu–Hf, Re–Os, and Sm–Nd). The basic requirements are (1) that the radioactive parent element is reasonably abundant geologically, (2) that, in the sample, the daughter element remains bound with the mineral in which it formed over time, and (3) that the latter can be distinguished reliably from any concentration of the same element that may have been originally present in the mineral or rock.

Zircon Samples

Zircon forms an ideal sample material with respect to the U–Pb decay scheme, because it is an abundant

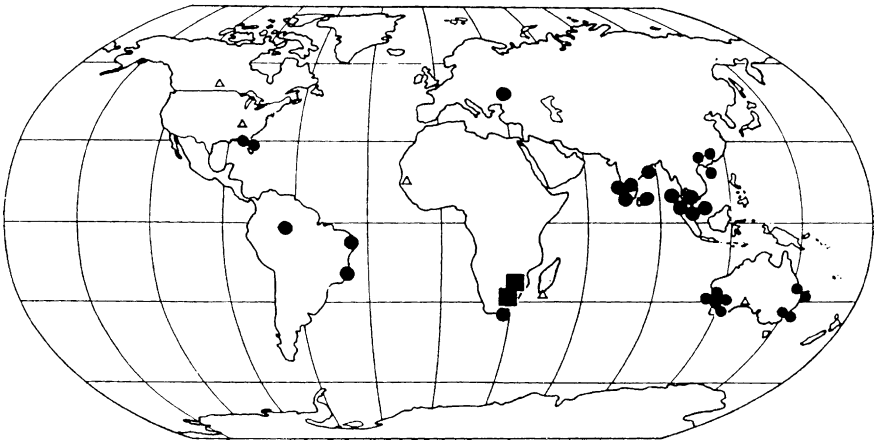


Figure 7 Sites of zirconium deposits throughout the world: (●) active zircon mines; (■) active baddeleyite mines; (△) deposits not yet worked. Reproduced with permission from Cary DD (ed.) (1994) *Industrial Minerals and Rocks*, 6th edn. Littleton, CO: Society for Mining, Metallurgy and Exploration.

Table 5 World production of zirconium mineral concentrates by year[a]

Country	*1986*	*1987*	*1988*	*1989*	*1990*
Australia	451 824	456 590	480 049	511 000	442 000
Brazil	15 116	18 131	28 029	32 970	33 000
China	15 000	15 000	15 000	15 000	15 000
India	10 000	15 000	17 000	18 000	18 000
Malaysia	12 633	17 828	25 671	18 704	4 279
South Africa, Republic of	140 000	140 000	150 000	180 000	180 000
Sri Lanka	4 000	4 000	3 000	3 000	3 000
Thailand	1 705	1 532	5 098	1 496	2 000
Former USSR	85 000	90 000	90 000	90 000	90 000
United States	W	W	117 606	118 388	102 073
Total	741 278	753 081	929 453	987 758	889 352

[a] W, Withheld; *Source: Minerals Yearbooks*, US Bureau of Mines.

accessory in a wide range of igneous, metamorphic, and epiclastic/pyroclastic sedimentary rocks. Its structure is such that it normally does, at the time of crystallization, accommodate trace quantities of U at levels sufficient for precise analysis, and the daughter Pb tends to be retained tightly. The double isotopic decay of ^{238}U to ^{236}Pb and ^{237}Pb provides a check on any loss of radiogenic Pb. Early on, in the 1950s, the method required milligram quantities (thousands of grains) of zircon, which were separated into fractions of varying magnetic susceptibility, dissolved in acid, and isotopically analysed by conventional thermal ionization mass spectrometry (TIMS) in ion-exchange columns. Improved techniques later allowed individual zircon crystals or parts of them to be analysed.

Sensitive High-Resolution Ion Microprobe Technique

In the 1980s, the sensitive high-resolution ion microprobe (SHRIMP) technique was successfully developed at the Australian National University, Canberra, based on secondary ion mass spectrometry (SIMS). In SHRIMP analysis, a focused beam of oxygen ions sputters the surface material of a mounted and polished zircon crystal into a high-voltage field, which accelerates a complex mix of ions through an electrostatic separator and then through a conventional magnet, as in the TIMS instrument. Experimentally difficult and time-consuming stages of dissolution and ion-exchange are eliminated. SHRIMP analysis has the advantages of being faster than TIMS analysis, is essentially non-destructive, and allows analysis of very small spots ($\sim$20 μm in diameter). A disadvantage is that SHRIMP analysis requires analysis by TIMS of a standard natural zircon crystal with perfect internal homogeneity and large enough for the crushed fragments to be supplied to and used by different laboratories. In practice, finding such a standard has proved difficult. In addition, the physics of the sputtering process is poorly understood and there may be an in-built bias in isotope ratios derived.

The internal complexities of zircons have been studied in great detail using SHRIMP, particularly using electron backscatter images. Age can vary significantly within a single zircon crystal. Euhedral prismatic zircons from granites commonly contain rounded cores inherited from source rocks, and high-grade orthogneisses contain patchy areas that yield

Figure 8 A dredge working on a deposit of rutile and zircon on Stradbroke Island off the Queensland coast. Such large and complex installations require capitalization running into millions of dollars. Photograph by GJH McCall.

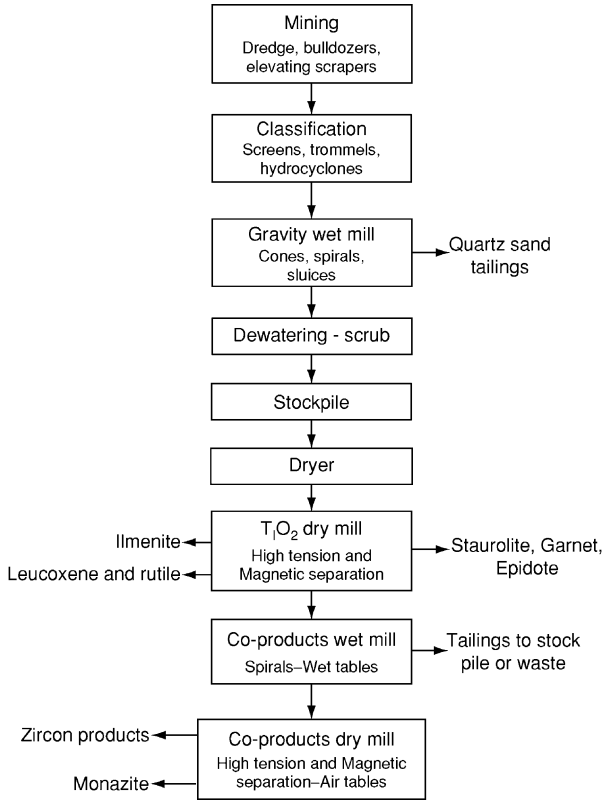


Figure 9 Simplified flow diagram showing the stages in processing a zircon-bearing placer deposit. Reproduced with permission from Cary DD (ed.) (1994) *Industrial Minerals and Rocks*, 6th edn. Littleton, CO: Society for Mining, Metallurgy and Exploration.

reset metamorphic ages, rather than the original igneous age. Such ages may have no geological meaning, but development of a combination of TIMS and SIMS methods, supported by field studies, goes beyond the

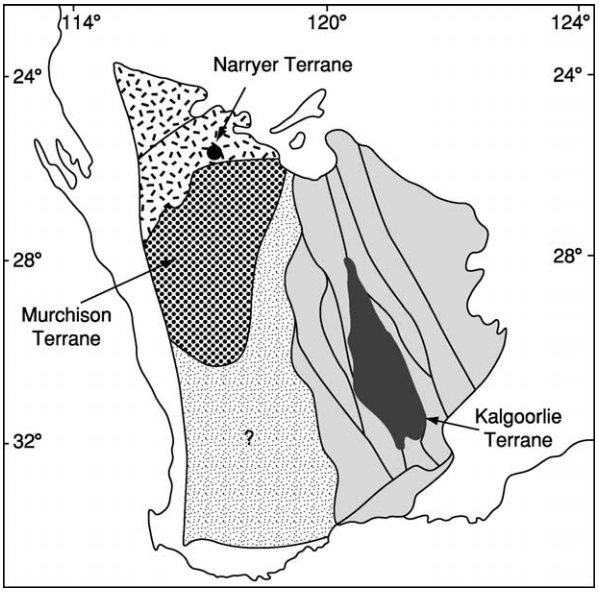


Figure 10 Location of the Narryer Gneiss Complex and Jack Hills, Western Australia.

limits of zircon geochronology based on igneous rocks, allowing complex metamorphic histories and the source rocks of sedimentary basins to be deciphered. This is a field of research that is really only just opening up.

Amelin and co-workers reported in 2001 on detrital zircons aged between 3.90 and 4.25 million years; these zircons came from the Jack Hills within the Narryer Gneiss Complex, Western Australia (Figure 10). The host rock is a metaconglomerate. Separation for SHRIMP analysis was possible for a large number of ancient grains. The experiments were carried out on SHRIMP-II at the Geological Survey of Canada, Ottawa, on 605 detrital grains from sample W74 of the metaconglomerate. Ages between 4000 and 4100 million years were found for 86 grains and 17 gave ages between 4100 and 4250 million years. Wilde and colleagues have found two zircon grains with an age of 4404 ± 8 million years using the same sample. Some grains are uniform but others show oscillatory zoning. Analysis on some spots within crystals gave no significantly different values. Some grains have visible uniform rims overgrown on zoned zircon, but, again, there is no discordance in age. A few had younger rims (3580–3650 My) and several 4100- to 4150-My-old grains have 4100- to 4150-My-old rims. These ancient zircons are not only the world's oldest dated minerals, they preserve a record of the terrestrial crust formed 100–400 My earlier than the oldest known rocks. The earliest known evidence for liquid water and a terrestrial hydrosphere comes from the 3.85-Gy-old sediments (including

banded ironstones) of south-west Greenland. The evidence derived from the Mount Narryer zircons (including oxygen isotope ratios) strongly implies that water existed on Earth at 4.4 Ga. Life may even have developed by this time, within the Hadean.

Overgrowths on Zircons

Dating of sedimentary rocks by radiometric methods has long been a difficult matter and has been achieved only to broad limits by dating igneous intrusions into them or concordant lava flows or tuffs close to the sedimentary stratum concerned. Authigenic overgrowths on zircon, a mineral that is formed at high temperature and is insoluble in groundwater, have long attracted attention. They are composed of xenotime, essentially yttrium phosphate (Figure 11). Such overgrowths occur in the Millstone Grit (Carboniferous, Namurian) of northern England. Though diagenetic glauconite has been radiometrically dated, it suffers from the defect that it is prone to postdepositional alteration. The SHRIMP technique has recently been applied to such xenotime outgrowths (xenotime contains traces of radioactive elements). The outgrowths have been shown to precede all other authigenic minerals such as quartz, and are thus thought to be very early diagenetic deposits. The xenotime is precipitated from seawater with other biogenic phosphates shortly after burial, in a zone of high bacterial activity and sulphate reduction. Xenotime outgrowths thus provide a method of obtaining what is virtually the age of deposition of a sedimentary rock, by radiometric methods. The method can be applied to rocks of any geological age, to date strata directly. This new method is especially important in the case of sedimentary rocks devoid of fossils.

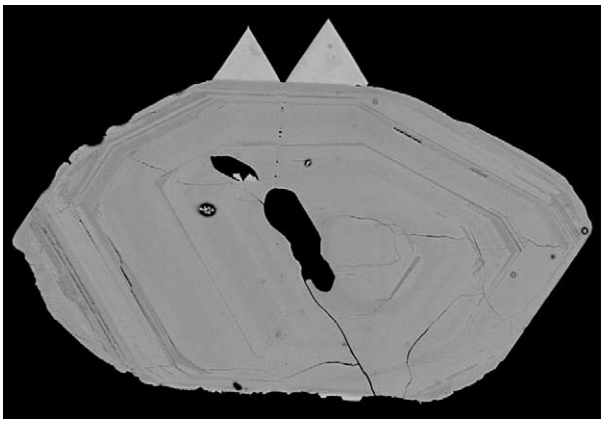


Figure 11 Two xenotime outgrowths on a zircon crystal. Reproduced from Glover J (2003) *Geological Journeys from Artifacts to Zircon*, pp. 105–106. Perth: Geological Society of Australia, Western Australian Division.

See Also

Analytical Methods: Geochronological Techniques. **Mineral Deposits and Their Genesis**. **Sedimentary Environments:** Alluvial Fans, Alluvial Sediments and Settings; Shoreline and Shoreface Deposits.

Further Reading

Amelin Y, Stern R, Wiechert U, Davis D, Lee D, Halliday A, and Pidgeon R (2001) Combined U–Pb, trace element, oxygen, Zr and Lu–Hf isotopic systematics of 3.90–4.25 Ga detrital zircons from Jack Hills, Western Australia. In: Cassidy KF, Dunphy JM, and Van Kranendonk MJ (eds.) *4th International Archaean Symposium – Extended Abstracts AGSO – Geoscience Australia Record 2001/37*, pp. 33–34. Perth, WA: Geoscience Australia.

Bauer J and Bouska V (1983) *A Guide in Colour to Precious and Semi-precious Stones*, pp. 108–109. London: Octopus Books.

Berry LG and Mason B (1959) *Mineralogy – Concepts, Descriptions and Determinations*. San Francisco: WH Freeman.

Cary, DD (ed.) (1994) Industrial Minerals and Rocks, 6th edn. Littleton, CO: Society for Mining, Metallurgy and Exploration.

Correia Neves JM, Lopez Nunes JE, and Sahama ThG (1974) High hafnium members of the zircon–hafnon series from the granite pegmatites of Zambezia, Mozambique. *Contributions to Mineralogy and Petrology* 48: 73–80.

Deer WA, Howie RA, and Zussman J (1982) *Rock Forming Minerals*, 2nd edn, vol. 1A, pp. 418–442. London: Longmans.

Evans AM (1993) *Ore Geology and Industrial Minerals*, 3rd edn. Oxford: Blackwells Science Publications.

Glover J (2003) *Geological Journeys from Artifacts to Zircon*, pp. 105–106. Perth: Geological Society of Australia, Western Australian Division.

Harben PW and Bates BC (1984) *Geology of the Nonmetallics*. New York: Metal Bulletin Inc.

Mitchell RS (1973) Metamict minerals: a review. *Mineral Record* 4: 177–182.

Wilde SA, Valley JW, Peck WH, and Graham CM (2001) Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Ga ago. *Nature* 409: 175–178.

Wilde SA, Valley JW, Peck WH, and Graham CM (2001) Evidence for the early growth of continents and oceans from >4 Ga detrital zircons. In: Cassidy KF, Dunphy JM, and Van Kranendonk MJ (eds.) *4th International Archaean Symposium – Extended Abstracts* AGSO – *Geoscience Australia Record 2001/37*, pp. 106–108. Perth, WA: Geoscience Australia.

MINING GEOLOGY

Contents

Exploration Boreholes

M Vaněček, Charles University Prague, Prague, Czech Republic

Introduction

Exploration boreholes are the primary tools used in the preliminary investigation of coal and industrial mineral deposits. Complex ore deposits are studied by means of boreholes in the initial stages of prospecting and exploration. The most frequently used technique used is core drilling. The reliability of the results of exploration drilling depends on core recovery, which is influenced by choice of method and equipment (e.g., type of flushing, use of special core barrels, bit design). Correct interpretation depends also on correct assessment of the borehole inclination.

Borehole Drilling

The drilling of boreholes for mineral exploration is practised with the aim to locate mineral deposits, to define their extent and reserves, to determine hydrogeological and engineering geological conditions, and to take samples for laboratory investigations (chemical assays, dressing characteristics, etc.). Drilling intercepts yield satisfactory data in the case of mineral deposits that are not very complex, such as sedimentary deposits and other deposits forming more or less homogeneous mineral bodies. Exploration of homogeneous mineral bodies by means of drilling is usually made in square or other regularly shaped grids. If any feature of the mineral body has an anisotropic distribution, the exploration grid should be rectangular, with shorter distances between boreholes orientated in the direction of the changing feature. Where ore deposits dip, the exploratory boreholes are arranged in profiles perpendicular to the strike of the mineral body. The entire ore deposit, however, cannot be explored in detail without complete mining operations.

Rotary core drilling is the most commonly used drilling technique for creating exploration boreholes. Drilling machines rotate an annular bit to cut a narrow kerf around a central core and thus obtain unaltered samples of the drilled rocks. The core is cut by a tungsten carbide or diamond-studded drill bit and is collected in the core barrel (**Figures 1** and **2**). When the depth of the ore body is relatively shallow, the core barrel containing the cut core is retrieved by pulling the entire length of the drill stem. If the target interval is deep and a considerable number of drill

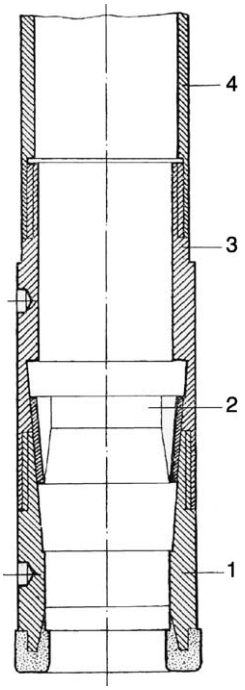


Figure 1 Conventional core drilling string. 1, Diamond bit; 2, core breaker; 3, core breaker tube; 4, core barrel.

Figure 2 Examples of diamond drill bits for exploration drilling. Photograph by Atlas Copco.

Figure 3 Underground exploration drilling (Simba H1354 for Echo Bay, Canada). Photograph by Atlas Copco.

stem rods are involved, then a wire–line system is more feasible. In this case, double-tube core barrels are used; the inner barrel is removed from the bottom of the borehole by means of wire and is thus retrieved without having to bring the drill stem rods to the surface. Use of the wire–line system is limited, depending on the hardness of the rocks being drilled. If the rock is very hard, drill bits may be destroyed in a short time and the entire drill stem will have to be retrieved frequently.

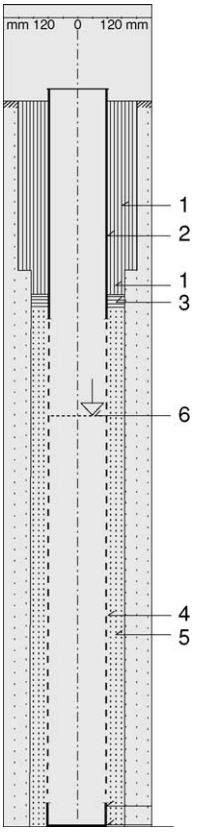


Figure 4 Detail of the hydrogeological borehole casing. 1, Cementation; 2, full casing; 3, sand; 4, perforated casing; 5, quartz pebbles (graining, 4–8 mm); 6, underground water level detected.

Special drilling machines have been designed for drilling exploration boreholes from locations below the surface (Figure 3). In comparison with surface-drilled boreholes, which may extend for some hundreds of metres, the length of subsurface-drilled boreholes usually does not exceed 50–100 m. These boreholes are drilled, often from one location, with different inclinations, from vertical to horizontal. Geological cross-sections are then based on results of fans of boreholes.

Assessment of mining conditions for mineral deposit development or extraction sometimes requires special hydrogeological approaches to exploration drilling. In these situations, the borehole casing is perforated at the underground water horizons (see Figure 4). The boreholes can be drilled using any of non-coring techniques (rotary drilling, percussion drilling, rotary–percussion drilling, etc.).

Core Recovery

One of the main problems encountered in exploration drilling is core recovery. There is a general expansion

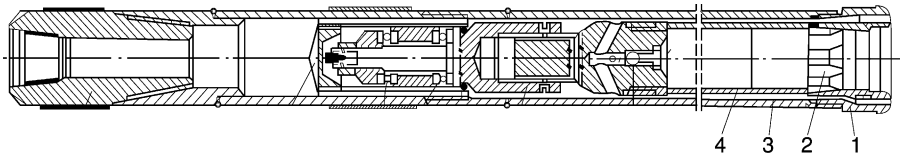


Figure 5 Double-tube core barrel for exploration of coal deposits. 1, Diamond bit; 2, core breaker; 3, outer tube of the core barrel; 4, inner tube of the core barrel.

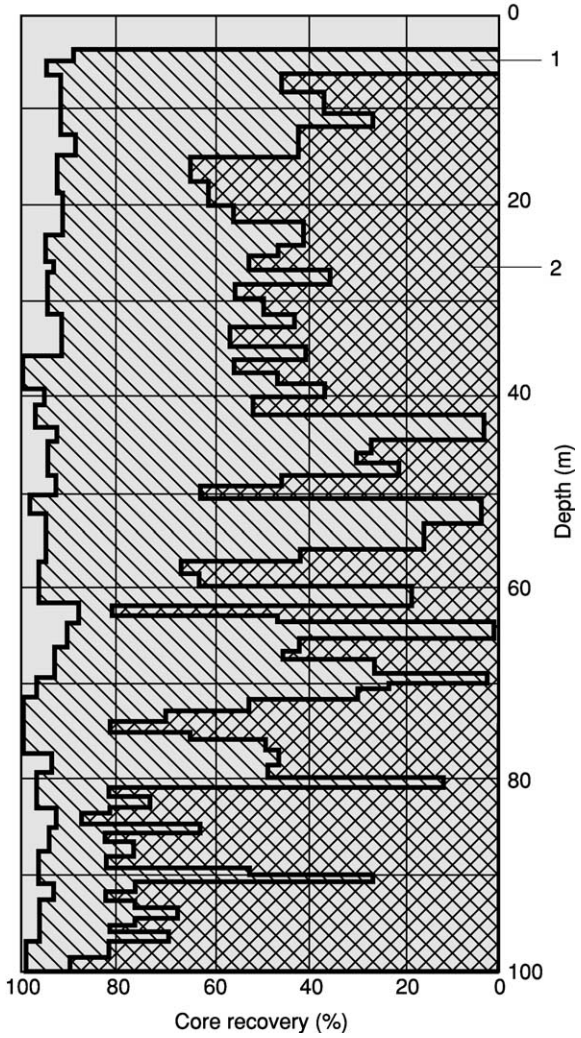


Figure 6 Comparison of the core recovery from two boreholes drilled in identical geological conditions. 1, Drilled using bentonite mud flushing; 2, drilled using water flushing. Modified with permission from Tkaný Z, Juránek O, and Brych J (1966) *Deep Diamond Drilling* (in Czech). Prague: Technical Literature Publ. House (SNTL).

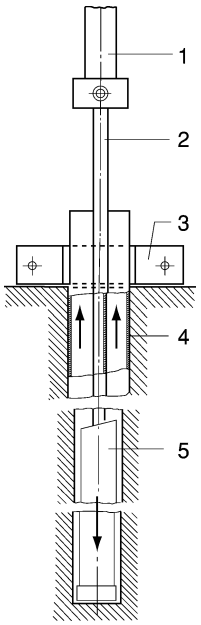


Figure 7 Detail of direct flushing tubular drill rod. 1, Drill head; 2, drill stem; 3, casing clamp; 4, casing; 5, core drilling string.

of core sample length on removal of the sample from the core barrel, related to fracturing of rocks during the drilling process. If soft rock intercalations exist in the borehole location, the core sample is likely to be lost. Because it may not be possible to determine the presence or position of soft rock prior to drilling, use of double-tube (Figure 5) or even triple-tube core barrels may increase the chances of obtaining 100% core sample recovery.

Core recovery may also depend on the type of flushing medium used. The effect of using water versus bentonite mud as the circulation fluid is shown in Figure 6. Circulation media (water, clay mud, or air) may be pumped into drill holes through the inner part of a tubular drill rod (Figure 7); in counterflush drilling, media are injected into drill holes through the annulus between the wall casing and the drilling pipe or, more often, between the inner and outer wall pipes (see Figure 8). Counterflush drilling, also called reverse-circulation drilling, has the distinct advantage over normal rotary drilling of providing relatively complete and uncontaminated samples. Though the method is considerably slower and more expensive than normal rotary drilling is, it is used increasingly in contemporary explorations of mineral deposits. Costs incurred in reverse-circulation drilling depend on geological formation hardness.

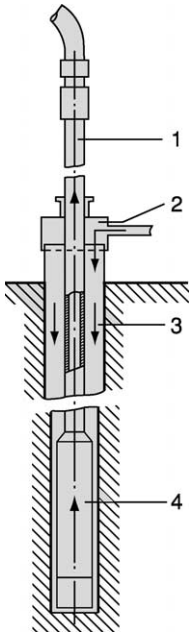


Figure 8 Detail of reverse circulation flushing drill. 1, Drill stem; 2, flush head; 3, casing; 4, core drilling string.

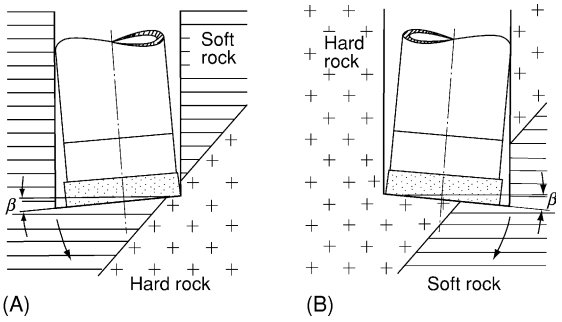


Figure 9 Change in the borehole inclination at the boundary between two layers. (A) Transition from soft to hard rocks. (B) Transition from hard to soft rocks.

Correct interpretation of exploration drilling results requires knowledge of the course of the drilled boreholes, which also relates to rock hardness. For example, if a drill bit reaches a boundary between two layers with different levels of hardness, the inclination of the borehole may change (see Figure 9) and the hole may become crooked and off-course.

Drill Bits

Drill bits of different geometry and design are used for core drilling. Surface-set bits contain a single layer of big diamonds, layered bits are made of diamonds of medium size and small size, and impregnated bits have diamond fragments distributed irregularly throughout the copper matrix of the crown. Core bits with tungsten carbide are made either with tungsten carbide inserts or with cutting edges hardened by welded-on grains of tungsten carbide. Non-coring diamond bits may be used when a core is not desired. The lack of core may be compensated by geophysical logging (recording changes in lithology, resistivity, temperature, porosity, hole diameter, etc., using special probes).

See Also

Engineering Geology: Site and Ground Investigation; Ground Water Monitoring at Solid Waste Landfills. **Geological Maps and Their Interpretation**. **Mining Geology:** Exploration; Mineral Reserves; Hydrothermal Ores; Magmatic Ores.

Further Reading

Anonymous (1963) *Symposium on Exploration Drilling*. Golden: Colorado School of Mines.

Bender F (ed.) (1986) *Angewandte Geowissenschaften*, vol. IV. Stuttgart: Enke Verlag.

Clark GB (1982) *Principles of Rock Drilling and Bit Wear. Colorado School of Mines Quarterly*, vol. 77(1 and 2). Golden: Colorado School of Mines.

Compton RR (1962) *Manual of Field Geology*. New York: Wiley.

Davenport B (1992) *Horizontal and Vertical Drilling*. New York: Osborne-McGraw-Hill.

Finkl CW Jr. (ed.) (1988) *Encyclopedia of Field and General Geology*, vol. XXV. New York: Van Nostrand Reinhold.

Kennedy JL (1983) *Fundamentals of Drilling: Technology and Economics*. Tulsa: PenWell Books.

Kužvart M and Boehmer M (1986) *Prospecting and Exploration of Mineral Deposits*, 2nd edn. Prague: Academia.

Lahee FH (1961) *Field Geology*, 6th edn. New York: McGraw-Hill.

Tkaný Z, Juránek O, and Brych J (1966) *Deep Diamond Core Drilling* (in Czech). Prague: Technical Literature Publ. House (SNTL).

Exploration

N C White, Brisbane, QLD, Australia

Introduction

Mineral exploration aims to discover deposits of minerals and rocks that can be used to meet the resource needs of society. It encompasses the search for industrial raw materials (e.g., clay, limestone, sulphur, salts, and fertilizer minerals and rocks), ores from which metals are extracted (e.g., iron, copper, and zinc ores), and gemstones (diamonds, sapphires, and opals), and includes the search for solid fuels (coal, oil shale, and uranium) but not liquid or gaseous fuels (petroleum and natural gas). Mineral exploration can be as basic as prospecting, using elementary techniques such as panning for gold, or it can be very sophisticated, involving the use of complex technology for data gathering and interpretation. This article considers our need for mineral resources, why different groups explore for minerals and the strategies they adopt and the tools they employ, what success and failure mean in the industry, the role of governments, and the future of exploration.

Our Need for Minerals

Human beings have always been dependent on natural materials to satisfy the needs of everday life. All these materials must come from Earth, either grown as plants or animals or extracted as rocks, minerals, or fossil fuels. Some of the rocks and minerals that are important in everyday life are readily available, but many are not, and we must search for them: that search is what exploration is about.

The importance of rocks and minerals to humankind is conveyed by terms such as 'stone age', 'bronze age', 'iron age'; the emphasis on stages in the evolution of technology and civilization is related to our mastery of different rocks and our understanding of the metals that can be extracted from them. In modern civilization, there is amazing diversity in the rocks and minerals we use and in the variety of ways in which we use them. Some mineral resources are used directly, with little processing apart from shaping, polishing, crushing, or sizing (e.g., rocks used for building, aggregate used for road-making, talc used in talcum powder, or gemstones used in jewellery). Many mineral resources are consumed in industrial processes that transform them into new products (e.g., clays used to make bricks or ceramics, silica sand used for glass-making, limestone used to make lime or cement, or phosphate rock used for superphosphate fertilizer). The large group of rocks we call ores (e.g., iron, copper, and gold ores) are important because of the metals they contain. Other rocks and minerals are used in industrial processes because they aid the processing of other materials, not because anything will be extracted from them (e.g., many rocks and minerals are used as flux in smelting).

Who Explores for Minerals, and Why?

The motivation for mineral exploration is the same as for many other human activities: either we need the materials we seek or, by satisfying other people's needs, we can earn money. Though exploration is necessary to maintain the supply of essential raw materials, individuals or corporations undertake exploration for a variety of reasons, but underlying them all is the belief that there will be a good commercial return from what is discovered. The commercial return may come from selling the deposits that are discovered, from setting up a company based on the discovery and trading the company's shares, from starting a mine and selling the products, or from using the products as feed in some downstream business based on the commodity, and profiting from sales of those new products. Companies of different sizes have different objectives and reasons for exploring, and target different commodities and styles of mineralization.

Prospectors, Small Syndicates, and Small Companies with No Mines

If the small-scale exploration group plan to mine what they discover, commodities such as gold, gemstones, and tin are favoured, because processing of the ore is simple, capital costs are low, and the unit value of the commodity produced is high, so small-scale operations may be viable. Work occurs in prospective areas that are known, typically close to where the group is based. If the small-scale group plans to sell the prospects they discover, whatever the market prefers at the time is favoured, because those targets are readily marketable. Work is undertaken in prospective areas that are familiar to them, or where their financial resources allow operations.

Small Exploration Companies with Mines or Mineral-Based Businesses

If small groups that own mines or businesses plan to use the product in their business, they will explore

for the raw materials they need for their current or planned operations. This will typically take place near current or planned operations to minimize transport costs. If the plan is to produce minerals for sale, small-scale owners will explore for commodities for which they believe they have the technical skills (mining, processing, marketing) to handle, and that can be sold profitably. Typically, work is undertaken in known regions, but may expand to other regions.

Large Mining and Exploration Companies

Some large companies focus on specific commodities or metals for which they perceive they have a technical and operational advantage, or for which wish to appeal to investors wanting commodity focus (e.g., Barrick and Newmont for gold, De Beers for diamonds, Cominco for zinc, Alcoa for aluminium, and INCO for nickel). They explore where they believe they have the greatest chance of success, based on geology and legal, commercial, and security criteria. Other large companies focus on any commodities meeting their commercial criteria: typically these will be commodities with very large global markets, no barriers to entry, low production costs, and major cash flows. These companies explore where they believe they have the greatest chance of economic success, based on geology and legal, commercial, and security criteria.

Mega-Mining Houses

This small group of very large mining houses has emerged only recently as a result of mergers and acquisitions (e.g., Anglo American, BHP Billiton, and Rio Tinto). For these corporations, exploration alone can no longer deliver their growth objectives because very few individual discoveries can meet their commercial criteria, and because explorers cannot reliably predict when they will make their next economically viable discovery. Depending on the level of competition, these companies may reduce their exploration expenditure and aim instead to acquire major discoveries made by successful smaller companies. If they acquire a small company that has grown by exploration success, in many instances they discard the successful exploration group.

Strategies for Exploration

The strategies and tools applied to exploration vary with the size of the group's exploration budget and its technical capabilities. Prospectors generally work as individuals or small syndicates. Corporations employ specialist explorers, mostly geologists, but include in-house specialist geochemists and geophysicists where there is sufficient demand for their services. Large exploration groups may comprise several hundred people, with technical specialists including geologists, geochemists, and geophysicists, and a wide array of technical support staff.

The sequence typically followed by explorers is summarized in Figure 1. The first major strategic issue they face is what to explore for and where to explore for it. There are many possible targets and commodities for which to search, and the choice can be as broad as 'anything we can make money from' or as narrow as one specialist commodity. For companies planning to use minerals in their own operations, the choice of commodities is already decided by what they need. For other explorers, the choice of targets for exploration can be driven by opportunity or by strategy. In either case, the main criteria are the perceived likelihood of success and the reward for success.

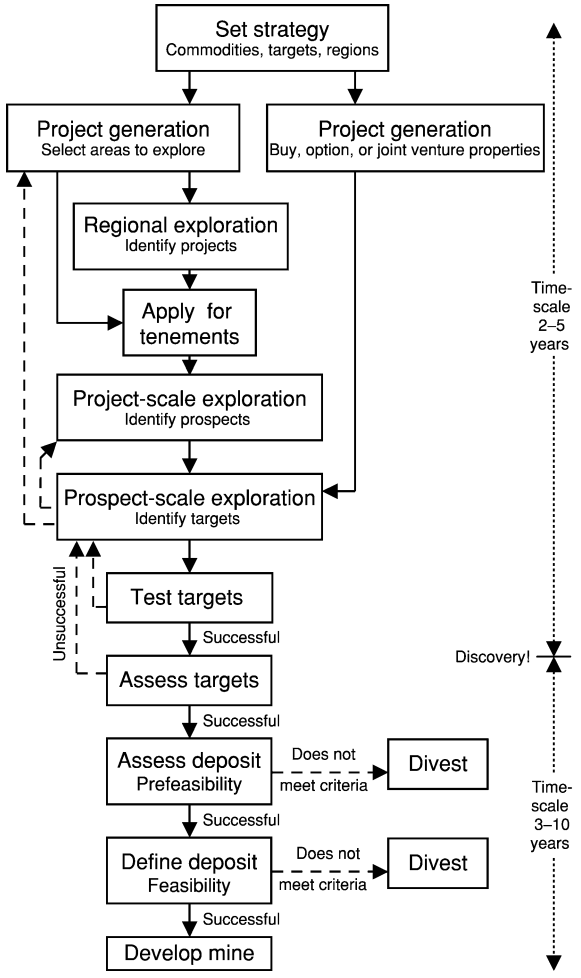


Figure 1 The sequence of stages from mineral exploration to mining.

Selecting commodities and targets is probably most crucial for large companies allocating substantial resources to exploration. They have the opportunity to explore for whatever they choose, and few restrictions on where they explore. However, because they spend a lot of money on exploration, their shareholders expect to see substantial returns for the investment. For a very large company, a small discovery that might seem to be a major success for a small company has no discernible impact on overall profitability or net capital worth. To justify large expenditures, large corporations need to focus on finding deposits that are very large and that can have a significant impact on profitability.

Once the exploration targets have been selected, the next decision is where to explore: locally, regionally, nationally, or globally. For small players, that decision is largely dictated by where they can afford to operate. It is easy to spend a lot of money on travel before any work is done, so small exploration groups typically choose a limited area of operations. To increase their chances of success, every exploration group, regardless of size, must balance between trying to be in all the best places and having as many good prospects as possible against the risk of spreading their effort too thinly and spending too much of their budget on access and support costs. The response of different explorers to these issues varies widely. At the peak of the exploration boom of the late 1990s, many exploration groups spread their activities thinly around the world, while others focused all their resources into a few priority regions. There is no right way, and either strategy can lead to success.

Even when decisions such as what to explore for and where to explore have been decided, there remain many important issues. Prospective regions may be very large, so the exploration group needs to select a region of manageable size on which to focus attention. The most important factor in making a discovery is favourable geology, but because the aim of exploration is economic success, not merely technical success, many other factors come into play. These factors include having secure title over the land to be explored, having access to the land, and having safe operating conditions, favourable legal and taxation laws, certainty of permission to progress from exploration to mining, availability of infrastructure and skilled workers, and many other issues. Previous exploration history must also be considered, because there is no point in exploring an area where previous work has been thorough, and new explorers can only repeat what has been done before. Availability of infrastructure has greatest impact for bulk commodities or industrial minerals, for which transport costs may be a substantial proportion of the sale price obtained, and the least impact for small-bulk, high-value commodities such as gemstones or precious metals, for which the product can easily be flown to market in a small plane.

Once a region has been selected for exploration, there may be opportunities to gain access to advanced projects by entering agreements with explorers already active in the area. Possible arrangements include outright purchase of an established project, option agreements (these involve progressive payments during exploration, with a major payment or share in profits or production if exploration is successful), and joint ventures involving working together to share costs and benefits. If arrangements of this kind are not possible, or the best opportunities are no better than can be obtained alone, then the best option is to explore alone.

This then raises further questions. Should the company focus its attention on known deposits and prospects, or get as close as possible to them (known as brownfields exploration)? Or should it search in other areas with apparently similar geology, but lacking known deposits (known as greenfields exploration)? The relative merits of the two approaches are a contentious issue. In brownfields exploration, it is easier to justify work and expenditures on the basis of being close to a known deposit, but how likely is it that a new deposit will be found close to an old one? With greenfields exploration, the apparent risk is higher because it involves stepping out beyond what is known, in the hope that there are new, unknown deposits to be found. Success is possible using either approach. One thing that is certain, however, is that brownfields opportunities ultimately are limited: new deposits cannot continue to be found by endlessly revisiting the same areas.

A major goal of exploration is to get to the point of testing a potential orebody as quickly and as cheaply as possible. Identifying a volume of rock that potentially hosts an orebody (the target) is an important step along that path. Identification of the target may result from the evaluation of previous work or application of a new exploration technique on a brownfields site, or from a systematic exploration program on a greenfields site. Most exploration starts with identification of a broad target area that has the potential to contain an orebody: the challenge is then to narrow the focus to a specific target that can be tested, hopefully leading to discovery. In systematic exploration, the starting region might be as large as hundreds of thousands of square kilometres, whereas the target usually has an area of much less than 1 km^2 (commonly 0.003 km^2 or less). Explorers can apply a wide array of tools and exploration methods to narrow down the area of search and define

targets for drilling. The skill in exploration lies in how those tools are used to discover an orebody.

Exploration Tools

The wide variety of exploration techniques available for exploration can be broadly grouped into three types: geological, geochemical, and geophysical. The selection of the technique or combination of techniques used varies with the target sought, the area to be explored, the geological conditions, the stage of exploration, the weathering regime, and factors such as location, topography, vegetation cover, climate, and social and cultural issues. Some techniques can be applied over large regions, whereas others can be applied only over small areas. Some require detailed ground access, whereas others can be flown by aircraft and require minimal ground access. The following discussion focuses mostly on exploration for metallic ore deposits, because for them, discovery of an economic tonnage and grade is typically the crucial step. For many bulk commodities (especially most industrial minerals), discovery is important, but the ultimate value of the find relates less to grade and more to marketability, which involves many different characteristics, such as physical properties, amenability for the ultimate application, and presence or absence of deleterious impurities. Thus, the relative importance of the early search and discovery stages, and the later testing and evaluation stages, changes with the commodity sought.

Geological Techniques

If the geological situation in which mineralization is likely to occur can be recognized by observation, then geological mapping of rock types, stratigraphy, or structure can be used (Table 1; Figure 2). Work on very prospective areas may involve more detailed studies, including petrography, fluid inclusion studies, or alteration mineralogy. Geological techniques are applicable only where geological observations can be made (i.e., where there is sufficient outcrop or where appropriate samples can be collected) or where geological inferences can be made indirectly (from aerial geophysical or spectral surveys, aerial photographs, satellite images, etc.).

Geochemical Techniques

Geochemical techniques involve the chemical analysis of geological materials such as stream sediment, soil,

Table 1 Main geological tools in mineral exploration

Tool	*Basis*	*Coverage*	*Applicability*	*Target*
Geological mapping	Systematic mapping of surface rock types, structures, hydrothermal alteration, etc.	Regional scale available globally; detailed coverage very limited	Requires reasonable distribution of exposure of bedrock, though other tools can be used to fill in missing information	An important tool in exploration for all deposit types; provides the basis for interpretation of all other datasets
Aerial photograph interpretation	A stereoscopic view of Earth's surface is provided by overlapping photographs; lineaments, colour, structures, weathering patterns, and vegetation can be used to generate a geological interpretation	Global coverage available with variable resolution and quality	Applicable if the surface is visible (no cloud cover); best in areas lacking thick vegetation cover	An important aid to geological mapping; can be used to locate orebodies that are exposed at the surface and show distinctive features
Satellite image interpretation	Satellite-based scanners generate detailed images of Earth's surface	Global	Available everywhere, but application varies with density of buildings and agriculture, vegetation, rainfall, weathering, etc. Most useful in arid, well-exposed regions	An important aid to geological mapping; can seldom be used to locate orebodies directly, but in favourable conditions in remote areas, can highlight alteration zones
Spectral data interpretation	A wide variety of spectral information is now available from satellites, aircraft, and ground surveys	Global for satellite systems, large areas from aircraft, small areas with ground instruments	Applicability varies with density of buildings and agriculture, vegetation, rainfall, weathering, etc. Most useful in arid, well-exposed regions	A useful aid to geological mapping; in favourable conditions, can locate hydrothermal alteration related to mineralization, or, rarely, large mineralized areas

Figure 2 Geological mapping in the Farallon Negro district, Argentina. All rocks show intense hydrothermal alteration related to a major porphyry copper–molybdenum–gold deposit called Agua Rica.

rock, and, in some cases, water, vegetation, or air (Table 2; Figure 3). Analytical results ('anomalies') may suggest prospective areas. Geochemical techniques are applicable only where the rock that hosts the ore, or material derived directly from it, is accessible. Sampling of overlying transported material (e.g., glacial debris) is not effective unless the source of the transported material is known, or unless chemical elements related to mineralization have migrated up through the transported cover. In areas of shallow cover, drilling may be used to collect samples from beneath the cover. Proper orientation surveys are essential for effective geochemical surveys. These involve testing the response of known mineralization in a similar nearby area to verify the effectiveness of the method under the prevailing conditions, before using it routinely to look for unknown mineralization.

Geophysical Techniques

Geophysical techniques rely on variations in the physical properties of the mineralized rocks or the rocks that surround them to indicate the location of mineralization directly or indirectly (Table 3; Figure 4). The physical response may be a direct property of the rock (e.g., its density, natural radioactivity, magnetic properties, or resistivity), or an induced response produced by exposing the rock to a physical stimulus such as an electrical or magnetic field. Geophysical techniques may indicate mineralization directly, or may allow the search area to be reduced to a favourable host rock or a favourable structural environment. With the increasing availability of large-scale surveys, often involving more than one technique, the data from geophysical surveys can be interpreted to produce geological maps and to identify favourable settings for mineralization, exactly as they would be from conventional geological maps. A major advantage of geophysical techniques is that they can detect responses from mineralization buried several hundred metres below the ground surface, at depths too great to be reflected in the surface geology or detected by geochemical surveys.

Recognizing Targets

Sometimes orebodies are exposed and their surface expression can be recognized by explorers, but as exploration continues, new surface targets become fewer and fewer, and increasingly there is a need to look for more subtle targets. That typically involves the use of multiple exploration tools. The data from all the techniques used are compiled onto maps, and the strength and location of anomalies in each technique are compared. Areas that stand out with coincident anomalies from more than one technique are usually priority targets for drill testing.

One additional geological tool universally used by modern explorationists is the ore deposit model. There are models available for a wide range of ore deposit types, and new ones are regularly developed. These models summarize all the important geological characteristics (including spatial relationships) of a particular deposit type. Models include large-scale aspects such as the tectonic and structural settings,

Table 2 Main geochemical tools in mineral exploration

Tool	*Basis*	*Coverage*	*Applicability*	*Target*
Soil geochemistry	Systematic sampling of soil, usually of a selected size fraction	Large to small areas, depending on sampling density	Areas with residual soils; a very commonly used method	Most metallic deposit styles
Rock chip geochemistry	Systematic sampling of exposed bedrock as composites of small chips	Small areas only	Requires extensive exposure of bedrock	Most metallic deposit styles
Bedrock geochemistry	Systematic sampling of the upper parts of bedrock, usually by drilling through shallow cover	Large to small areas, depending on sampling density and depth of cover	Areas of strong near-surface leaching, or covered by transported overburden	Most deposit styles
Stream sediment geochemistry	Systematic sampling of stream sediments, usually of a selected size fraction	Mostly large areas; application to small areas may be limited by the stream pattern density	Widely applicable in regional exploration; requires bedrock outcrop or shallow subcrop	Most metallic deposit styles
Water geochemistry	Sampling of water in streams or wells and bore holes	May be applied over large areas; sample density typically low	Depends on access to water samples in streams or bores	Limited application for base metal and uranium deposits
Soil gas geochemistry	Systematic sampling and analysis of gases trapped in soil	Mostly small areas; sampling density is high	Suitable for areas with deep soil and transported cover	Locates structures and may help find some sulphide-rich deposit styles
Plant geochemistry	Sampling of one plant species or part of one plant species, followed by chemical analyses (mostly after ashing)	Mostly over restricted areas	Best in areas with high plant density but low plant diversity (usually temperate to cold areas)	Most base metal deposit styles

Figure 3 Geochemical sampling in eastern Taiwan. Sediments are dug from the creek bed, sieved to remove the largest pieces, and the finer fraction is bagged and sent to a laboratory for chemical analysis.

Table 3 Main geophysical tools in mineral exploration

Tool	*Basis*	*Coverage*	*Applicability*	*Target*
Magnetics	Measures spatial variations in the intensity of Earth's natural magnetic field	Very large areas covered by aerial surveys; smaller areas by ground surveys	Widely used as an aid to geological mapping; also a direct detection method for some deposits: applicable wherever rock magnetic properties vary	All deposits with a well-defined geological setting in magnetically variable rocks; direct detection of deposits containing magnetic minerals (e.g., iron ore, chromite, diamonds, some volcanic massive sulphide deposits, some porphyry copper deposits)
Radiometrics	Measures spatial variations in the intensity of natural radiation from potassium, thorium, and uranium	Very large areas covered by aerial surveys; smaller areas by ground surveys	Widely used as an aid to geological mapping; also aids direct detection of some deposits: applicable wherever there are variations in abundance of natural radioactive elements	All deposits with a well-defined geological setting; direct detection of deposits containing uranium or thorium minerals, or deposits with strong potassium enrichment in alteration zones (e.g., some volcanic massive sulphide deposits, some porphyry copper deposits, some epithermal gold deposits)
Gravity	Measures spatial variations in the intensity of Earth's natural gravitational attraction	Very large areas covered by aerial gravity gradiometer surveys; smaller areas by ground surveys	Huge areas covered by very low-density ground surveys; limited coverage with high-resolution aerial surveys; detailed ground surveys only in areas of low topographic relief: applicable wherever rock densities vary	Direct detection of deposits with strong density contrast to surrounding rocks (e.g., iron ore, volcanic massive sulphide deposits, kimberlites)
Electromagnetics	Measures the induced response in the Earth from an applied electromagnetic field	Very large areas covered by aerial electromagnetic surveys; smaller areas by ground surveys	Can be used as an aid to geological mapping; also aids direct detection of some deposits: applicable wherever there are variations in natural rock conductivities	Direct detection of deposits with high concentrations of conductive minerals (e.g., volcanic massive sulphide deposits, manganese deposits) or where there are associated anomalous conductivity zones (e.g., some porphyry copper deposits, kimberlites)
Induced polarization	Measures the induced electrical polarization response in minerals from an applied electrical field	Ground technique mostly applied over small areas, but large surveys also possible	Detects the presence of polarizable minerals (mainly pyrite, pyrrhotite, galena, chalcopyrite, etc.)	Direct detection of deposits with high concentrations of polarizable minerals (e.g., volcanic massive sulphide deposits, porphyry copper deposits)
Resistivity	Measures spatial variations in the resistivity of the rock mass being surveyed	Mostly applied over small areas	Detects variations in resistivity	Sulphide-rich mineralization has low resistivity compared to the surrounding rock; applied to direct detection of sulphide-rich masses (e.g., massive sulphide deposits)
Seismic	Records underground surface reflections produced by explosions or mechanical thumpers at the surface	Applied over huge areas for petroleum exploration; seldom used in mineral exploration	Indicates surfaces in the bedrock (e.g., faults, beds, igneous layers)	In mineral exploration, mainly used for layered complexes (e.g., Bushveld) or for engineering purposes

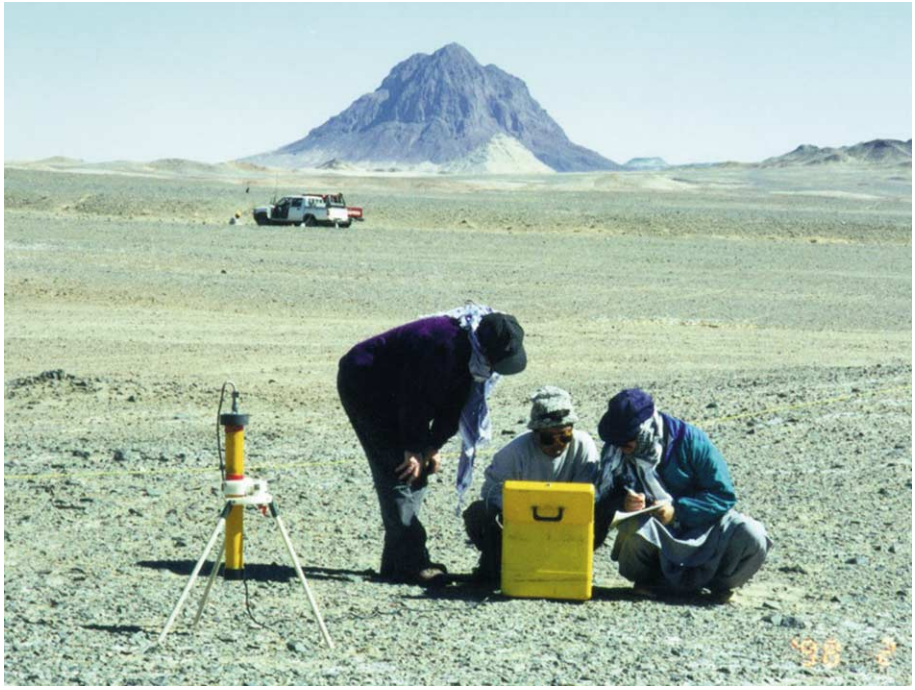


Figure 4 A geophysical survey being conducted in western Pakistan. This is an electromagnetic survey, in which electrical pulses are used to induce a response from conductive rocks in the ground. Photograph by R Irvine, reproduced with permission.

down to small-scale features such as the mineralogy of the ore, and details such as the chemistry of fluid inclusions. In most but not all cases they include genetic models summarizing how geologists believe the deposits form. They also reflect the kind of geological, geochemical, and geophysical expressions an orebody of this type should have, and so they help in the interpretation of exploration data.

Testing Targets

Once possible mineralization has been located or a target identified, it is then necessary to test it to confirm that it is what is sought, and how commercially significant it is (Figure 5). The cheapest approach in near-surface situations is to dig trenches across the mineralization to create complete exposure and to allow detailed sampling. If the target is more than a few metres below the surface, drilling is needed (Figure 6). Drilling is the most reliable means to investigate a target and is used to define the grade and extent of the mineralization. If testing confirms that potentially economic mineralization has been found, then a discovery has been made (Figure 7). Further work is then required to quantify the full significance of the discovery.

The assessment stage that follows a discovery involves gathering huge the amounts of data needed to assess whether the mineralization can be mined economically. Typically, widely spaced drilling is used first to assess the overall size of the mineralized body, and then more detailed infill is used to determine the shape, distribution, tonnage, grade, and metallurgical properties. If it is found that the indicated tonnage and grade could potentially justify starting a mine, a new stage of data-gathering follows; typically projects pass through several stages, at each of which it is reassessed to see if another more detailed stage of expensive assessment work is warranted. These stages typically include prefeasibility, full feasibility, and, if the outcomes are positive, finally a decision to go ahead and develop a mine. Depending on the commodity, the planned scale of operation, and the location of the deposit, developing a mine might cost anything from a few million dollars up to several billions of dollars. (*see* **Mining Geology:** Exploration Boreholes).

Success and Failure in Exploration

Exploration is a technical activity with a commercial objective. It is common to see these two aspects, the technical and commercial, confused in claims of success. In the case of junior companies, it is important that any success (not necessarily a commercial success) is publicly highlighted, because it is important for these groups to be seen as successful, to increase their ability to raise further funds for ongoing

Figure 5 Exploration success! The geologists are standing around the drill hole that discovered the high-grade Cannington silver–lead–zinc orebody in western Queensland, Australia. A drill rig can be seen on the skyline, where drilling is being done further along the same orebody. The deposit was brought into production and quickly became the world's largest producer of silver.

Figure 6 A helicopter being used to move a diamond drilling rig in the Kimberley region of Western Australia. The helicopter allows the drill to be positioned with minimal disturbance and environmental damage.

exploration. By contrast, it is sometimes in the interests of major mining companies to play down expectations from their exploration, and their successes often go unmentioned, or are announced only in a low-key way. Too much publicity can increase the level of competition for ground near the discovery, and overly enthusiastic expectations may mean shareholders are ultimately disappointed if a technical discovery does not turn into a commercial success.

The most important justification for exploration is that it supplies society with the mineral products it needs. This benefit, however, is realized only if the value of the products is greater than the cost of discovery and production. Real success in exploration and mining must therefore be judged in commercial, not technical, terms. In this respect, the role of the large exploration and mining companies has particular importance. They have financial and technical

Figure 7 Examining trays of drill core from the Eloise copper–gold deposit in western Queensland, Australia. Diamond drilling techniques are used, which allows deep holes to be drilled, and the core recovered allows detailed geological examination.

resources that small companies lack, and they are not so dependent on the day-to-day demands of the share market. This means they can pursue longer term goals and are more likely to discover and develop long-life mines that contribute stability to the long-term supply of major mineral commodities. Some junior companies are fast and economically efficient explorers and play an important role by supplying new exploration opportunities to larger companies and by bringing into production deposits that would be too small for major companies.

The Role of Governments in Exploration

In most parts of the world, governments regulate exploration and mining. Governments control the granting of titles, the conditions that are applied, and regulations governing safety and environmental protection. Governments also determine the availability of regional data and the cost of accessing it, and they set the levels of taxation and other imposts and concessions that are applied to the industry. Direct government involvement in exploration and mining is not regarded favourably these days, because having a government acting as both regulator and competitor involves them in a conflict of interest that may undermine trust in the regulatory process.

Although favourable geology is by far the most important determinant of how prospective a region is for exploration in a purely technical sense, the potential for commercial success from exploration is strongly influenced by the regulatory and financial environment in which it operates, as laid down by governments. Countries with relatively poor geology may still be attractive for exploration if the legal, economic, and regulatory environment is attractive; conversely, even the best geology does not compensate for a bad environment, because confidence in the stability of an operating environment is very important for explorers and miners.

The Future of Exploration

Mineral exploration is an essential part of the process of supplying the resource needs of societies: while that demand continues, so will mining and the exploration that is needed to sustain it. But exploration faces many challenges, and is currently undergoing major changes.

With the rise in environmental awareness in the second half of the twentieth century, mining was widely criticized for its environmental impact. Much of that criticism was justified, and there have been major changes, such that the modern mining industry can now operate with minimal environmental impact in the areas surrounding mines. However, it is inescapable that any process that takes material from one place to another must have an environmental impact. Mineral exploration has also come under attack by environmentalists, but most parts of the exploration process have very little to no environmental impact. Most opposition to exploration is either misinformed about what is actually done, or more commonly is based on fear of what might happen in the unlikely event that a discovery is

made, leading to mining. In fact, most exploration efforts leave no long-term trace and make no discoveries. Even when targets are identified, on further testing less than 1% prove to be of economic interest.

The mining industry as a whole, including exploration, has been criticized for not returning a satisfactory profit on money invested, but such analyses can only be sensibly done on a case-by-case basis. Over the past decade, major changes have occurred in financial markets and in how businesses operate, with emphasis on short-term financial performance. Because exploration cannot guarantee very quick returns, it lost favour with investors, with the result that in the late 1990s, exploration expenditure dropped drastically, and investment in mining developments were also sharply reduced in response to low commodity prices. If this situation were to continue, there would eventually be a shortage of some commodities, which should cause prices to rise, making exploration and mining of those commodities attractive again.

The last few years of the twentieth century represented a period of dramatic change for the mining and exploration industry. These changes are still being played out, but a new industry structure and way of operating have to be found. Throughout the world, all people seek a better life, and to fulfil that desire requires more and better houses, vehicles, roads, and everyday things. All those things have to be made from some material commodity, and most commodities can be obtained only by mining. Mining will ultimately die without exploration, so, as long as populations continue to increase and people strive for a better standard of living, requiring the use of mineral commodities, the future of mineral exploration is assured.

See Also

Analytical Methods: Geochemical Analysis (Including X-Ray). **Economic Geology**. **Geochemical Exploration**. **Geological Maps and Their Interpretation**. **Minerals:** Definition and Classification. **Mining Geology:** Exploration Boreholes; Hydrothermal Ores; Magmatic Ores. **Quarrying**.

Further Reading

Boyle RW (ed.) (1971) *Geochemical Exploration, CIMM Special Volume 11*. Quebec: Canadian Institute of Mining, Metallurgy, and Petroleum.

Dobrin MB and Savit CH (1988) *Introduction to Geophysical Prospecting*, 4th edn. New York: McGraw-Hill.

Evans AM (1995) *Introduction to Mineral Exploration*. London: Blackwell Science Ltd.

Harben PW and Kuzvart M (1997) *Industrial Minerals – Global Geology*. Surrey, UK: Industrial Minerals Information Ltd.

Levinson AA (1980) *Introduction to Exploration Geochemistry*, 2nd edn. Wilmette, Illinois: Applied Publishing.

Peters WC (1987) *Exploration and Mining Geology*, 2nd edn. New York: John Wiley & Sons.

Rose AE, Hawkes HE, and Webb JS (1979) *Geochemistry in Mineral Exploration*, 2nd edn. New York: Harper and Row.

Sillitoe RH (1995) *Exploration and Discovery of Base- and Precious-Metal Deposits in the Circum-Pacific Region during the Last 25 Years, Resource Geology Special Issue 19*. Tokyo: The Society of Resource Geology.

Sillitoe RH (2000) *Exploration and Discovery of Base- and Precious-Metal Deposits in the Circum-Pacific Region – A Late 1990s Update, Resource Geology Special Issue 21*. Tokyo: The Society of Resource Geology.

Telford EM, Geldart LP, Sheriff RE, and Keys DD (1976) *Applied Geophysics*. Cambridge, UK: Cambridge University Press.

Van Blaricom R (1992) *Practical Geophysics II for the Exploration Geologist*. Spokane: Northwest Mining Association.

Vanecek M (ed.) (1994) *Mineral Deposits of the World; Ores, Industrial Minerals, and Rocks, Developments in Economic Geology*. vol. 28. Amsterdam: Elsevier.

Mineral Reserves

M Vaněček, Charles University, Prague, Czech Republic

Introduction

Mineral reserves and mineral resources are terms that cannot be separated from each other. Their definitions are derived from the grade of geological certainty of the basic information about the mineral accumulation and from expected profitability of extracting and marketing the mineral commodity. In all main classifications (i.e., the United Nations, the USA, and the Russian classifications) the definitions of mineral resources and mineral reserves are slightly different, but in all cases the reserves form only a better known part of the resources. Procedures for mineral reserves calculation can be divided into the scalar-geometrical methods and into geostatistical methods. For the most part, in mineral statistics,

world reserves and reserve base of different minerals are introduced.

The terms mineral reserves and mineral resources are inseparably bound to each other. One of the possible definitions for resource is: "something in reserve or ready if needed". This can be made precise for mineral resources as follows: "valuable mineral deposits of an area that are presently recoverable and may be so in the future, including known ore bodies and potential ore". Mineral reserves usually represent a part of mineral resources that could be defined with respect to the quantity and quality of mineral bodies.

The evaluation of reserves/resources of a mineral deposit is based on reserve calculation procedures that would include description and analysis of main geological characteristics of the deposit, such as tonnage, grade, shape and thickness of ore bodies, and depth of mineralization etc., and on a feasibility study that analyzes expected profitability of extracting and marketing the mineral commodity in a given economy and at a given time.

Methods of reserves calculation and of economic analyzes of mining activities have developed over a long period of time. The various connoisseurs in prospecting, mining and metallurgy were the very first to be able to evaluate the worth of a mineral accumulation. In 1910, during the XI, International Geological Congress in Stockholm, the iron ore resources of the world were classified into different groups of exactitude and the symbols A, B, and C were used for the first time. At present time a special branch of economic geology, called mineral economics, deals with the reserves/resources evaluation and classification.

Due to the needs of industrialization, different systems of reserves classification and calculation have arisen. So far as the classification, i.e., group segregation, of mineral reserve tonnage is concerned, we find two schools of thought in existence. Governmental agencies, for example different geological surveys, are mostly concerned with the determination of the future potential of mineral resources. On the other hand, a reserve analysis made for or by a private enterprise usually is designed to show the various mineral tracts classified on the basis of their currently mineable nature. The classification terminology used by the governmental agencies is as follows: measured, indicated, and inferred resources. On the other hand, in mining geology there are used more readily the terms developed ore, proved ore, probable ore, and possible and/or extension ore. Single members of both classifications are not direct equivalents. For example, the developed ore may be only a part of measured or proved reserves that are prepared for immediate withdrawal by mining. Some probable ore may be the essential counterpart of some measured reserve or/and of some indicated resource. Possible ore, which is sometimes called future ore, and extension ore, are to all intents and purposes the equivalent of inferred reserve.

At present, three different reserves/resources classifications are mainly used: the United Nations International Framework Classification, the Classification System of the U.S. Bureau of Mines and U.S. Geological Survey, and the Classification of the State Commission for Reserves of the Russian Federation.

The UN Framework Classification provides information about the stage of Geological Assessment, the stage of Feasibility Assessment, and the degree of Economic Viability. Geological Study is subdivided into four consecutive stages of geological assessment, which are: Reconnaissance, Prospecting, General Exploration, and Detailed Exploration. Feasibility Assessment is subdivided into three consecutive stages which are: Geological Study, Prefeasibility Study, and Feasibility Study/Mining Report. There are two categories of Economic Viability: economic, and potentially economic, which are only quoted in the stages of Mining Report/Feasibility Study and Prefeasibility Study (**Table 1**).

With respect to a variety of meanings in the use of the terms 'reserve' and 'resource' they are redefined as follows: the Total Mineral Resource is defined as the naturally occurring concentration of mineral raw material of economic interest and with specified geological certainty. A Mineral Reserve is the economically mineable part of the Total Mineral Resource as demonstrated by Feasibility Assessment. The Remaining Mineral Resource is the balance of the Total Mineral Resource that has not been identified as a Mineral Reserve. Only the groups 'proved' and 'probable' are reserves, whereas the group 'inferred' belongs to resources. The UN Framework Classification is mostly applied in Western Europe and in some European countries with 'economies in transition' (**Figure 1**).

The U.S. Resource/Reserve Classification for Minerals defines Resource as a concentration of naturally occurring solid, liquid, or gaseous material in or on the Earth's crust in such form and amount that economic extraction of a commodity from the concentration is currently or potentially feasible. Identified Resources are resources whose location, grade, quality, and quantity are known or estimated from specific geological evidence. Identified resources include economic, marginally economic, and subeconomic components. To reflect varying degrees of geologic certainty, these economic divisions can be subdivided into 'measured', 'indicated', and 'inferred'. For the sum of measured plus indicated, the term 'demonstrated' is used. The part of an identified resource

Table 1 United Nations International Framework Classification for Reserves/Resources. Solid fuels and mineral commodities

UN International Frame-work	*National system*			
	Detailed exploration	*General exploration*	*Prospecting*	*Reconnaissance*
Feasibility study and/ or mining report	1 Proved Mineral Reserve (111)	Usually not relevant	Usually not relevant	Usually not relevant
	2 Feasibility Mineral Resource (211)	Usually not relevant	Usually not relevant	Usually not relevant
Prefeasibility study	1 Probable Mineral Reserve (121) + (122)	1 Probable Mineral Reserve (121) + (122)	Usually not relevant	Usually not relevant
	Prefeasibility Mineral Resource (221) + (222)	Prefeasibility Mineral Resource (221) + (222)	Usually not relevant	Usually not relevant
Geological study	1–2 Measured Mineral Resource (331)	1–2 Indicated Mineral Resource (332)	1–2 Inferred Mineral Resource (333)	? Reconnaissance Mineral Resource (334)

Economic viability categories: 1 = economic; 2 = potentially economic; 1–2 = economic to potentially economic (intrinsically economic); ? = undetermined.

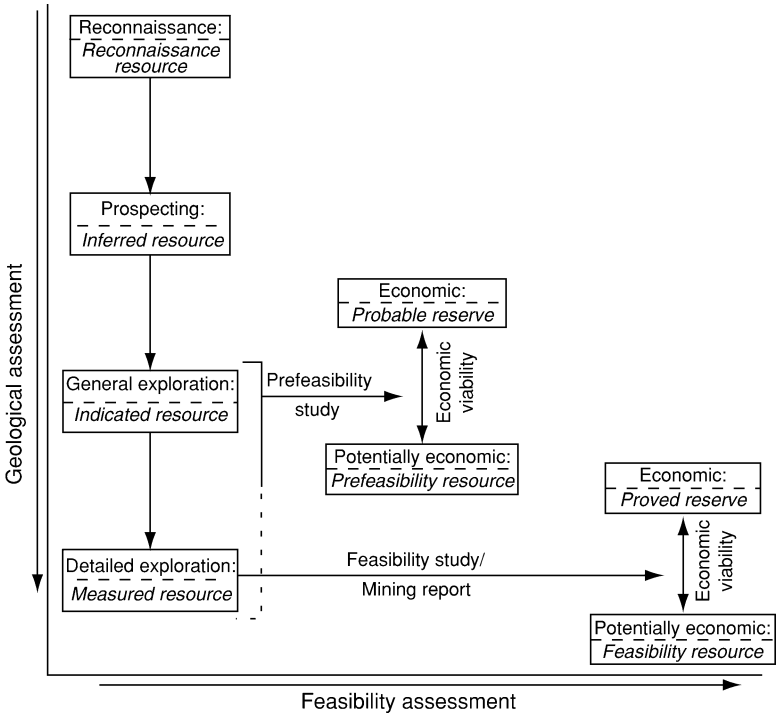


Figure 1 Schematic illustration of the UN Framework Classification.

that meets specified minimum physical and chemical criteria related to current mining and production practices, including those of grade, quality, thickness, and depth, is called Reserve Base. Reserves are that part of the reserve base which could be economically extracted or produced at the time of determination. The whole U.S. system of resource/reserve classification is illustrated by Tables 2 and 3. The U.S. Resource/Reserve Classification for Minerals seems to be the most widespread, not only in the USA, but in most of world.

In the classification of the State Commission for Reserves (GKZ) of the Russian Federation reserves and resources are distinguished. According to the grade of geological certainty the reserves are divided into the explored ones (categories A, B, C 1), and the estimated ones (category C 2). From the point of view of the economic importance, the reserves are

Table 2 Major elements of mineral-resource classification, excluding *reserve base* and *inferred reserve base*

	Identified resources			*Undiscovered resources*	
	Demonstrated			*Probability range*	
Cumulative production	*Measured*	*Indicated*	*Inferred*	*Hypothetical*	*Speculative*
Economic	Reserves	Reserves	Inferred reserves	—	—
Marginally economic	Marginal Reserves	Marginal Reserves	Inferred marginal reserves	—	—
Subeconomic	Demonstrated subeconomic resources	Demonstrated subeconomic resources	Inferred subeconomic resources	—	—
Other occurrences	Includes non-conventional and low-grade materials	Includes non-conventional and low-grade materials	Includes non-conventional and low-grade materials	Includes non-conventional and low-grade materials	

Table 3 Reserve base and inferred reserve base classification categories

	Identified resources			*Undiscovered resources*	
	Demonstrated			*Probability range*	
Cumulative production	*Measured*	*Indicated*	*Inferred*	*Hypothetical*	*Speculative*
Economic	Reserve base	Reserve base	Inferred reserve base	—	—
Marginally economic	Reserve base	Reserve base	Inferred reserve base	—	—
Subeconomic	Reserve base	Reserve base	Inferred reserve base	—	—
Other occurrences	Includes non-conventional and low-grade materials	Includes non-conventional and low-grade materials	Includes non-conventional and low-grade materials	Includes non-conventional and low-grade materials	

Table 4 Russian classification of reserves

	Grade of geological certainty		*Group of economic importance*	
Reserves	Explored	A	Economic	Subeconomic
		B	Economic	Subeconomic
		C_1	Economic	Subeconomic
	Estimated	C_2	Economic	Subeconomic
Resources				
	Prognostic	P_1		
		P_2		
		P_3		

subdivided into economic and sub economic. In the Russian mining industry a classification of the explored reserves according to their mineable nature is used as follows: developed, prepared for extraction, and ready to extraction. Resources are divided into three groups of prognostic resources (P 1, P 2, P 3) depending on the grade of geological certainty. Main patterns of the Russian reserves/resources classification are shown on Table 4. The Russian reserves classification is used mostly in the states of the former USSR and also in some Eastern European countries.

The general procedure of a mineral reserve calculation is the same regardless of the concept that is used in making the ultimate classification, since it follows that no classification of any mineral reserve can be visualized completely until all necessary working data have been evaluated. Two main groups of traditional resource estimations can be distinguished:

1. The scalar-geometric methods or methods of geometrization
2. The spatial ('geostatistical') methods.

In the first case a complex natural shape of mineral bodies are geometrized into defined geometric bodies, the volume of which can be calculated. The following methods are distinguished:

1. Methods of blocks
2. Methods of geological sections
3. Methods of isolines.

Methods of blocks can be divided depending on a geometrical arrangement of the exploration points

(i.e., boreholes, shafts, etc.) into triangular, regular squares, rectangular, polygonal, and irregular or geological blocks. These methods differ slightly in ways of determination of volume and grade of mineral bodies. The average thickness is calculated as an arithmetical mean in all cases with an exception of the polygonal method where thickness and grade are taken from the borehole lying at the geometrical centre of each polygon. In all other kinds of blocks an average grade of the mineral body is calculated as an arithmetical mean weighted by thicknesses. Polygonal blocks and geological blocks are applied if the arrangement of the exploration points is irregular. Contours of geological blocks are derived from the geological situation (tectonic elements, facial changes, changes of thickness, etc.).

Methods of sections can be divided as follows:

1. Longitudinal sections (along the axis of the body)
2. Cross-sections.

Sections may be vertical, horizontal, and inclined, each other being parallel and non-parallel. Owing to the application of geological sections the natural shape of the body is nearly preserved. Volumes of blocks can be calculated either between two sections or using area of only one section. The first way is applied in the case of similar in shape and near in dimension geological sections, the second one is used if the shape of the mineral body is very irregular and/or if due to geological reasons a connection between two sections is not assumed. The average grade of blocks is calculated as an arithmetical mean weighted by the areas of geological sections of the body. Areas of the sections are measured by means of instruments (planimeter, scanning, etc.) or average thickness may be used.

For calculation by means of isolines there are used isopachs (i.e., isolines of thicknesses), isohypses (isolines of altitude), grade, density and their combination can be used. Volume of mineral bodies is given by contours of isolines and can be calculated using the Simpson formula. If a combination of isopachs, of isolines of grades, and of isolines of density is used the computed volume is equal to the reserve.

Geostatistics, in their most generally accepted use, are concerned with the study of the distribution in space of useful values for mining engineers and geologists, such as grade, thickness of accumulation, including a most important practical application to the problems arising in mineral deposit evaluation. Spatial or geostatistical methods are working with statistical processing of primary data and with description of a statistical character of the distribution of main properties of the deposit in space by means of variograms. The discovered statistical characteristics that allow the prediction of the distribution of studied properties in the mineral body are used for the computing of reserves. One of the most known and used geostatistical methods is so called 'kriging', a method developed for gold reserves computing in South Africa.

In the international statistical overviews of the world reserves and of the regional resources of a mineral commodity, the terms 'resources' and 'reserves' base are usually given. For the metallic ores the reserve are given mostly in terms of metal content. Reserves of naturally occurring non-metals and fossil fuels are expressed in the tonnage of raw mineral material. Quantities, qualities, and grades may be expressed in different terms and units to suit different purposes, but usage must be clearly stated and defined.

Mineral reserves can be defined also as: "a quantity of mineral raw materials in the Earth's crust which is estimated as a result of geological exploration works or during their extraction". They could be calculated for a single deposit, for an ore field or district, for a basin, for a region, for states, for continents, and for the whole world.

World mineral reserves of each raw mineral commodity are immense and their quantity is subject of changes due to the development of geological knowledge, in conjunction with considerable mining activity. For example, world resources of iron ores are estimated to exceed 800 billion metric tonnes of crude ore containing more than 230 billion tonnes of iron. World reserves in iron content are estimated to be 71 000 million metric tonnes, the reserve base being 160 000 million metric tonnes. Annual world output of iron ore mines varies by about 1 billion metric tonnes per year. World reserves of phosphate rocks are recorded as high as 12 billion metric tonnes, reserve base being 37 billion metric tonnes. World resources of coal are estimated to be about 15×10^{12} metric tonnes, of which about 1067 billion metric tonnes are recoverable reserves. In 2001 the annual world mining output of coal was 3630 metric tonnes. These few examples may give an idea as to the order of huge quantities of world mineral resources and reserves.

See Also

Military Geology. **Mining Geology:** Exploration. **Petroleum Geology:** Reserves.

Further Reading

Agterberg F P (1974) Geomathematics: mathematical background and geoscience applications. In: *Developments in Geomathematics,* vol.1, p. 596. Amsterdam, London, New York: Elsevier.

Bárdossy G, Szabó I R, and Varga G (2003) *A New Method of Resource Estimation for Bauxite and Other Solid Mineral Deposits*, pp. 57–64. New York: Springer Wien.

Carr D D and Herz N (eds.) (1989) *Concise Encyclopedia of Mineral Resources*, p. 426. Oxford-New York: Pergamon Press.

Clark I (2000) *Practical Geostatistics*, pp. 68–85. Columbus, Ohio: Greyden Press.

Goovaerts P (1997) *Geostatistics for Natural Resources Estimation*, p. 483. New York: Oxford University Press.

Jewbali A and Mousset-Jones P (2002) A survey of sampling and resource/reserve estimation practices in the surface gold mining industry. *Mining Resources of England* 11(2): 183–195.

Mineral Commodity Summaries 2003 (2003) *U.S. Geological Survey*, p. 199. Washington: U.S. Department of the Interior.

Rudawski O (1986) *Mineral Economics*, p. 192. Amsterdam: Elsevier.

U.N. Economic and Social Council (1999) *United Nations International Framework Classification for Reserves/Resources*. New York: U.N. Economic and Social Council.

Wellmer F W (1983) *Klassifikation von Lagerstättenvorräten mit Hilfe der Geostatistik*, pp. 9–43. Schriftenreihe der GDBM, H.39, Weinheim: GDBM (GDBM = Society of German Miners and Metallurgists).

Working Group of IMM in Conjunction with EFG and IGI (2000) *European Code for Reporting of Mineral Exploration Results, Mineral Resources and Mineral Reserves*. (The European Code). Consultation draft.

Hydrothermal Ores

M A McKibben, University of California, CA, USA

Introduction

Hydrothermal ores are natural, economically valuable concentrations of metallic minerals whose origins can be directly tied to the physical and chemical actions of hot aqueous fluids within the Earth's crust. Hydrothermal ores are ideally distinguished from those that form by other geological processes, including magmatic ores (formed by segregation from a melt), sedimentary ores (formed by sedimentation or precipitation from ambient surface waters), and residual ores (formed by surficial weathering). Such strict genetic distinctions are not always possible, however; for example, magmatic-hydrothermal ores (such as skarns and pegmatites) may arise from a combined process whereby an aqueous phase exsolves from a crystallizing magma. The origins of some diagenetic ores likewise may fall within a realm between purely sedimentary and hydrothermal processes. So while the defining criteria for classifying minerals and ores as hydrothermal is evidence of transport and precipitation from a largely aqueous fluid phase at elevated temperatures, there is always the caveat that transitional processes can take place.

For millennia, humans have observed connections between areas of active subaerial volcanism and the mineralizing actions of hydrothermal fluids, manifested in surficial features such as fumaroles, hot springs, geysers, mud pots, and sinter and travertine terraces. More recently, marine scientists have discovered similar active hydrothermal features on the ocean floors near mid-ocean ridge volcanic centres, particularly 'black smokers' and their associated mounds of accumulated hydrothermal minerals. In all of these instances, it is obvious that minerals containing elements of economic significance are periodically concentrated by such volcanic-related hydrothermal phenomena.

Chemical and physical clues gathered from older ore deposits, as well as deep commercial and scientific drilling into active hydrothermal systems over the last few decades, have provided additional details of hydrothermal processes that are only hinted at in surficial manifestations. These field-based data on modern and ancient systems, coupled with essential theoretical and experimental laboratory studies of hydrothermal processes, have added to our knowledge of the full spectrum of hydrothermal processes that have taken place within the Earth's crust through geological time. Thus, although hydrothermal ores are often associated with volcanism, geologists now recognize that some of these ores can be generated from processes that are peripheral to or entirely unrelated to magmatic processes.

Hydrothermal Fluids

Because of their high temperatures and pressures, hydrothermal fluids in the Earth's crust can contain relatively elevated concentrations of dissolved salts (such as NaCl) and gases (such as H_2S) (Table 1). Dissolved salts and gases can arise from a variety of sources, including evaporative concentration of

Table 1 Representative chemical compositions (mg kg^{-1}) of geothermal fluids and brines in selected geothermal fields of the Salton Trough

Field					
Well	*Salton Sea*	*Imperial*	*Cerro Prieto*	*East Mesa*	*Heber*
Temperature (°C)	*52–14*	*L2–28*	*M-5*	*6-1P*	*5*
Depth (m)	*330*	*275*	*300*	*190*	*195*
Constituent[a]	*2500–3220*	*3290–4270*	*1200*	*2164*	*1800*
Na	54 800	50 466	5004	6362	4019
Ca	28 500	18 140	284	759	750
K	17 700	9555	1203	1124	333
Fe	1710	3219	<1	NA	NA
Mn	1500	985	1	NA	NA
SiO_2	>588	465	569	257	237
Zn	507	1155	NA	NA	NA
Sr	421	1500	NA	NA	41
B	271	217	11	NA	4
Ba	210	2031	NA	NA	4
Li	209	252	13	NA	7
Mg	49	299	<1	9	2
Pb	102	>262	NA	NA	NA
Cu	7	>1	NA	NA	NA
Cd	2	4	NA	NA	NA
NH_4	330	NA	NA	NA	6
Cl	157 500	131 000	9370	11 668	7758
Br	111	NA	31	NA	NA
CO_2	1580	30 000	2400	NA	186
HCO_3	NA	NA	NA	221	NA
H_2S	10	>47	180	NA	1
SO_4	53	NA	4	51	66
TDS	26.5%	25.0%	1.6%	2.2%	1.3%

[a]The compositions given have been corrected for loss of gases and steam during wellbore phase separation and therefore represent the *in situ* reservoir compositions.

NA, not analysed.

TDS, Total Dissolved Solids in weight percent.

Modified from McKibben MA and Hardie LA (1997) Ore-forming brines in active continental rifts. In: Barnes HL (ed.) *Geochemistry of Hydrothermal Ore Deposits*, 3rd edn., pp. 875–933. New York: Wiley-Interscience.

surface waters, dissolution of evaporites, water–rock interaction, and exsolution from magmas.

The solvent (dissolving) properties of hydrothermal fluids impart their ability to create ore deposits. The high dielectric constant of water and the ability of metal cations to bond with aqueous anions (such as chloride and sulphide) give natural aqueous fluids in the Earth's crust the potential to dissolve, transport and re-deposit the elements that make up hydrothermal ore minerals.

Water vapour is far less effective than liquid water in transporting most valuable elements. The temperature and pressure conditions under which aqueous liquids can effectively transport metals depend not only on the physical properties of water but also on the amounts of dissolved minerals and gases in the water. Salts such as NaCl dramatically increase the critical temperature and pressure of hydrothermal fluids, allowing brines to exist up to magmatic conditions as very effective liquid transporters of metals.

Metals are often solubilized and carried in solution as metal–anion complexes, such as $ZnCl^+$ and $Au(HS)_2^-$. Precipitation of these metals as ore minerals occurs when physico-chemical conditions destabilize the complexes in favour of the minerals.

Cooling, boiling, depressurization, oxidation, fluid mixing and fluid–rock reaction are examples of common processes which can trigger ore mineral precipitation from fluids. For example, native gold can be deposited during shallow effervescence and oxidation of hot spring waters:

$$2Au(HS)_2^- + 7.5O_{2(g)} + 2HCO_3^- = 2Au_{(s)} + CO_{2(g)} + 4HSO_4^-$$

Likewise, Zn sulphide can precipitate when chloride brines react with carbonate wall rocks such as limestone:

$$ZnCl^+ + H_2S_{(aq)} + CaCO_{3(s)} = ZnS_{(s)} + Ca^{2+} + H_2CO_{3(aq)} + Cl^-$$

In addition to reacting chemically, hydrothermal fluids will flow in response to thermal gradients, density gradients, and gradients in hydrological head or pressure. Flow rates will be highest in zones of maximum permeability, such as fracture and fault zones in brittle rocks, as well as in more porous lithologies such as sandstones and carbonates. Concentrations of ore minerals are thus formed in loci where a number of chemical and physical environmental factors have acted in concert to promote both sustained fluid flow rates and effective precipitation rates of ore minerals.

Hydrothermal Minerals

Economic geologists divide hydrothermal minerals into ore minerals (economically valuable) and gangue minerals (economically worthless). While some hydrothermal ore minerals occur as native metals or alloys, most consist of a metal combined with another non-metallic element or anion group – examples are shown in Table 2. Gangue minerals usually consist of common silicate, oxide, carbonate and sulphate minerals that do not contain sufficient concentrations of valuable metals. Nonetheless, gangue minerals can be just as important as ore minerals in providing scientific clues to the origins of hydrothermal ore deposits, particularly in regard to their age and the origins of the water in the ore-forming fluid.

Table 2 Examples of common hydrothermal ore minerals

Elements	Copper	Cu
	Gold	Au
	Mercury	Hg
	Silver	Ag
Alloys	Electrum	(Au,Ag)
Oxides	Cassiterite	SnO_2
	Cuprite	CuO_2
	Magnetite	Fe_3O_4
	Pyrolusite	MnO_2
	Uraninite	UO_2
Sulphides	Argentite	Ag_2S
	Arsenopyrite	FeAsS
	Bornite	Cu_5FeS_4
	Chalcocite	Cu_2S
	Chalcopyrite	$CuFeS_2$
	Cinnabar	HgS
	Cobaltite	(Co,Fe)AsS
	Covellite	CuS
	Galena	PbS
	Molybdenite	MoS_2
	Realgar	AsS
	Sphalerite	ZnS
	Stannite	Cu_2FeSnS_4
	Stibnite	Sb_2S_3
Sulphosalts	Enargite	Cu_3AsS_4
	Pyrargyrite	Ag_3SbS_3
	Tennantite	$(Cu,Zn)_{12}(As,Sb)_4S_{13}$
Tellurides	Calaverite	$AuTe_2$
	Sylvanite	$(Au,Ag)Te_2$
Tungstates	Scheelite	$CaWO_4$
	Wolframite	$(Fe,Mn)WO_4$
Carbonates	Cerussite	$PbCO_3$
	Smithsonite	$ZnCO_3$
Sulphates	Anglesite	$PbSO_4$

The crystal habits and textures of hydrothermal ore and gangue minerals are controlled by many factors, including fluid saturation states, nucleation/precipitation kinetics and rock surface textures. Fluids can move through brittle rocks most easily by virtue of interconnected fracture and pore permeability, and precipitation of ore minerals in such open spaces coupled with sustained flow rates (nutrient supply) often promote the development of abundant crystal faces. A straightforward textural sequence of precipitation may be recorded by bands or crusts of crystals deposited one after the other along a fracture or pore surface (Figure 1). Vugs or cavities lined with pristine euhedral crystals may remain as a consequence of the eventual cessation of precipitation or fluid flow; such mineral textures can look quite spectacular and are often a reason that people initially become interested in collecting and studying hydrothermal minerals.

Less efficient chemical flow through rocks can occur by the process of diffusion, where fluid components migrate in response to concentration gradients rather than fluid flow (advection). For example, many hydrothermal veins exhibit the development of alteration envelopes, areas of immediately adjacent wall rock in which new minerals have formed via diffusion and infiltration of chemical components from the hydrothermal fluid as it flows through the fracture (Figure 2). New hydrothermal phases such as silicates, carbonates and sulphides often form in these envelopes at the expense of primary phases in the wall rock. Some phases in alteration envelopes, such as potassium feldspars, can prove quite useful in making

Figure 1 Banded, crustiform sphalerite (ZnS, yellow to brown) and marcasite (FeS_2, green) with later pore-filling, cross-cutting galena (PbS, grey) in polished slab of hydrothermal ore from Olkusz mine, Silesian–Cracovian lead–zinc district, Poland. US 10 cent coin for scale.

Figure 2 Vein of quartz-molybdenite (MoS_2, grey) ore surrounded by alteration envelope of pink alkali feldspar in polished slab of altered porphyry granitoid, Nevada Moly (Hall) mine, near Tonopah, Nevada. US 10 cent coin for scale.

Figure 3 Porphyry Cu–Mo ore showing cross-cutting silica veins, from Gross Peak pit, Ithaca Peak (Mineral Park) mine, Arizona. US 10 cent coin for scale.

radiogenic age determinations of the age of formation of the veins.

At elevated temperatures rocks become ductile rather than brittle and open fractures and pores cannot be sustained. Under these conditions, hydrothermal mineral deposition occurs more commonly by replacement rather than by open-space filling. Extreme instances of wholesale hydrothermal replacement of wall rock by ore and gangue minerals can accompany metasomatism, a metamorphic process which substantially changes the original wall-rock composition. Both fluid flow and diffusion occur during metasomatism. Skarns are probably the best example of metasomatic hydrothermal ore deposits, where wall-rock carbonate is replaced by coarse calc–silicate minerals. Such deposits typically form near the contact zones of hydrous silicate magmas, with heat and silica provided by the cooling melt.

The textures of hydrothermal ores, especially their cross-cutting relationships, are an important aspect of the study of hydrothermal ore deposits (Figure 1, Figure 3). The term paragenesis is used for the inferred sequence of mineral precipitation based on textural studies of ores. Because most hydrothermal ores consist of a mixture of opaque and non-opaque minerals, the use of reflected-light microscopy in addition to transmitted-light microscopy is essential in their study.

After raw ore is mined, it is ground up and processed in such a manner as to separate the ore and gangue minerals, often by taking advantage of the chemical, density, magnetic, or surface adsorption differences among the particular minerals. For example, cyanide solutions can be used to preferentially dissolve gold from ore, and many base metal sulphide minerals can be efficiently separated from silicate minerals by their different wetability characteristics in oil and water. Once separated, the ore minerals are then smelted and refined to produce the pure metal.

Often there are practical economic constraints in processing ores that may only be revealed by very careful microscopic study of the paragenesis. For example, some hydrothermal gold ores contain native gold or electrum encapsulated in later silica, making grinding and processing to release the gold much more difficult and expensive. In the case of one famous California gold mine, the mine plant and smelter were designed before this practical paragenetic constraint was realized, forcing expensive last-minute changes to the processing circuit!

Hydrothermal Alteration

Because the total volume of rock subject to fluid flow and diffusion is typically much greater than that subject to more localized ore precipitation, hydrothermal alteration accompanying ore mineralization is often a larger and more obvious target for exploration than the ore deposit itself. Economic geologists therefore

look for characteristic alteration mineral assemblages that tend to develop in wall rocks around hydrothermal ore mineralization. Common alteration types in aluminosilicate wall rocks are listed below. For each type, the set of alteration minerals that appears in a rock is a function of its bulk composition.

Advanced argillic:
- alunite, sericite, kaolinite, dickite, pyrophyllite, pyrite

Phyllic:
- sericite, pyrite

Argillic:
- sericite, kaolinite, halloysite, montmorillonite, calcite, biotite, chlorite, pyrite

Propylitic:
- sericite, montmorillonite, zeolite, epidote, calcite, biotite, albite, K-feldspar, ankerite, chlorite, Fe sulphides, Fe oxides, siderite

Potassic:
- sericite, calcite, anhydrite, ankerite, chlorite, biotite, K-feldspar, pyrite, Fe oxides, siderite

Other more monomineralic alteration types include silica and carbonate alteration.

Because the formation of both ores and alteration envelopes are a function of permeability and fluid flow patterns, the geometry of the alteration assemblages can sometimes give clues to the likely location of additional ore during exploration and drilling, as well as aid in reconstructing the palaeohydrology of the ore-forming system. Potassic, phyllic and advanced argillic alteration often mark the highest-temperature, most intense part of a hydrothermal system close to the ore, whereas propylitic alteration often marks the peripheral regions. Silica alteration forms mainly during cooling and tends to occur in the hydrological discharge regions of the system, whereas carbonate alteration typically occurs during heating in the recharge regions.

On the basis of empirical data and proposed genetic models for the origins of different ore types, many national geological surveys have published exploration models that aim to assist in the discovery of new deposits. These idealized models emphasize characteristic alteration assemblages, geochemical fingerprints, geological structures, rock associations and geological ages of major ore deposit types. The models should be viewed as works in progress, subject to revision as new data and theories materialize.

Origins of Ore-Forming Hydrothermal Fluids

The origins of all of the constituents in an ore-forming hydrothermal fluid can be complex and may not be the same, but usually such a fluid can be classified in terms of the origin of its main component, water. The water can be (1) derived directly from magma, (2) derived from ground or connate water heated by magmatic processes, or (3) derived from water heated by non-magmatic processes such as subsidence, tectonism and metamorphism. The following is a list of examples of hydrothermal ore types grouped on this basis.

Magma-hydrothermal fluids:
- pegmatite
- skarn
- porphyry base metal (early stage)

Magma-heated surficial and subsurface waters (including seawater):
- volcanogenic massive sulphide (black smoker)
- porphyry base metal (late stage)
- epithermal vein

Waters heated mainly by tectonism, metamorphism and subsidence:
- Mississippi Valley type (Pb–Zn carbonate)
- stratiform/stratabound base metal
- greenstone belt (orogenic) gold

The origins of water in a hydrothermal ore deposit can sometimes be inferred from the geological setting, but often it also must be based on stable and radiogenic isotope ratios, elemental geochemical signatures, and fluid inclusion compositions in the ore and gangue minerals.

Magma-Hydrothermal Fluids

Magma-hydrothermal fluids arise from the crystallization of magmas that contain water. Magmas can contain significant amounts of dissolved water, sometimes exceeding 10% by weight. They can acquire water initially from the partial melting of 'wet' source rocks containing hydrous mineral phases (amphiboles, micas), and they can also acquire water subsequently via assimilation of hydrous wall rocks during their ascent. In either case, such waters will often exhibit H and O isotopic values that reflect equilibrium exchange with magma at high temperature. Silicic magmas tend to hold more dissolved water than mafic magmas, because of the ability of water molecules to interact with highly polymerized silicate melts.

As a hydrous magma cools and crystallizes, the early precipitation of largely anhydrous silicate liquidus phases (orthosilicates, feldspars) forces the dissolved water to be contained in a progressively shrinking mass of melt. Eventually, the diminishing melt may become saturated in water, forcing exsolution of a separate aqueous fluid (often violently) upon further crystallization. Aqueous fluid exsolution also can be

triggered by a sudden drop in confining pressure on the magma in response to rapid ascent, stoping, seismic events, or collapse of an overlying volcanic edifice.

This aqueous exsolution process, traditionally termed 'resurgent boiling', can result in injection of the magma-hydrothermal fluid into the crystallized outer carapace of the magma as well as the surrounding wall rock. The aqueous fluid often carries incompatible large ion lithophile elements as well as volatile components such as CO_2, HCl, HF, $B(OH)_3$, SO_2 and H_2S. Metal ions may be carried along with the halides and sulphur, typically as metal chloride or sulphide complexes.

As this exsolved fluid reacts with the extraneous rocks and pore fluids it encounters, it cools and evolves chemically; it may also undergo further phase separations (boiling and gas exsolution). Because the stabilities of aqueous metal complexes can be strongly affected by changes in temperature, pressure, pH and other chemical parameters, the cooling, reacting fluid can precipitate hydrothermal ore minerals in open fractures and pores, as well as deposit ore minerals by replacement in suitably reactive wall rock.

Economic geologists look for a number of lines of evidence to link a specific ore deposit to a magma-hydrothermal origin. Field evidence may show spatial and temporal proximity of the ores to a source pluton; for example, veins, breccias or pegmatites which begin in a pluton and then branch out into and cross-cut surrounding wall rock. Hydrothermal alteration developed in the surrounding wall rock may show concentric mineralogical and geochemical zonation around a likely source pluton (Figure 4). Evidence can also include remnants of the magma-hydrothermal fluid itself, preserved as aqueous fluid inclusions trapped in the ore and gangue minerals. Such inclusion fluids will typically have very high salinities, high homogenization temperatures and daughter minerals indicative of their origin. Stable isotope ratios of H and O in the minerals may also yield values and calculated equilibration temperatures consistent with a magmatic-hydrothermal origin.

Many magma-hydrothermal ore deposits (such as the numerous skarn and porphyry base metal deposits found in the south-western United States, Chile, Papua New Guinea, and Indonesia) exhibit evidence of early exsolution of a magma-hydrothermal fluid followed by late-stage heating and entrainment of surrounding groundwaters as the pluton

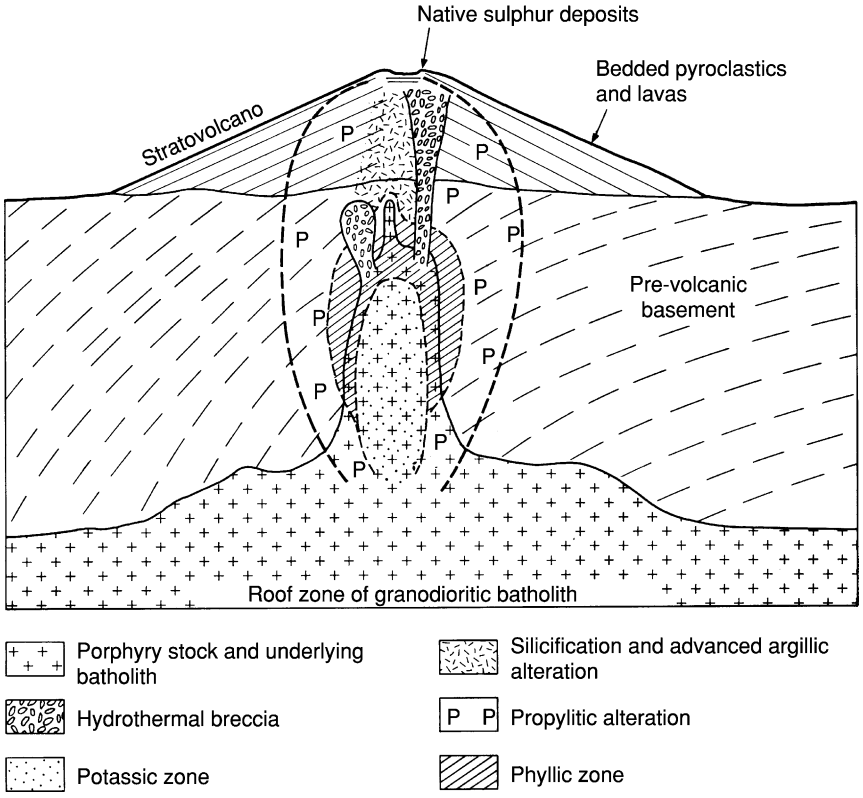


Figure 4 Diagramatic representation of a simple porphyry copper system. (Reproduced with permission from Evans AJ (1993) *Ore Geology and Industrial Minerals: An Introduction*, 3rd edn. Oxford: Blackwell.)

cools. This may result in redistribution of the early-formed ore metals and development of a larger-scale hydrothermal convective meteoric system more akin to the magma-heated systems discussed in the next section.

Examples of some major magma-hydrothermal ore deposits include Grasberg (porphyry and skarn Cu in Indonesia), Bingham (porphyry and skarn Cu in Utah), King Island (skarn W in Tasmania), and Chiquicamata (porphyry Cu in Chile).

Magma-Heated Surface, Ground and Ocean Waters

Some hydrothermal ores owe their origins to waters that were mainly heated by, but not directly exsolved from, magmas. In many cases any aqueous geochemical signature from magma degassing or exsolution is simply swamped out by the availability of shallow water from other sources. Such non-magmatic water sources include meteoric waters, groundwaters, lake waters and ocean waters. There are specific types of plate tectonic settings where such waters often become heated by proximity to magmatic processes.

In continental settings, shallow (epizonal) magmas often ascend into close proximity to surface and subsurface waters in volcanic arcs. Being topographically elevated, active continental volcanic arcs can receive abundant meteoric precipitation and are sufficiently porous and permeable (due to recurring tectonism and volcanism) that rapid recharge of their subsurface waters occurs. Crater lakes, caldera lakes, and tectonic lakes (formed from landslide blockage of drainages) provide additional reservoirs of surface water for recharge in arcs. At higher elevations and latitudes, snow and glaciers on the flanks of volcanoes can provide recent meteoric water upon melting.

This combination of shallow heat sources plus readily available meteoric water in permeable rocks creates ideal conditions for shallow convective hydrothermal systems, in which cold meteoric recharge waters descend, become heated by magmas, and rise buoyantly to discharge at or near the surface. Fumaroles, hot springs, geysers and mud pots at the surface attest to such fluid circulation cells at depth. As the heated meteoric waters circulate through the wall rocks and even the outer crystallized portions of a magma, they can acquire dissolved salts and metals and become ore-forming fluids, re-precipitating the metals as ores when they cool, boil or react with shallower groundwaters and rocks. Such shallow vein and disseminated hydrothermal ores are often termed epithermal on account of their shallow depth of formation. Examples include Creede, Colorado and Hishikari, Japan.

Arc magmas do not passively contribute only heat to overlying epithermal systems. Exsolving magmatic gases such as SO_2 and H_2S can become entrained in epithermal circulation systems, oxidizing to sulphuric acid near the surface. This results in a common association between some epithermal ores and acidic alteration (micas, clays, sulphates) of shallow wall rocks. Metals from magmas may also be exsolved into shallow epithermal systems, so they really should be viewed as hybrid systems with variable degrees of chemical input from magmatic and non-magmatic sources. Epithermal ores with evidence for a high contribution of components (especially acidic sulphur gases) from shallow degassing magmas are called high sulphidation state ores; they typically form in settings where a crater lake sits right above a shallow volcanic magma. Examples include Summitville, Colorado and El Indio, Chile.

Figure 5 schematically shows examples of many types of settings for hydrothermal systems in which epithermal and mesothermal gold deposits are formed.

It should not be surprising that magma-hydrothermal and epithermal hydrothermal processes can overlap in time and space, or that transitions between these processes should be common. Economic geologists use tools such as fluid inclusions and stable isotopes (O, H, S) to evaluate the relative roles of magmatic and meteoric components in the genesis of specific ore deposits. It is not uncommon to find deep, largely magma-hydrothermal ore deposits (such as porphyry base metal or skarn deposits) beneath temporally equivalent epithermal deposits (such as gold and silver vein and disseminated deposits) in shallower wall rock and volcanics.

Continental rifts are another plate tectonic setting where magmas and non-magmatic waters can interact to form hydrothermal ores. In such rifts, extension, subsidence and rapid sedimentation take place in conjunction with bimodal volcanism, providing opportunities for shallow basaltic and rhyolitic magmas to interact with lakes and groundwaters. Moreover, evaporation of closed-basin surface waters in rifts can lead to the development of very saline brines, which when heated by magmatism have tremendous capacity to dissolve, carry and re-precipitate ore metals derived from sedimentary and volcanic sources. Subsurface vein and disseminated ores can form in rift sediments and volcanics, along with overlying stratiform lenses and layers of ores formed when upwelling brines are injected or expelled parallel to shallow sediment layers. The Salton Trough of western North America (Table 1) and the Red Sea are modern examples of such settings.

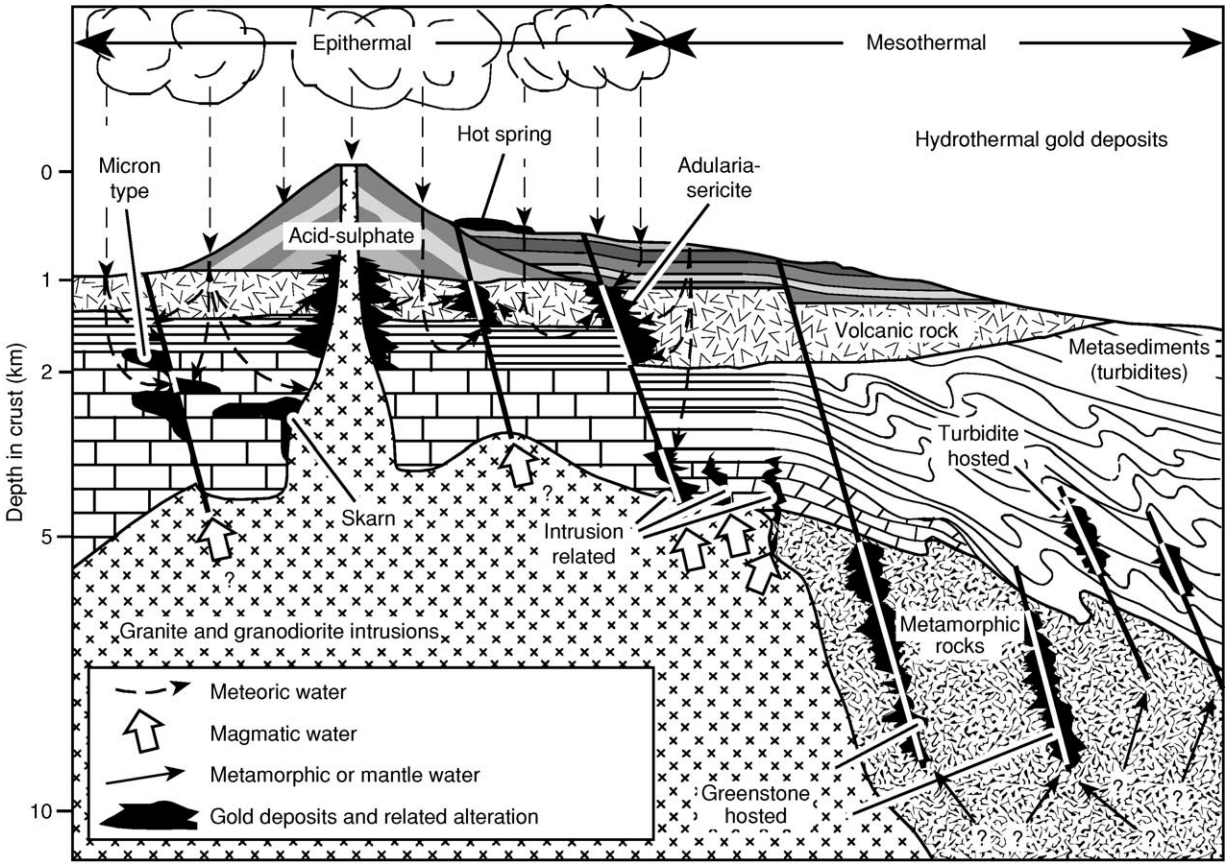


Figure 5 Schematic illustration of the wide variety of occurrences of hydrothermal gold deposits in continental crust. The terms epithermal and mesothermal refer to shallow-crustal and mid-crustal level deposits, respectively. (Reproduced with permission from Kesler SE (1994) *Mineral Resources, Economics and the Environment.* New York: MacMillan.)

In oceanic settings, volcanic arcs and rifts can lead to the formation of hydrothermal ores similar in principle to those found on continents. The main differences are the pervasiveness of seawater as a source of water and the dominance of oceanic crust and basaltic magmatism. Mid-ocean ridges, where roughly two-thirds of the Earth's heat flow occurs, develop large-scale hydrothermal convection systems that flux tremendous amounts of seawater through hot oceanic crust. The cold, oxidized metal-deficient and sulphate-rich seawater reacts with the hot basaltic crust and becomes a hot, reduced metal- and H_2S-rich ore-forming fluid. The metals that have been mobilized from the oceanic crust are re-deposited as vein and disseminated sulphide ores in the subsurface of the ridge structure, as well as expelled from submarine hot springs (black smokers) to form mounds and precipitates of metal sulphide on the seafloor. Only within the last few decades have earth scientists become aware of the spectrum of oceanic hydrothermal deposits that form at oceanic ridges, seamounts and back-arc basins. Many volcanogenic massive sulphide (VMS) deposits, long known to have spatial associations with altered mafic rocks, are now recognized as remnants of ancient oceanic ridge and back-arc hydrothermal systems that have since been accreted to or obducted onto the continents. The Cu–Pb–Zn districts of Kuroko, Japan and Jerome, Arizona are ancient examples; modern settings include massive sulphide deposits forming along the Juan de Fuca ridge in the north-eastern Pacific.

Hydrothermal Fluids Not Directly Affiliated with Magmatic Processes

Hydrothermal fluids and ores can be formed via heating mechanisms that have no direct association with magmatic processes. Basin subsidence through the normal geothermal gradient can heat sediment pore fluids and induce free convection, enhancing their ability to carry and redistribute metals. Such processes may be responsible for unconformity-related basinal uranium deposits such as those found in Saskatchewan (Canada) and Northern Territory (Australia). Rapid subsidence may also induce

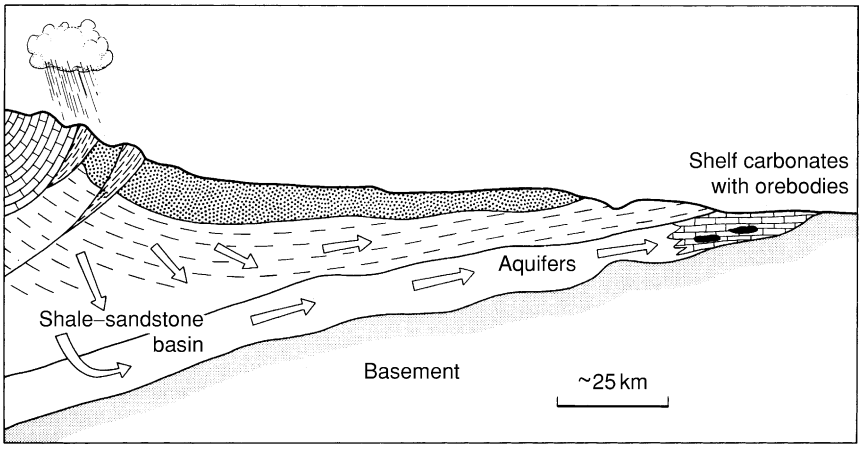


Figure 6 Schematic model for formation of carbonate-hosted Mississisppi Valley type Pb–Zn ores, via lateral gravity-driven flow of basinal brines from an upland region. (Reproduced with permission from Evans AJ (1993) *Ore Geology and Industrial Minerals: An Introduction*, 3rd edn. Oxford: Blackwell.)

compaction and overpressuring of basin sediments, promoting the flow of warm pore fluids.

Pore fluids in deforming sedimentary basins adjacent to uplifting forelands can be forced to flow laterally in response to gravity, a scenario that is best developed during continental plate collisions. A widely accepted theory for the origin of many carbonate-hosted Mississippi Valley-type (MVT) Pb–Zn deposits in central and eastern North America is westward gravity-driven flow of warm basinal brines, caused by Late Palaeozoic collisions between the ancient North American and European continents (Figure 6). Compositional and colour banding ('sphalerite stratigraphy'), fluid inclusions, and other geochemical features of the ores are remarkably consistent over long distances between the Pb–Zn deposits, supporting the occurrence of regional fluid-flow events. Gravity-driven lateral flow of basinal brines has emerged as the most effective mechanism that can account for the thermal and mass balances in many MVT systems.

Overthrusting and consequent sediment compression in fold-mountain belts may also drive the flow of deeper, warmer fluids into shallower or more peripheral environments. Shale-hosted stratiform copper deposits such as the Kupferschiefer (Europe) and White Pine (North America) may have formed via this type of mechanism.

Dehydration reactions associated with regional metamorphism can also provide sources of hot fluids, although they are initially low in salinity and high in CO_2. Such fluids may have been responsible for hydrothermal gold deposits associated with CO_2 metasomatism, found in shear zones and faults in Archaean and younger greenstone belts. Earthquake-induced seismic pumping could provide a mechanism of fluid flow along fault zones.

See Also

Economic Geology. **Gold**. **Minerals:** Native Elements; Sulphides. **Mining Geology:** Hydrothermal Ores. **Tectonics:** Hydrothermal Activity.

Further Reading

Barnes HL (1979) *Geochemistry of Hydrothermal Ore Deposits*, 2nd edn. New York: Wiley-Interscience.

Barnes HL (1997) *Geochemistry of Hydrothermal Ore Deposits*, 3rd edn. New York: John Wiley.

Barton PB, Rye RO, and Bethke PM (2000) Evolution of the Creede Caldera and its relation to mineralization in the Creede mining district, Colorado. In: Bethke PM and Hay RL (eds.) *Ancient Lake Creede; Its Volcano-Tectonic Setting, History of Sedimentation, and Relation to Mineralization in the Creede Mining District*, vol. 346, *Special Paper – Geological Society of America*, pp. 301–326.

Berger BR and Bethke PM (1985) *Reviews in Economic Geology, vol. 2: Geology and Geochemistry of Epithermal Systems*. El Paso, Texas: Economic Geology Publishing Company.

Cox DP and Singer DA (1986) *Ore Deposit Models*. US Geological Survey Bulletin 1693.

Craig JR and Vaughan DJ (1994) *Ore Microscopy and Ore Petrography*, 2nd edn. New York: Interscience.

Evans AJ (1993) *Ore Geology and Industrial Minerals: An Introduction*, 3rd edn. Oxford: Blackwell.

Giggenbach WF (1992) Magma degassing and mineral deposition in hydrothermal systems along convergent plate boundaries. *Economic Geology* 87: 1927–1944.

Guilbert JM and Park CF Jr (1986) *The Geology of Ore Deposits*. New York: WH Freeman.

Hedenquist JW and Lowenstern JB (1994) The role of magmas in the formation of hydrothermal ore deposits. *Nature* 370: 519–527.

Humphris SE, Zierenberg RA, Mullineaux LS, and Thompson RE (1995) *Seafloor Hydrothermal Systems: Physical, Chemical, Biological and Geochemical Interactions*. AGU Geophysical Monograph 91, Washington, DC: American Geophysical Union.

McKibben MA and Hardie LA (1997) Ore-forming brines in active continental rifts. In: Barnes HL (ed.) *Geochemistry of Hydrothermal Ore Deposits*, 3rd edn., pp. 875–933. New York: Wiley-Interscience.

Pirajno F (1992) *Hydrothermal Mineral Deposits: Principles and Fundamental Concepts for the Exploration Geologist*. Heidelberg: Springer-Verlag.

Rye RO (1993) The evolution of magmatic fluids in the epithermal environment: the stable isotope perspective. *Economic Geology* 88: 733–753.

Thompson JFH (1995) *Magmas, Fluids and Ore Deposits*. Mineralaogical Association of Canada, Short Course Series 23.

White DE (1981) *Active Geothermal Systems and Hydrothermal Ore Deposits*, pp. 392–423. Economic Geology Seventy-Fifth Anniversary Volume. El Paso, Texas: Economic Geology Publishing Company.

White NC and Herrington RJ (2000) Mineral deposits associated with volcanism. In: Sigurdsson H (ed.) *Encyclopedia of Volcanoes*, pp. 897–912. San Diego: Academic Press.

Magmatic Ores

J E Mungall, University of Toronto, Toronto, ON, Canada

Introduction

Magmatic ore deposits may be defined as rocks of igneous origin (*see* **Igneous Processes**), which can profitably be mined for their constituent chemical elements. For example, the worth of annual global production from magmatic ore deposits exceeded $10 billion in 2001, and dominates or contributes significantly to world supplies of elements representing more than a third of the periodic table. This article summarises the petrogenetic controls on the genesis of magmatic ore deposits, with an emphasis on the application of geochemical models. Readers are encouraged to consult the Further Reading Section at the end of this article for detailed descriptions of the multitude of types of magmatic ore deposits (Figure 1).

Fundamental Controls

The generation of a magmatic ore deposit depends upon the successful operation of four fundamental processes. Elements that are normally widely distributed at low concentrations must be extracted from a large volume of rock by a natural melting event. The resulting magma must be collected into or channelled through a relatively small volume of the Earth's crust, such as a conduit or magma chamber. While within the conduit or magma chamber the magma must become saturated with a phase within which the element of interest is highly concentrated. The phase containing high concentrations of the target element must then be mechanically sorted from the remainder of the magma, and collected in sufficient quantity and at sufficient grade to constitute an economically attractive ore deposit (*see* **Economic Geology**). The study of magmatic ore deposits is, therefore, inextricably tied to studies of the partitioning of elements between coexisting phases in magmas, and of the processes of intrusion and differentiation of magmatic systems (Figure 2).

Element Partitioning

The key to a quantitative approach to the description of the evolution of a magmatic system lies in the distribution of chemical components among the coexisting phases. The distribution of an element between two phases at equilibrium may be described through the use of a partition coefficient D, which is defined as the ratio of the concentration of the element in one phase (e.g., mineral or sulphide melt) to its concentration in another (e.g., silicate melt) (*see* **Minerals:** Sulphides); for example, the partition coefficient for Cu between sulphide and silicate melts can be expressed as

$$D_{Cu}^{sulphide/silicate} = \frac{C_{Cu}^{sulphide}}{C_{Cu}^{silicate}} \qquad [1]$$

where C denotes concentration of the subscripted element in the superscripted phase. Partition coefficients can be used in quantitative thermodynamic calculations if the element in question is a trace element whose concentration varies through ranges too small to affect D (Henry's Law is obeyed). If the element is a major or stoichiometric constituent of one of the phases, then D loses its utility as a constant. Values of

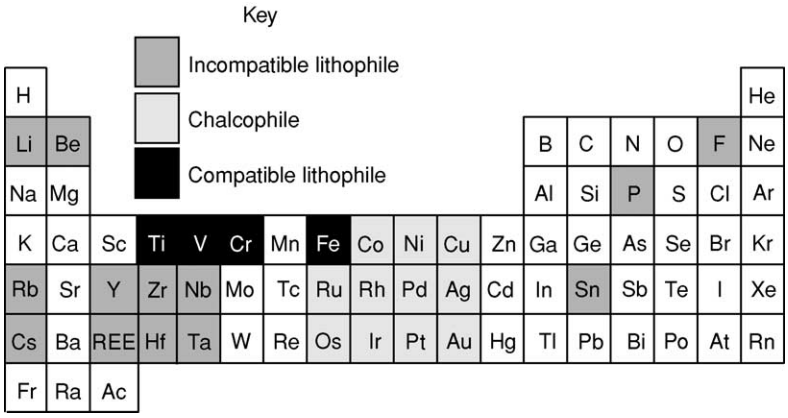


Figure 1 Elements extracted in significant quantities from magmatic ore deposits.

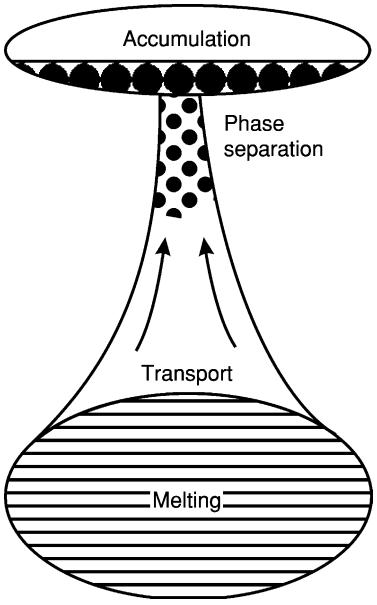


Figure 2 Processes required to generate a magmatic ore deposit.

D quoted in this contribution are intended to be used not as thermodynamic constants but rather as approximate indicators of the affinity of elements for various magmatic phases.

If two phases coexist at equilibrium in a closed system, eqn [1] can be combined with mass balance constraints to give concentration of an element in the melt as a function of the degree of partial melting:

$$\frac{C_i^{melt}}{C_i^{bulk}} = \frac{1}{D_i^*(1-F)+F} \qquad [2]$$

where C_i^{bulk} is the concentration of element i in the entire system, F is the fraction of melt present and D_i^* is a bulk partition coefficient for element i between solids and melt, obtained by summing over all solids the products of the individual phase/melt D and the weight fraction of the phases in the total solid assemblage.

Two processes that may occur during the differentiation of a magma are of critical importance to the generation of magmatic ore deposits. The first is the fractional removal of crystalline phases as the magma loses heat by assimilation of wall rock or by conduction. The second is the separation of an immiscible Fe-S-O liquid if the magma exceeds its carrying capacity for sulphide. Both processes govern the concentration and eventual segregation of ore metals from magmas after melting has occurred.

The effect of fractional crystallization on the concentration of a trace element with constant solid/liquid partition coefficient D_i^* is described by the Rayleigh distillation equation:

$$C_i^{melt} = C_i^o F^{(D_i^*-1)} \qquad [3]$$

where C_i^o is the initial concentration of an element and F is the proportion of liquid remaining. Application of eqn [3] to a highly incompatible element in the limiting case of complete solidification would cause the singular occurrence of infinite concentration of a trace element in the last vestiges of melt. What occurs instead is that, at some point, the concentration of a trace element rises to the point at which a mineral that concentrates that element begins to crystallize, effectively increasing D and avoiding the singularity. Depending on the solubility of the concentrating phase, and the value of the element in question, an ore deposit may be generated simply by the intrusion and solidification of a magma that contains sufficient concentration of the element solely through the action of fractional crystallization from a magma which originated by small degrees of partial melting of a

fertile source. Metals which begin to crystallise in a concentrating phase, when the bulk concentration of the element in the magma itself is has not yet attained ore grade, must undergo a final concentrating stage in which the mineral phase carrying the element is physically separated from the magma and collected in one place to form an ore body.

The effect of separation of sulphide melt from silicate melt is usually described using the assumption of batch equilibration of the two phases in a closed system:

$$C_i^{sulphide} = \frac{C_i^{bulk}(R+1)}{R + D_i^{sulphide/silicate}} \quad [4]$$

where R is the ratio of the mass of silicate melt divided by that of sulphide melt, and C and D are defined as above. Equation [4] can be derived from eqn [2] by redefining some of the terms, and is effectively identical to the equilibrium crystallization equation.

Based upon their relative affinity for the three principal types of phase present in magmatic systems, the elements shown in Figure 1 may be classed into three main groups (*see* Table 1). The chalcophile elements are partitioned strongly into sulphide melt at equilibrium with oxides or silicates, whereas the lithophile elements are those which reside in silicate or oxide phases in preference to metal or sulphide liquids. The lithophile elements may be subdivided into those which are partitioned into oxide or silicate minerals (the compatible lithophile elements) and those which are preferentially retained in silicate melt (the incompatible lithophile elements).

There is considerable overlap between the three groups of elements listed above; for example, V shows variable compatibility because it is an incompatible lithophile element during melting of the mantle, but becomes highly compatible when a magma is saturated with ilmenite or magnetite. Pd is strongly chalcophile, but in the absence of sulphide melt it behaves as an incompatible element.

Incompatible Lithophile Elements

Some combination of very small degrees of partial melting and fractional crystallization to very small residual melt fraction is required to elevate the concentrations of highly incompatible lithophile elements from their trace abundances in the bulk silicate Earth to the weight percent levels necessary for economic extraction. The principal host silicate magmas are granitic pegmatites or highly evolved alkaline or peralkaline felsic plutons. Carbonatites are a related class of rocks derived from carbonate magmas, either through liquid immiscibility with felsic alkaline rocks, or directly by extremely small degrees of partial melting of carbonate-rich mantle.

Incompatible lithophile element deposits may be subdivided into the Li-Cs-Ta (LCT) and the Nb-Y-F (NYF) classes. Deposits of the LCT class are almost exclusively granitic pegmatites associated with highly fractionated intrusions of reduced ilmenite-series (S-type) granites, whereas the NYF class tend to be derived from alkaline or peralkaline felsic intrusions.

LCT pegmatites are thought to originate by fluid-absent melting of pelitic or semipelitic metasedimentary schists in Abukuma-type regional metamorphic belts. Pelitic sediments are generally rich in alkali elements including Li and Cs, and also contain large amounts of F and B. Incongruent melting of white micas and complete melting of tourmaline produces water-poor peraluminous magmas rich in the fluxing components P, F, and B. The compositional undercooling caused by degassing of H_2O as water-rich magmas rise through the crust, causes them to

Table 1 Summary of partitioning behaviour of metals

Element group	*Examples*	D_i^*	*Concentrating phases*	D_i	*Magmatic environments of ore deposition*
Incompatible lithophile	Li, Be, F, Rb, Y, Zr, Nb, Cs, REE, Hf, Ta	<0.1	Melt, rare-element minerals	na	Pegmatite fields alkaline intrusions carbonatites
Variable compatibility	P, Ti, V, Cr, Fe	<0.1	Oxide minerals, apatite	10–100	Stratiform intrusions, podiform bodies immiscible oxide melts
Incompatible chalcophile	Rh, Pt, Pd, Au, Cu	<0.1	Sulphide melt	2000–50 000	Magmatic sulphides
Compatible chalcophile	Os, Ir, Ru, Ni	1–10	Olivine, oxide minerals	500–50 000	Chromitites magmatic sulphides

D_i^* is the bulk partition coefficient between mantle minerals and silicate melt during conditions typical of the generation of basaltic magmas. D_i is the partition coefficient for the phase which ultimately concentrates the element (e.g., sulphide melt for chalcophile elements, oxides for Ti, V, Cr). Elements with variable compatibility are those which are generally incompatible during melting but which are concentrated in minerals that crystallize from basaltic to intermediate magmas.

freeze at their eutectic before they can undergo extensive fractional crystallization. In contrast, the water-poor magmas parental to LCT pegmatites do not exsolve a free vapour phase, and do not evolve toward the water-saturated eutectic in the haplogranite system. Instead, LCT magmas are high-variance assemblages that can continue to evolve down to temperatures as low as 450°C, while continously becoming enriched in the fluxing elements F, B, and P as well as the trace elements including Li, Cs and Ta. The residual liquid may be enriched in the lithophile elements to such an extent that its bulk composition may be of ore grade. This magma can then be injected into the country rocks surrounding the cooling pluton and crystallized as swarms of rare-element pegmatite dikes (Figure 3).

The NYF suite of pegmatites and related plutons are probably derived by small degrees of partial melting of upper mantle similar to the sources of ocean island basalts, to produce alkaline magmas or magmas transitional from alkaline to tholeiitic affinity. Such small degree melts are enriched in incompatible elements like Nb, Zr, and Y, and can evolve through a prolonged process of fractional crystallization to peralkaline silica-saturated (pantelleritic) or silica-undersaturated (phonolitic) products. Deposits of the NYF suite of elements may form either directly by crystallization of the rare-element-rich magma. Alternately, the incompatible lithophile elements (e.g., Be, Nb, Y, Zr) may be transported by peralkaline fluids expelled from the cooling pluton, to be deposited where they mix with meteoric or formational waters surrounding the pluton.

There is an immense variety of rare-element pegmatites and related rocks, and there are undoubtedly a number of deposits whose genesis differs in significant ways from the models described here. However, the key requirements for any rare-element magmatic ore deposit must be the essential combination of small degrees of partial melting of fertile source rocks, and extreme fractional crystallization, both processes acting to concentrate these elements in progressively smaller volumes of silicate magma.

Carbonatites (*see* **Igneous Rocks:** Carbonatites) are igneous rocks that contain more than 50% carbonate minerals (*see* **Mining Geology:** Mineral Reserves). They host a great diversity of ore deposits, but the primary commodities extracted from these enigmatic rocks are rare earth elements (REE) and Nb (for which carbonatites are the principal source) as well as Fe, P, F, V, Zr, and Cu. Carbonatite may occur as small, usually concentrically-zoned plutons, or as lavas or pyroclastic deposits of regional extent.

Since carbonate is immiscible with alkaline CO_2-rich magmas, almost any alkaline or peralkaline magma has the potential to form an immiscible carbonatite liquid, given sufficient excess CO_2. Associated carbonatite and alkaline silicate magmas are found in tectonic settings including hotspots and convergent margins. The primary carbonatite magma is thought usually to be calcitic, and concentrates elements with high ionic field strength such as Nb, P, Zr, and F relative to coexisting silicate melt. Carbonatites evolve by fractional crystallization to dolomitic compositions which may become saturated with ore minerals such as pyrochlore, magnetite, and apatite. Late-stage carbonatite magmas may evolve to extremely sodic or ferroan compositions, and can crystallize important volumes of REE minerals such as monazite and bastnaesite within core zones or dykes cutting across carbonatite-alkaline rock complexes.

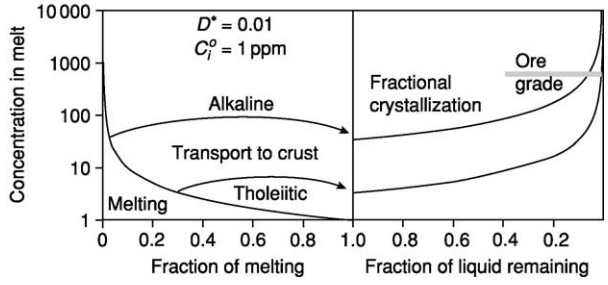


Figure 3 Formation of incompatible element deposits of the NYF class. Melting curve was calculated with eqn [2]; fractional crystallization with eqn [3], for a hypothetical element similar to Nb, Y, or Zr. Arrows represent the extraction of magmas from the mantle to the crust. Only alkaline magmas derived by small degrees of partial melting (left panel) can evolve through fractional crystallisation to very high concentrations of rare lithophile elements (right panel).

Oxides

Cumulates

Oxide minerals hosted by igneous rocks constitute important or principal sources of V, Ti, Cr and, to a lesser degree, Fe. There is considerable controversy surrounding the origins of many magmatic oxide deposits, between proponents of models for their deposition as cumulates from silicate magmas, or as oxide magmas derived by immiscibility with silicate magmas.

There is little doubt that the extremely refractory mineral chromite is deposited as a cumulus phase from silicate melts. Chromite deposits occur either as stratiform horizons in layered intrusions or as podiform accumulations in dunitic rocks. Mixing of two magmas cosaturated with olivine and chromite

will lead to oversaturation of the hybrid magma with chromite alone. Since primitive magmas commonly have only olivine and chromite on their liquidus, each magma recharge event in a large mafic magma chamber will lead to the temporary cessation of olivine accumulation, and the formation of a crystal layer composed solely of chromite. The thickness of the layer that forms will depend upon the temperatures and proportions of the magmas involved as well as the height of the magma column above the cumulate pile. Large layered intrusions, such as the Bushveld intrusion of South Africa, commonly contain numerous chromitite layers in their ultramafic basal series, the largest of which can be mined.

The mode of formation of podiform chromitites has been a more difficult conundrum than that of stratiform bodies. Podiform chromitites are found as wispy, nodular, or layered masses entirely enclosed within pipe- or dyke-like bodies of dunite in the harzburgitic mantle section of ophiolite complexes. Arai recently proposed that the same magma mixing mechanism used to explain the formation of stratiform chromitites can be applied to podiform chromitites. In his model, the more primitive magma is carried by porous flow through the dunite pipe. Melts that have migrated into successively lower-temperature environments farther from the core of the pipe become more evolved. Hybrid, chromite-oversaturated magma forms in places where the flow regime allows some of the cooler, more evolved melt to be drawn back into the core of the system. If the system reaches a steady state, chromitite pods can form and grow to considerable size in the zone of mixing between the two magmas.

Titanomagnetite can be deposited as stratiform horizons in large layered intrusions in much the same manner as chromite. Several processes have been proposed to explain how magnetite can arrive on the liquidus of a basaltic magma, to form monomineralic oxide layers in gabbroic cumulate rocks. The simplest involves magma mixing, analogous to the chromite saturation mechanism. Others include sudden changes in pressure or oxygen fugacity. Whatever the cause, once magnetite begins to crystallize from a basaltic magma, the transition metals Ti and V experience a sudden change from incompatible to strongly compatible behaviour. The first layer of magnetitite to form can constitute an ore of Ti and V, with Fe as an important byproduct. Because magmas form with a range of V and Ti contents, and because the partitioning behaviour of V is extremely sensitive to oxygen fugacity, there is a wide possible range of V abundance in magnetitites, and only a few are of economic grade. Stratiform accumulations of magnetitite are the world's primary repositories of V ore.

Immiscible Oxide Liquids

There are a number of deposits of Fe-Ti oxide (ilmenite and titanomagnetite with or without accessory apatite) worldwide that show textural and structural features, suggesting that they were emplaced as oxide magmas or lavas. There appears to be little doubt that the world's largest Ti deposits (Lac Tio, Canada; Tellnes, Norway) consist of intrusive sills and dykes composed of ilmenite, magnetite, and apatite. More controversial are the magnetite deposits of Kiruna, Sweden, and the El Laco deposit in Chile. In oxide rocks that have mesoscale textures and structures that perfectly replicate shallow subvolcanic and extrusive rocks, including pyroclastic flows, bombs, and lava flows, both deposits present textures closely resembling dendritic quench textures seen in slags. On the other hand, no experimental study has yet demonstrated that iron oxides can exist as liquids at reasonable magmatic temperatures. Furthermore, there is abundant evidence for marginal metasomatic replacement of common lavas and pyroclastic rocks by the same mineral assemblage as is found in the allegedly magmatic oxide rocks, raising the possibility that the oxide deposits may have formed by hydrothermal replacement of pre-existing silicate extrusive rocks. Controversy will likely continue for many years to come.

Chalcophile Elements

The upper mantle is thought to contain about 250 ppm of sulphur in the form of sulphide minerals. Under the conditions of formation of basaltic magmas, the sulphide will be entirely molten, so that the chalcophile elements will be sequestered within the sulphide melt. Under these conditions, the highly chalcophile elements will not enter the basaltic magma in significant quantities, and even the more moderately chalcophile elements such as Cu and Ni will be impoverished in the basalt. Since sulphide is sparingly soluble in basaltic magmas, progressively higher degrees of partial melting will dissolve progressively more of the sulphide. Upon approximately 25% partial melting of the mantle, all sulphide originally present will have been dissolved by the silicate melt, whereupon the chalcophile elements will now reside in a system containing only silicate minerals and silicate melt. Most chalcophile elements behave as incompatible lithophile elements in the absence of a free sulphide phase, so that at high degrees of partial melting of the mantle the chalcophile elements can be extracted into basaltic magmas and transported toward the surface. Such high degrees of melting are generally thought to be related to the melting of

mantle plumes (*see* **Mantle Plumes and Hot Spots**), and result in the extraction of picritic or komatiitic magmas (*see* **Igneous Rocks:** Komatiite).

Suprasubduction zone magmas typically remain sulphide-saturated during melting due to the introduction of sulphide to their source regions by the same fluids that induce partial melting (Figure 4). At best, arc magmas initially saturated in sulphide can generate low-grade Ni-Cu deposits with exceedingly low precious-metal contents.

Once a magma separates from its source and begins to rise toward the surface, it will fall into a state of sulphide-undersaturation due to the inverse dependence of sulphide solubility on pressure in silicate melts. When deeply-sourced mafic or ultramafic magmas reach the Earth's surface, they typically contain less than half the amount of sulphide necessary to induce sulphide saturation.

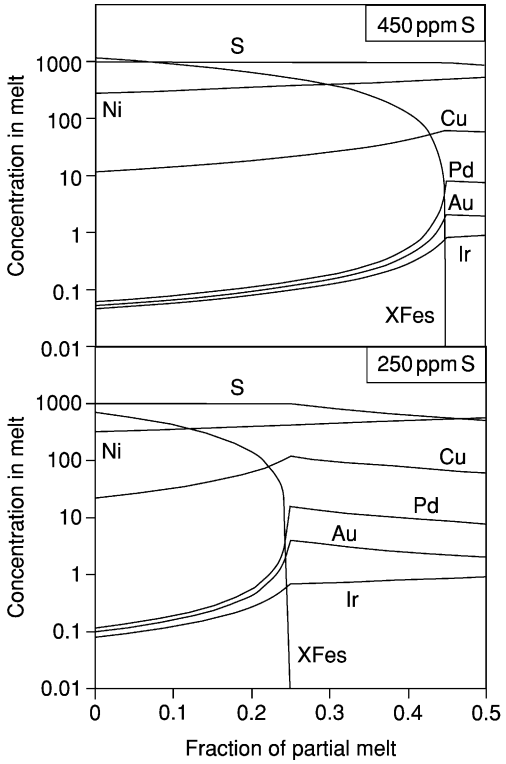


Figure 4 Concentration of chalcophile elements in basaltic magma during partial melting of mantle containing either 450 ppm S (metasomatized sub-arc mantle; upper panel) or 250 ppm S (plume head melting; left panel). Chalcophile element concentrations increase as the amount of sulphide remaining in the source (XFes) diminishes. At the moment when sulphide is completely consumed, incompatible chalcophile elements reach their maximum possible concentrations, whereas compatible chalcophile elements (Ni, Ir) continue to increase with further melting.

Magmatic Sulphide Deposits

Magmatic sulphides represent accumulations of immiscible sulphide liquid separated from silicate magmas, and account for the majority of the value of metal produced from magmatic ore deposits worldwide. Although all magmatic sulphide deposits are intrinsically polymetallic concentrations of the base metals Fe, Ni, Cu, and the precious metals Au, Ag, and the platinum group elements (PGE; especially Pt and Pd), it is convenient to categorise them as either precious metal or base metal deposits, each of which yields the elements of the other group as important byproducts.

Precious metal sulphide deposits generally contain less than 1 modal % sulphide, whereas base metal sulphide deposits contain 5 to 100 modal % sulphide. The distinction is an arbitrary one, but it is a consequence of the extreme sensitivity of precious metal concentrations to the mass ratio of silicate to sulphide melt that equilibrate during sulphide segregation, as illustrated in Figure 5. Precious metal sulphide deposits typically form where very small amounts of sulphide have separated from silicate melt and collected in stratiform bodies within large mafic intrusions. Base metal-dominated magmatic sulphide deposits form by the collection of immiscible sulphide liquid

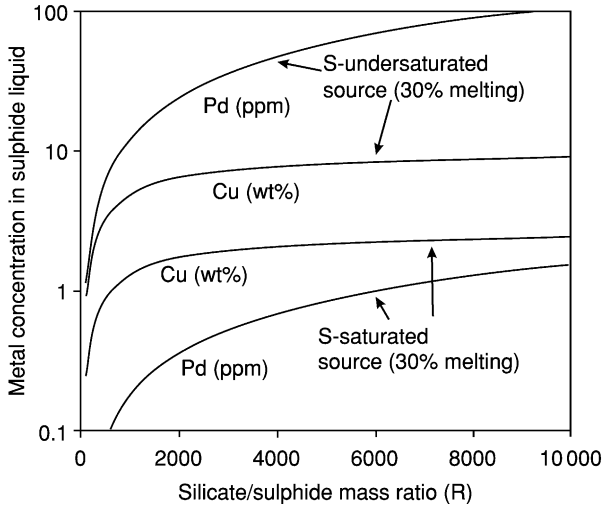


Figure 5 Concentrations of Cu and Pd in sulphide liquid equilibrated with basaltic liquids as functions of the silicate/sulphide mass ratio (R in eqn [3]). The curves are calculated assuming that the initial basalt composition resulted from 30% partial melting of the upper mantle, using the compositions for the sulphide-saturated and sulphide-undersaturated cases shown in Figure 4. Sulphide-rich systems ($100 < R < 1000$) can generate large quantities of base-metal-rich sulphide which contains relatively low PGE contents. At very high values of R the resulting sulphide is PGE-rich, but correspondingly smaller in volume. Magmas derived from sulphide-saturated mantle sources can only generate low-grade base-metal deposits without significant PGE.

from mantle-derived mafic, or ultramafic silicate magmas in lava channels, magma conduits, or places where conduits feed into larger bodies. The Sudbury mining camp, Canada, the world's largest collection of Ni-Cu-PGE deposits, is a unique exception. The Sudbury Igneous Complex and its base-metal sulphide ore deposits formed by the batch equilibration and gravitational segregation of sulphide liquid with several thousand km^3 of shock-melted continental crust following a large meteorite impact into a shallow marine basin filled with sulphidic sediments.

Base Metal Sulphide Deposits

A more common path to sulphide saturation in fertile, previously sulphide-undersaturated magmas is the assimilation of crustal rocks rich in sulphide. This has the combined effect of lowering the temperature of the magma while sharply increasing the amount of sulphur it contains. Addition of large quantities of sulphide can drive a magma into a state of extreme sulphide-oversaturation, so that the sulphide equilibrates at a relatively low silicate/sulphide ratio ($100 < R < 5000$; eqn [3]). The result can be the generation of sulphide in sufficient quantity to separate and form economically attractive tonnages of the base metal sulphides, however, at these relatively lower silicate/sulphide ratios, the concentrations of the precious metals in the resulting sulphide tend to contribute a relatively small part of the total value of the ores.

Emplacement of magmatic sulphides The dominant mode of transport of magmas in the crust is by the emplacement of dykes and sills. Figure 6 provides schematic illustrations of the most important mechanisms for sulphide segregation. Two phenomena related to dyke and sill emplacement are of particular importance in the generation of magmatic sulphide deposits. These are the formation of long-lived conduits, and the formation of channelized lava flows. Because thin, tabular sheets of magma cool rapidly, flow along dykes and sills tends to become focused along their widest portions, which then enlarge themselves by heating and assimilating their wall rocks. The result is the generation of long-lived blade-shaped or tubular magma conduits in which flowing magma cools relatively slowly. Similarly, large fissure-fed lava flow fields are thought to develop distributary systems resembling deltas, in which deeply incised lava channels proximal to the vent feed smaller branching channels and distal sheet or pillowed flow lobes. Within either a conduit or a lava channel, it is possible for a steady state to be reached, in which the temperature of magma passing a given point remains constant for extended periods of time. If the rate of supply of sulphide by degassing or erosion of the channel floor or conduit walls, and the rate of cooling and crystallization also approach a steady state, then sulphide liquid first reaches saturation in the magma always as it passes a single point within the conduit or channel. Sedimentation of sulphide droplets from the flowing lava, perhaps also with a co-saturated phenocryst phase (olivine or pyroxene), can thus lead to highly efficient sorting of the phase and the elements it concentrates from a very large volume of magma into a small, localized volume of rock.

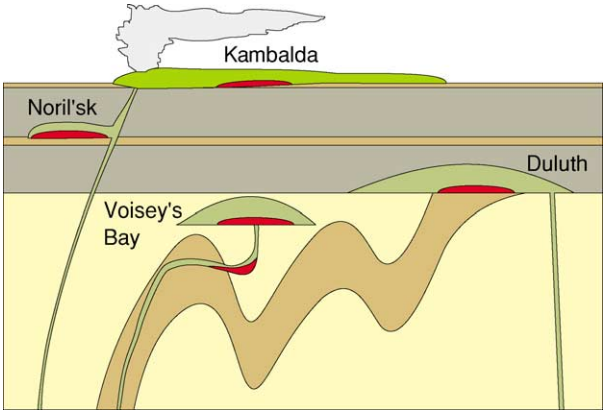


Figure 6 Cartoon of principal mechanisms for the generation and segregation of magmatic base-metal sulphides. The image shows a cross-section of the upper crust, with brown areas representing sulphur-rich crustal rocks which act as sources of sulphur for dykes and plutons of mantle-derived magma (darker green) or lava flows (lighter green). Sulphide liquid collects where it formed (Duluth), in traps within conduits or where conduits enter magma chambers (Voisey's Bay), in subvolcanic sills (Noril'sk), or in lava channels cutting through sulphidic sediments (Kambalda).

A related possibility with a similar result is that a magma which has become saturated with sulphide liquid may carry the exsolved sulphide suspended as small droplets in the rapid and turbulent flow within the conduit. A transition within the dyke to laminar flow or a sudden drop in flow velocity due to its passage through a magmatic breccia or a widening of the conduit will then cause the entrained sulphides to settle together in one place within the conduit.

In practice it may be impossible to distinguish between the two sulphide collection mechanisms described above. Classic examples of valuable sulphide deposits accumulated from flowing magma in conduits or lava channels are given in Table 2, with the Sudbury Camp for comparison.

If sulphides are deposited in the place at which sulphide saturation occurred, without being transported and collected from a large volume into a small concentration, then a low-grade disseminated

Table 2 Characteristics of major base-metal sulphide deposits*

Deposit or camp	Structural setting	Host magma type	Grades		Production and reserves (Mt)
			Ni %	Cu %	
Sudbury, Canada	Meteorite impact crater	Quartz diorite	1.2	1.0	>1600
Thompson, Canada	Magma conduit	Komatiite	2.5	<0.5	90
Voisey's Bay, Canada	Magma conduit	Picrite	2.8	1.7	32
Kambalda, Australia	Lava channel	Komatiite	3.5	<0.5	50
Perseverance, Australia	Lava channel	Komatiite	0.6	<0.5	270
Duluth, USA	Base of large intrusion	Continental tholeiite	0.2	0.7	>4000
Noril'sk, Russia	Magma conduit	Continental tholeiite	2.7	2.1	>500
Pechenga, Russia	Magma conduit	Ferropicrite	1.0	<0.5	30
Jinchuan, China	Magma conduit	Continental tholeiite	1.1	0.7	>500

*Data sources: Production and reserves from www.minecost.com; and Eckstrand, O.R., Magmatic nickel-copper platinum group elements (Eckstrand OR, Sinclair WD, and Thorpe RI (eds.) Geology of Canadian Mineral Deposit Types, Geological Survey of Canada, *Geology of Canada* 8: 583–605).

deposit like those of the Duluth Complex (USA) will result.

Precious metal magmatic sulphide deposits An initially sulphur-undersaturated magma may become sulphur-saturated because crystallization of silicate minerals causes an enrichment of sulphur in residual melt while cooling simultaneously lowers the solubility of sulphide. Highly evolved basaltic magmas may thus passively evolve into a state of sulphide saturation in large magma chambers, causing the separation of extremely small fractions of sulphide melt, as may have occurred in the Great Dyke of Zimbabwe and the Munni Munni intrusion of Australia. Recharge of a large magma chamber by a fresh pulse of magma is another popular explanation for the sudden onset of sulphide saturation in large magma chambers (Figure 7). However, the magma body does become saturated with sulphide, removal of sulphide melt at high R value (eqn [4]) from a silicate magma barely saturated in sulphide may efficiently scavenge precious metals from the entire magma chamber into a single thin stratabound layer. The sulphides thus generated will also be highly enriched in the base metals, but at total sulphide abundances on the order of 1 modal %, the deposits are only valuable for their precious metals.

An alternate possibility for the generation of stratiform sulphide horizons in a layered intrusion involves migration of deuteric fluids in the cooling cumulate pile (Figure 6). As the initially vapour-undersaturated intercumulus melt cools and approaches its solidus it will eventually exsolve an aqueous fluid. Orthomagmatic fluids are capable of dissolving considerable amounts of precious metals and sulphur as they rise through the cumulate pile. If such fluids remain sulphide-saturated and equilibrate with the small quantities of sulphide that exist within the intercumulus liquid, then they will carry

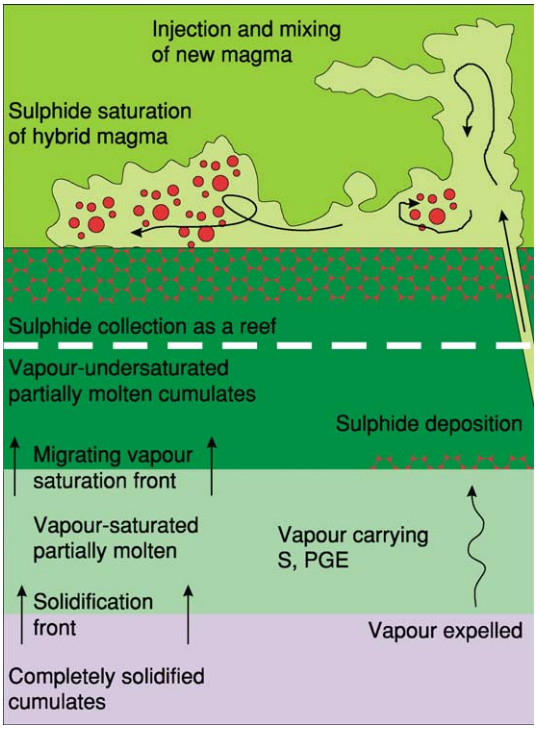


Figure 7 Schematic view of two popular models for the generation of reef-type PGE sulphide mineralization in layered intrusions. The orthomagmatic model proposed by Campbell and co-workers is illustrated above the white dashed line. The deuteric fluid model proposed by Boudreau and coworkers is illustrated below the line.

considerable amounts of dissolved sulphide and precious metals upward through the pile of cumulates. When such fluids reach the top of the vapour-saturated zone and encounter vapour-undersaturated intercumulus liquid, they will redissolve within it. Dissolution of the vapour phase will cause reprecipitation of the entrained sulphide and precious metals at the level of the transition to vapour-undersaturated conditions. As a vapour-saturation surface migrates

upward through a cumulus pile it will bring with it a sulphide- and precious metal-rich zone akin to a chromatographic front. Such fronts may be frozen in place by the eventual solidification of the cumulate pile, to form stratiform PGE-rich sulphide horizons.

Although there is a potential for the generation of a precious metal sulphide deposit wherever small volumes of sulphide liquid reach equilibrium with large volumes of formerly sulphur-undersaturated silicate magma, nearly all production from precious metal sulphide deposits comes from just two large layered intrusions; the Bushveld complex of South Africa and the Stillwater intrusion of the USA. About half of world production of Pd is generated as a by-product of base metal sulphide mining at Noril'sk in Russia.

See Also

Economic Geology. **Igneous Processes**. **Igneous Rocks:** Carbonatites; Komatiite. **Mantle Plumes and Hot Spots**. **Minerals:** Sulphides. **Mining Geology:** Mineral Reserves.

Further Reading

Boudreau AE, Mathez EA, and McCallum IS (1986) Halogen geochemistry of the Stillwater and Bushveld complexes: evidence for transport of platinum-group elements by Cl-rich fluids. *Journal of Petrology* 27: 967–986.

Campbell IH and Naldrett AJ (1979) The influence of silicate:sulphide ratios on the geochemistry of magmatic sulphides. *Economic Geology* 74: 1503–1505.

Campbell IH, Naldrett AJ, and Barnes SJ (1983) A model for the origin of the platinum group rich sulphide horizons in the Bushveld and Stillwater complexes. *Journal of Petrology* 24: 133–165.

Crowson P (2001) *Minerals Handbook 2000–2001*. Mining Journal Books.

Drew LJ, Qingrun M, and Wiejun S (1990) The Bayan Obo iron-rare-earth-niobium deposits, Inner Mongolia, China. *Lithos* 26(1/2): 43–65.

Eckstrand OR, Sinclair WD, and Thorpe RI (eds.) (1995) *Geology of Canadian Ore Deposit Types*. Geological Survey of Canada.

Keays RR, Lesher CM, Lightfoot PC, and Farrow CEG (eds.) (1999) *Dynamic Processes in Magmatic Ore Deposits and their Application in Mineral Exploration*. Short Course Volume 13. Geological Association of Canada.

Kolker A (1982) Mineralogy and geochemistry of Fe-Ti oxide and apatite (nelsonite) deposits and evaluation of the liquid immiscibility hypothesis. *Economic Geology* 77: 1146–1158.

London D (1992) The application of experimental petrology to the genesis and crystallization of granitic pegmatites. *The Canadian Mineralogist* 30: 499–540.

Naldrett AJ (1989) *Magmatic Sulphide Deposits*. New York, Oxford: Clarendon Press–Oxford University Press.

Naslund HR (1983) The effect of oxygen fugacity on liquid immiscibility in iron-bearing silicate melts. *American Journal of Science* 283: 1034–1059.

Reynolds IM (1985) The nature and origin of titaniferous magnetite-rich layers in the Upper Zone of the Bushveld Complex: a review and synthesis. *Economic Geology* 80: 1089–1108.

Roberts RG and Sheahan P (eds.) (1988) *Ore Deposit Models*. Geoscience Reprint Series 3. Geological Association of Canada.

Whitney JA and Naldrett AJ (eds.) (1989) *Ore Deposition Associated with Magmas*. Society of Economic Geologists. *Reviews in Economic Geology* Vol. 4.

MOHO DISCONTINUITY

P Giese, Freie Universität Berlin, Berlin, Germany

Introduction

A century ago, instrumental seismology began to reveal the internal structure of Earth. Analysing seismic data in 1909, following an earthquake 39 km south-east of Zagreb, the Austro-Hungarian meteorologist Adrija Mohorovičić discovered the boundary between the crust and mantle of south-eastern Europe. During the following decades, it turned out that this discontinuity, now known as the 'Mohorovičić discontinuity' (or simply 'Moho-discontinuity' or 'M-discontinuity' or Moho), is present worldwide in continental as well as in oceanic regions. The following discussion focuses on the morphology, structure, and age of the European Moho-discontinuity and the bearing that these features have on the structure and evolution of the crust. The data are largely based on the results of national and international projects carried out during the past 50 years. The European Geotraverse (EGT) project, from 1983 to 1990, involved studies of the European lithosphere along a north- and south-trending corridor from the North Cape of Scandinavia down to the northern coast of Africa (Tunisia), an ambitious

enterprise, in that it attempted to resolve a geodynamic evolution from ~3.5 billion years ago to the present. The EGT project was followed by the EUROPROBE project, which aimed at analysing the structure, properties, and evolution of 'deep Europe' and exemplary cases of crustal accretion, thickening, and extension. The summary of data from these projects covers the geophysical tools used, the petrological significance of the data, and the main structural features of the European Moho-discontinuity.

Geophysical Tools

Geophysical methods are indispensable for the recognition and interpretation of structures in Earth's crust and mantle. Among the numerous geophysical methods suitable to explore structure and composition of the lithosphere, seismic techniques provide the best tools for detailed crustal and upper mantle investigations (*see* **Seismic Surveys**). In explosion seismology, the use of deeply, nearly vertical, reflected waves (P waves) in the subcritical range provides detailed information about the internal structure of the crust and the uppermost mantle (Figure 1A). Another method is based on recording P waves termed 'PmP' wide-angle arrivals; these are recorded as overcritical reflected waves from the Moho-discontinuity and/or as diving waves generated by a strong velocity gradient zone in the model of the crust/mantle transition (Figure 1B). Tomographic studies using travel times of waves generated by earthquakes or by big explosions allow determination of the velocity field of the lithosphere (Figure 1C). During the past decade, a new 'receiver function method' has been developed (Figure 1D). It is based on earthquake records of broadband stations and makes use of P-to-S converted phases that are created at the Moho-discontinuity or other prominent boundaries in the lithosphere. For the study of the lithosphere/asthenosphere system, analysis of surface waves is another powerful seismic tool. The data discussed here are mainly based on recordings of P waves (compressional waves).

Nature and Stability of the Moho-Discontinuity

From a seismological point of view, the Moho-discontinuity is defined as the depth at which the velocity v_P of compressional waves increases from normal crustal velocity values of $<7.2\,km\,s^{-1}$ to typical upper mantle values of $>7.6\,km\,s^{-1}$. Modern studies see the Moho-discontinuity as a sandwich-like mix of crust and mantle materials (i.e., low- and high-velocity materials). In addition, the Moho-discontinuity is not only a compositional boundary,

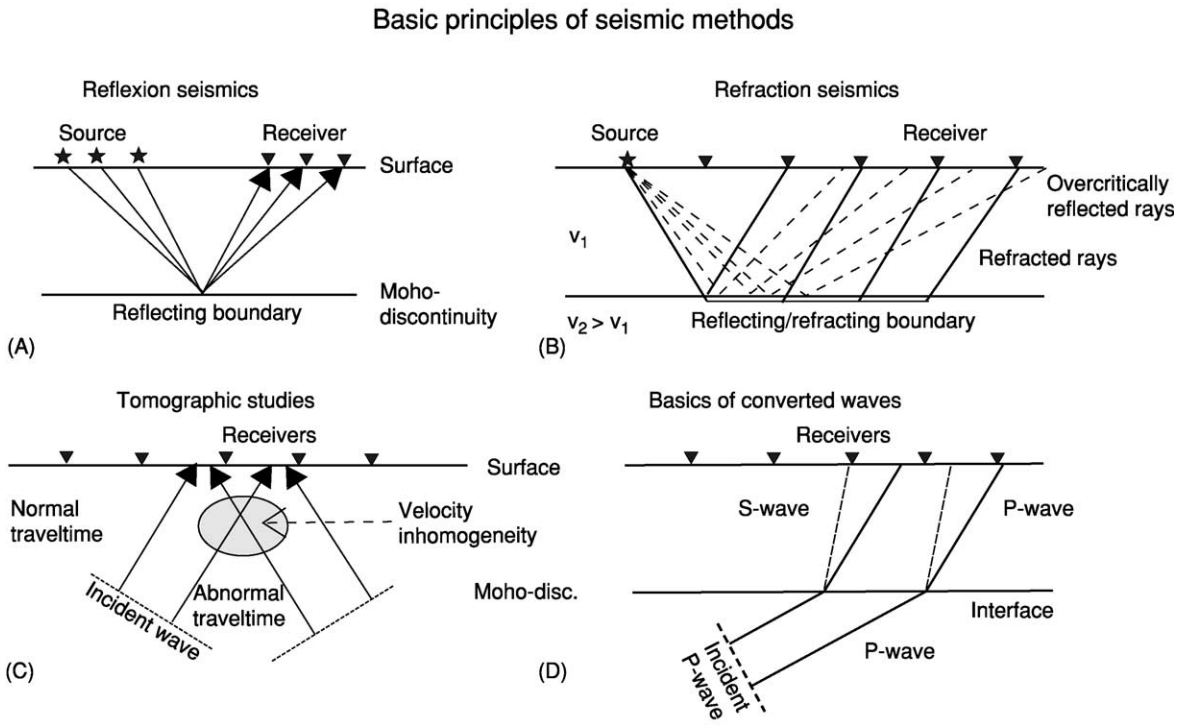


Figure 1 Basic principles of deep seismic sounding methods: (A) reflection seismology, (B) refraction seismology, (C) tomographic seismology, and (D) receiver function analysis.

but is additionally characterized by an abrupt change in the length scale of heterogeneities between the lower crust and upper mantle, and therefore also images various geodynamic processes (Figure 2). In many regions, the Moho depth derived from reflected waves generated at the base of the lower crust coincides with the depth of the Moho-discontinuity as determined from seismic refracted and wide-angle reflected waves. The difference between both results does not exceed a few kilometres. On the other hand, petrologists define the Moho-discontinuity between the crust and the uppermost mantle as a petrological boundary. The lower crust consists of high-temperature/high-pressure metamorphic rocks (granulite facies) with an average mafic composition approaching that of a primitive basalt (*see* **Ultra High Pressure Metamorphism**). As evidenced by mantle fragments brought to the surface by volcanism (xenoliths), the lower crust shows an interlayering of felsic and intermediate lithologies, causing a high degree of seismic reflectivity in the lower crust, as observed in many regions. The layer beneath the Moho-discontinuity, the uppermost mantle, is built up of rocks rich in olivine (*see* **Minerals:** Olivines).

The ambiguity between the geophysical and petrological definitions of the Moho-discontinuity causes problems of interpretation. If lower crustal rocks such as pyroxene granulite or gabbro are subducted to depths below 50–60 km, they are metamorphosed into the high-pressure rock type, eclogite, which shows the same density (3300 kg m^{-3}) and the same P-wave velocity (>7.6 km s^{-1}) as normal mantle rocks (peridotites) show. From this geophysical perspective, eclogites would be interpreted as upper mantle rocks underlying the Moho-discontinuity. Geological, petrological, and geophysical studies reveal that the Moho-discontinuity is not a permanently stable boundary and that it can be destabilized during all stages of the orogenic cycle by tectonic, magmatic, and metamorphic processes:

- Collision of continental blocks causes crustal thickening, associated with the formation of a surface topography and the generation of crustal and lithospheric roots. Intrusion of magmas may also cause crustal thickening.
- When the crust of a thickened mountain belt is reduced by erosion, the crustal root will be buoyantly uplifted until the crust has attained the same thickness as in the surrounding areas.
- When deeper levels of the crust are heated by the underlying mantle, crustal rocks may become mechanically soft and the crustal root may spread laterally over the Moho-discontinuity, much like oil on water.
- A pseudo-thinning of the crust may be caused by metamorphic processes in the lower crust, such as the transformation of gabbroic rocks into eclogite. Such 'pseudo-mantle' rocks might represent portions of subducted and metamorphosed lower crust, so that the true (petrological) crustal thickness would be higher than expected from the study of the recent seismic velocity distribution. Furthermore, such high-density layers may delaminate and sink downward into the deeper lithosphere and asthenosphere due to their negative buoyancy.

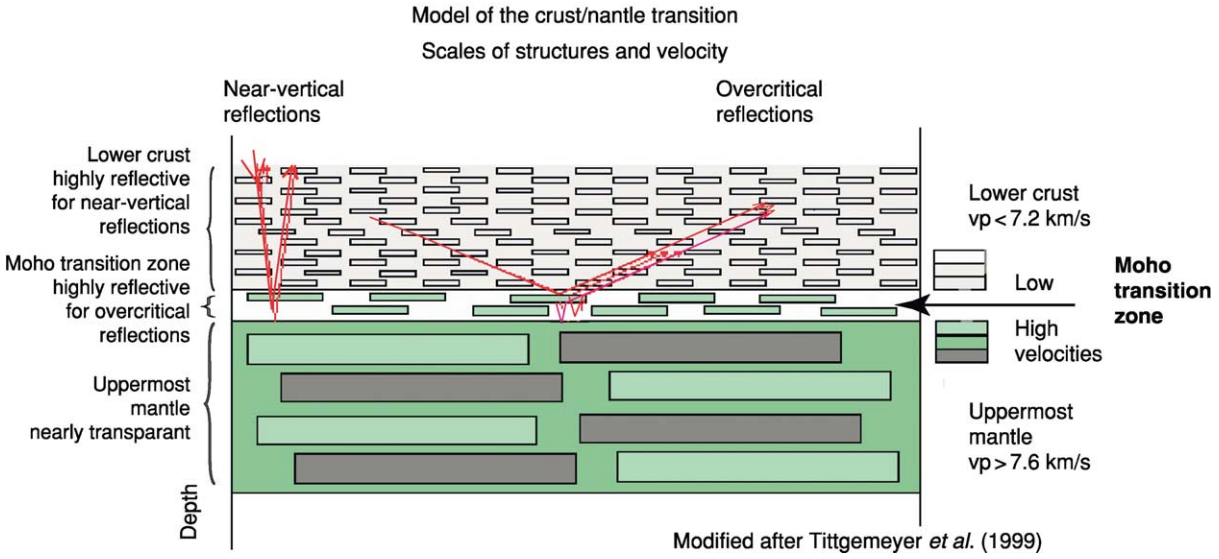


Figure 2 Material properties around the continental Moho-discontinuity and their response to seismic signals. The thickness of the transition zone between the lower crust and the upper mantle may vary between ~100 m and a few kilometres. Modified from Tittgemeyer *et al.* (1999).

- The thickened orogenic crust may be thinned by syn- to post-orogenic extension driven by mantle convection, which creates new continental rifts and oceanic basins.

This brief overview demonstrates that the Moho-discontinuity is a sensitive indicator for the detection of past and recent geodynamic processes.

Main Tectonic Units of the European Continent

Present-day Europe forms the western part of the large continental Eurasian Plate. It is bordered in the east by the mountain belt of the Urals and in the north and west by the Atlantic Ocean, which opened in the Triassic. The southern border of Europe is formed by the Mediterranean Sea, the Black Sea, and the Caspian Sea (Figure 3). The growth of the European continent started in the northern region of the Baltic Shield, and the European crust becomes younger and younger from north to south and from east to west. Five main tectonic domains can be distinguished: (1) The East European (Russian) Platform and the Baltic Shield, the north-western part, were formed during Precambrian times, 3.6 Ga to 660 Ma (*see* **Europe:** East European Craton). The western margin of the East European Platform is formed by the complexly structured Trans-European Suture Zone (TESZ), which separates the Variscan mobile belts of Central Europe from the stable platform of Eastern and Northern Europe. This suture zone extends over 2000 km from the North Sea down to the north-western Black Sea. (2) The north-western margin of the Baltic Shield is marked by the Caledonide Orogen, which was formed during the Late Silurian to Early Devonian (*see* **Europe:** Scandinavian Caledonides (with Greenland)). The Caledonides continue into northern parts of England and Ireland (*see* **Europe:** Caledonides of Britain and Ireland). (3) On the eastern side of the East European Platform, the Uralian Orogen (380–250 Ma) marks the tectonic boundary

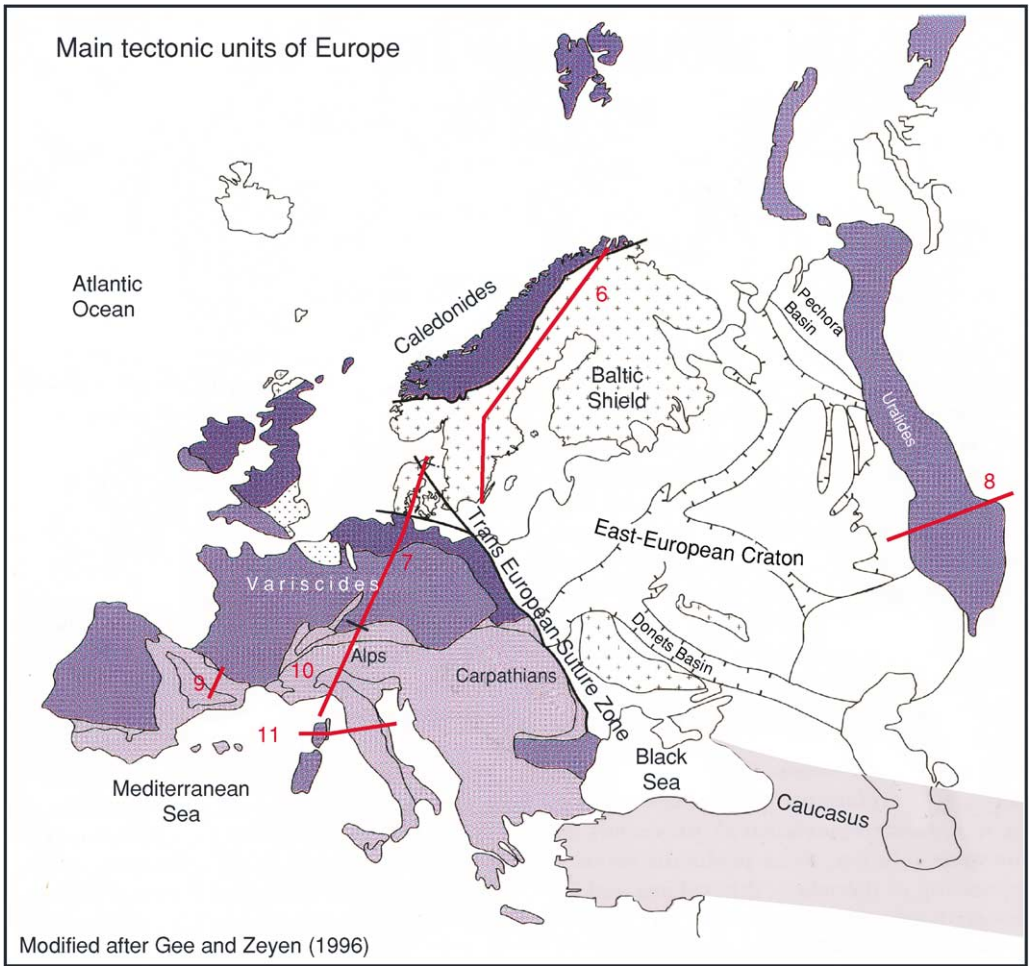


Figure 3 Main tectonic units of Europe. The position of the crustal cross-sections described in this paper is marked by red lines. Modified from Gee and Zeyen (1996).

between the East European and the Siberian Plates (*see* **Europe:** The Urals). (4) Central and Western Europe were formed during the Cadomian (~620–540 Ma) Orogeny and severely were overprinted by the Variscan (or Hercynian) Orogeny (400–295 Ma) (*see* **Europe:** Variscan Orogeny). (5) In Southern Europe, the Cenozoic Alpine–Mediterranean mountain belts resulted from the convergence between Africa and Europe during Mesozoic to Cenozoic times (220 Ma). The Mediterranean Sea contains relicts of the oceanic crust of the Tethys and young back-arc basins that are partly built up by young oceanic crust (*see* **Europe:** Mediterranean Tectonics). Besides these compressional structures, there are also some prominent extensional regions in Europe. The formation of the North German and Polish basins and the evolution of the Central European Rift System were generated by post-Variscan extensional processes (*see* **Europe:** Permian to Recent Evolution).

Main Features of the Moho-Discontinuity in Europe

During the past five decades, a large amount of seismic data (active and passive seismology) has provided detailed information on regional and local crustal structures and resulting geophysical anomalies. Figure 4 is a contour map of the depth of the Moho-discontinuity of the Eurasian plate, extending from the Atlantic Ocean in the west to the Siberian Platform in the east. A more detailed Moho contour map of Central, Western, and Southern Europe is presented in Figure 5, which is based on seismic data measured in Europe from 1958–1988. This map comprises large parts of Pre-Caledonian, Caledonian, Variscan, and Alpine domains in Europe. There are very pronounced variations of crustal thickness within and between the different tectonic units. The most prominent structural features are as follows:

- The crust of the Baltic Shield and the East European Platform shows a rather large thickness, with an average of 40–50 km.
- The Trans-European Suture Zone is one of the most important geological boundaries in Europe north of the Alpine Orogen. This suture zone forms a graben-like structure with a thickened crust and a depth of the Moho-discontinuity between 45 and 55 km.
- In contrast to the Baltic Shield and the East European Platform, the depth of the Caledonian and Variscan Moho-discontinuity in Central and

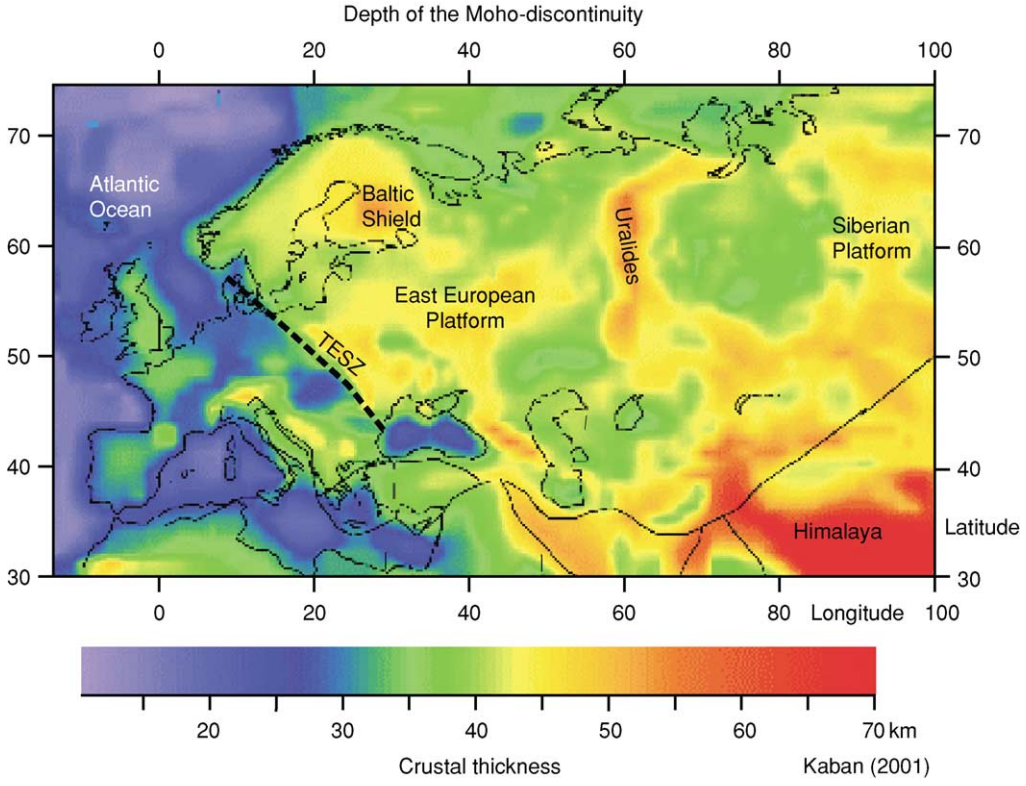


Figure 4 Crustal thickness of the European continent and the West Siberian platform, derived from numerous seismic profiles (mainly from deep seismic sounding studies) and gravity data. Modified and reproduced from Kaban (2001).

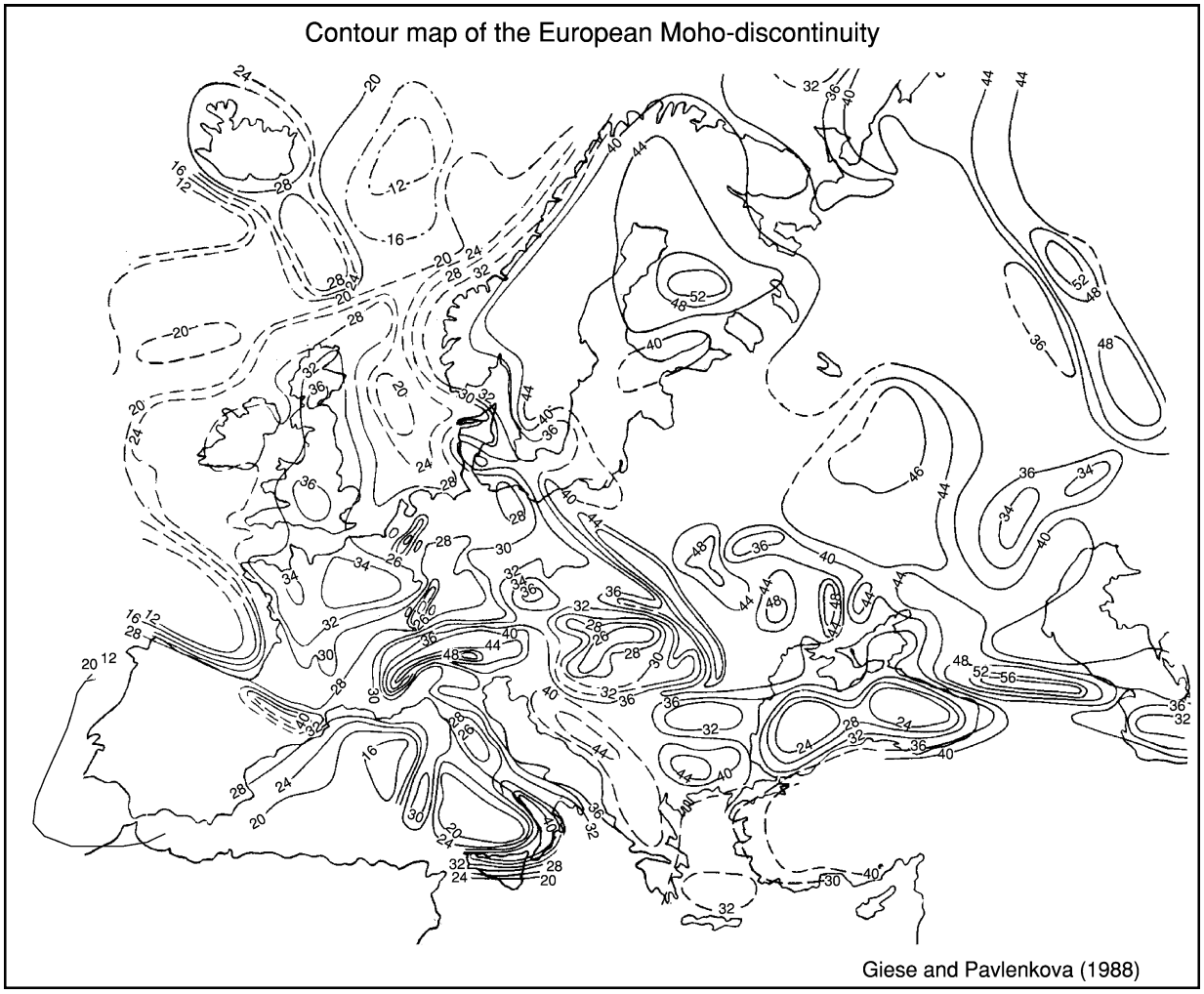


Figure 5 Morphology of the Moho-discontinuity in Central, Western, and Southern Europe. The contours (spacing, 2 to 4 km resp) reveal the main geological structures of young orogens such as the Alps, the Apennines, and the Pyrenees. Giese and Pavlenkova (1988).

Western Europe measures only 30–35 km. No crustal roots are recognizable.

- The picture is distinctly different in the Variscan Uralides, where the crustal root or parts of the root have been preserved since the Late Palaeozoic. In the root zone, the Moho-discontinuity has been detected at a depth of 50–60 km.
- The southern (Mediterranean) region of Europe presents a very heterogeneous set of crustal structures. This region is built up of a system of several Cenozoic orogenic belts with thick mountain roots up to 60 km deep.

Other regions were formed by extensional processes and are characterized by a shallow Moho-discontinuity. Extension in continental crust is documented in the Permo-Mesozoic basins of Northern Germany and Northern Poland, in the Cenozoic European rift system. The Cenozoic Pannonian Basin with a crustal thickness of only 24 km was formed by back-arc spreading. Oceanization of continental crust is also observed in back-arc regions of the western Mediterranean, where the Moho-discontinuity is found at a depth of only 10 km. Atlantic Ocean crust borders the western and north-western margin of the European continent. Oceanic crust is also present in the central Black Sea and in the Bay of Biscay, which was formed by counterclockwise rotation of the Iberian peninsula with respect to mainland Europe.

Main Structural Features of Pre-Variscan and Variscan Europe

Northern and Eastern Europe

Precambrian Europe is built of the Baltic Shield and the East European Platform (*see* **Europe:** East European Craton; Scandinavian Caledonides (with

Greenland)). The Baltic Shield is a collage of cratonic blocks and island arcs. The orogenic belts resulting from multiple collision have been levelled out by erosion at the surface and partial re-equilibration of the Moho-discontinuity. Seismic studies reveal variations of crustal thickness ranging from 40 km in Sweden to more than 55 km in southern Finland. This anomalously large crustal thickness is mainly caused by a relatively thick lower crust (15–20 km). Figure 6 shows a north to south crustal cross-section through the Baltic Shield between the North Cape and the southern tip of Sweden at the Baltic Sea. Along this section, the average crustal thickness is about 40 km. Lateral inhomogeneities with several distinct offsets of the Moho-discontinuity may be related to fault zones in the upper crust. These probably reflect amalgamation of microplates that were derived from the fragmentation of Late Archaean megacontinents and that were welded together along Proterozoic collisional belts. The large thickness of the Baltic Shield and the Eastern European Craton requires an average crustal density of 2900–3000 kg m^{-3} By comparison, the average crustal density in Central Europe amounts only to about 2800 kg m^{-3}. Thinned crust is present only in the Dniepr–Donets Basin, which is a product of intracratonic rifting in the Devonian and Early Carboniferous times. Parts of the Caledonian belt are exposed along the coast of Norway and in Scotland and central England. Another branch of the Caledonides is suspected beneath the Northern German and Polish basins. The recent average crustal thickness is about 28–34 km and no orogenic roots are visible (Figure 5). As in the Variscides, the Caledonian crust is distinctly thinner than that of the Baltic Shield, which again suggests crustal thinning by delamination of lower crust.

The Trans-European Suture Zone

The Trans-European Suture Zone is one of the most prominent geological boundaries in Europe. It separates the East European Platform from the mobile Variscan Domain in Central Europe. This zone, with its graben-like structure and a crustal thickness of 40–50 km, extends from the western coast of the Black Sea up to the western Baltic Sea. Here this system splits into two branches, the northern Sorgenfrei–Tornquist Zone (STZ) and the southern Trans-European Fault Zone (Figure 3). This zone is well recognizable in the contour map of the Moho-discontinuity in Figure 5.

The Variscides of Central and Western Europe

The main part of Central and Western Europe has a Late Proterozoic (Cadomian) fundament, but the present-day structure is controlled by a strong Variscan overprint (*see* **Europe:** Variscan Orogeny). This region extends between the North Sea and Baltic Sea to the north and the Alpine mountain belts to the south and, in the west to east direction, between the Atlantic Ocean and the Trans-European Suture Zone (Figure 3). The upper section of Figure 7 shows the crustal structure of the transition from the Baltic Shield to the Variscan Domain in Central Europe (*see* **Europe:** Variscan Orogeny). The lower section of shows the crustal structure of the area between the

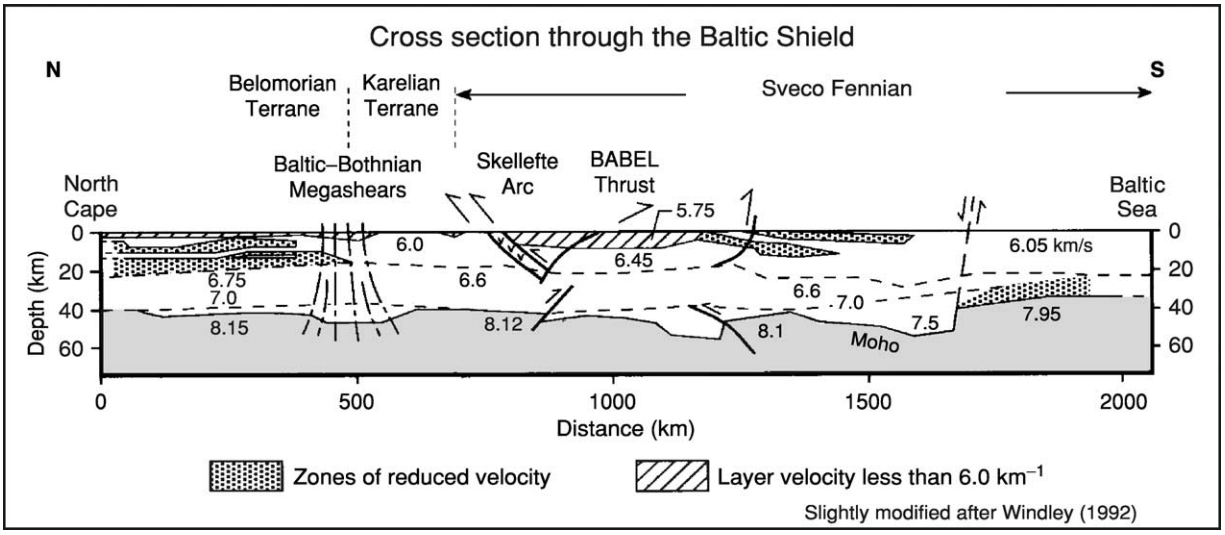


Figure 6 N–S cross section through the Baltic Shield with the main crustal structures, displayed as an isoline velocity model (vertical exaggeration: 5:1). Modified after Windley *et al.* (1991) in Blundell *et al.* (1992).

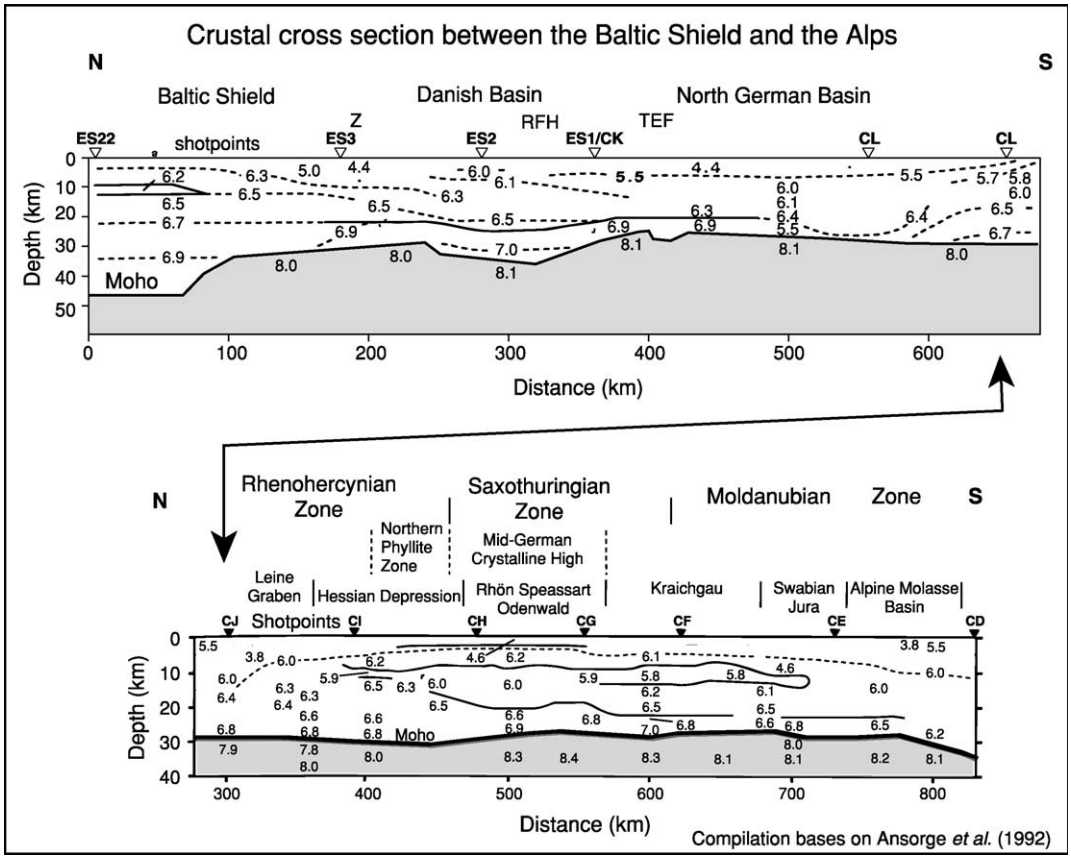


Figure 7 (Upper part) Structure of the transitional zone between the Baltic Shield and the Variscan crust adjacent to the south, which has undergone Permo-Mesozoic extension. STZ, Sorgenfrei–Tornquist zone; RFH, Ringkbing-Fyn High; TEF, Trans-European Fault. (Lower part) North–south section from the northern margin of the Rhenohercynian zone to the northern margin of the Alps. Although the internal crustal structure is very heterogeneous and the section cuts prominent suture zones, the crustal thickness is rather constant (28–30 km). Modified from Ansorge *et al.* (1992).

North German Basin and the northern margin of the Alps. Along this profile, the recent Moho-discontinuity lies at a depth between 28 and 30 km The depth increases to 34 km at the northern margin of the Alps. The present-day morphology of the Moho-discontinuity does not show any significant relicts of Variscan roots. Velocities in the uppermost mantle change significantly from a minimum value of 7.8 km s^{-1} in the north to as much as 8.4 km s^{-1} under the southern margin of the Mid-German Crystalline High; from there it decreases again to a more normal value of 8.1 km s^{-1}. The Variscan orogenic systems were formed during Late Palaeozoic times by the collision of numerous continental fragments or terranes of different size with the Baltic Shield and the East European Platform. Because seismic reflection profiles provide evidence of important crustal stacking, compressional deformation must have formed thick and deep crustal roots, comparable with the mountain roots of the recent Mediterranean orogens. These roots have not survived. Destruction of the Variscan roots is usually referred to as 'collapse' of the orogen. This comprehensive term summarizes a variety of processes, including erosion and isostatic uplift; gravitational spreading of hot, low-viscosity materials; extension by active rifting; and delamination of dense lower crust and mantle lithosphere.

Spectacular Crustal Root of the Uralides

The Ural Mountains are the product of collision of the East European Craton with the Siberian Shield (*see* **Europe:** The Urals). Figure 8 shows the results of a seismic profile across the southern Uralides. The Moho-discontinuity, well defined by seismic reflection data, dips from the west and east under the central part of the belt, down to a depth of 50 km. Wide-angle reflections in the central part of this belt reveal a still existing crustal root, with the Moho-discontinuity situated at a depth of 50–60 km. This unusual preservation of an orogenic root may be explained by the small density contrast (only 100 kg m^{-3}) between the root and the surrounding

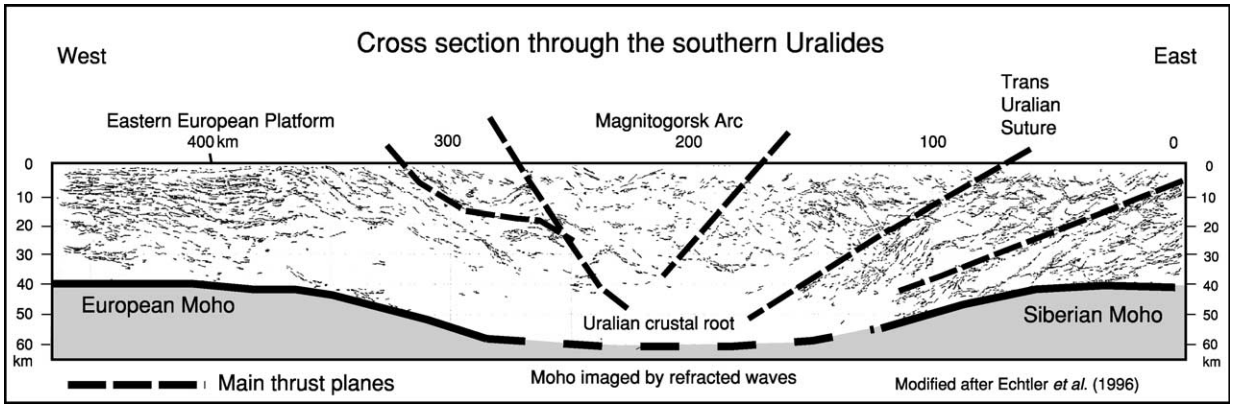


Figure 8 (Upper part) West- to east-directed section across the southern Uralides, based on seismic reflection and refraction measurements. Note the bilateral symmetry of reflections dipping from either side under the thick root. Modified from Echtler *et al.* (1996).

mantle, which was not sufficient to drive buoyant lateral spreading. In addition, there is no indication of post-orogenic active rifting in the Uralides.

The Post-Variscan Basins of Central and Western Europe

The post-Variscan regions of Central and Western Europe are characterized by an average crustal thickness of only 28–32 km (Figures 5 and 7). In these regions, the post-Variscan evolution is dominated by extensional tectonics, with formation of a new Moho-discontinuity at depths around 30 km. The main extensional elements are the Permo-Carboniferous basins in northern Germany and Poland and the Cenozoic European Rift System (*see* **Europe:** Permian to Recent Evolution). In the North German and Polish basins, the thickness of the post-Variscan sediments may attain 10 km, and the base of the crust has been detected at depths between 23 and 30 km. This implies crustal thinning, decreasing from a thickness of ~30 km to 13–23 km. Basin development started in Early Permian times and was probably caused by a combination of strike–slip and extension along the south-western margin of the East European Platform, and thermal subsidence subsequent to an Early Permian episode of mantle-derived magmatism. A second magmatic event occurred during the Cretaceous.

The Cenozoic European Rift System traverses Western and Central Europe up to the southern border of the Baltic Shield. Over a distance of about 1800 km, this north–south-oriented rift system is composed of a sequence of separated segments (Figure 5). This rift system starts in southern France in the Rhône Graben. Northwards, the Limagne Graben, west of the Rhône Valley, is a very significant segment of the rift that traverses the Massif Central and is flanked by volcanic chains. Further north, the Bresse Graben system continues from the Rhône Valley. Both the Limagne Graben and the Bresse Graben systems terminate again in the Burgund transform fault system, which links the grabens of south-eastern France with the Rhine Graben near Basel. The Rhine Graben (Figure 5) between Basel and Frankfurt/Main forms the central segment of the European Cenozoic Rift System. At the northern end of the Rhine Graben, at a triple junction near Frankfurt, the rift system splits into two branches. To the north, it follows the Hessen depression and the Leine Graben near Göttingen and disappears beneath the North German Basin. The Oslo Graben forms the northernmost segment of this rift system. The western branch starts at the northern end of the Rhine Graben near Frankfurt and continues in a north-west direction through the Rhenish Massif into the Lower Rhine Basin and disappears beneath the North Sea.

As shown in Figure 5, the total crustal thickness along the Central European Rift System is distinctly reduced, from 30 to 32 km outside the rift to about 28 km on top of the rift. Even the splitting of the rift near Frankfurt is recognizable in the contour map of the Moho-discontinuity. (Figure 5). The creation of the layering in the lower crust has been widely discussed in the literature. Deep seismic reflection profiles suggest that the layering of the lower crust resulted in an extensional regime, and that magmatic intrusions caused partial crustal melting and the transformation of deep crustal mafic rocks into eclogite. The unusual large amplitudes of the Moho reflections and the presence of anomalously high seismic velocities in the lowermost crust beneath Tertiary rifts (e.g., the Rhine and Rhône Grabens) imply that the Moho-discontinuity and part of the layering

of the lower crust are comparatively young features, related to interactions between the lower crust and the mantle.

Crustal Structure of the Mediterranean Region

Towards the south, Europe is flanked by Mesozoic–Cenozoic orogenic systems created by the convergence and collision of the European continent with the African/Arabian Plate. During their evolution, since the early Mesozoic, the Alpine–Mediterranean orogenic belts underwent a more or less complete cycle of extension, subsidence, subduction, collision, topographic buildup, extensional collapse, and back-arc spreading associated with the creation of new oceanic basins. The morphology of the Moho-discontinuity shows that this discontinuity was strongly involved in these compressional and extensional deformation processes.

All Mediterranean mountain belts show the typical subdivision into three main tectonic zones, each with its own specific crustal structures: the foreland (fore-arc), the central zone, and the hinterland (back-arc). In all the studied cases (Alps, Apennines, Betic Cordillera, Carpathians, Caucasus, Hellenides, and Pyrenees), this basic picture is clearly expressed in the asymmetric geometry of the Moho-discontinuity. The Moho-discontinuity of the foreland dips under the central zone and in some cases even beyond, under the hinterland, down to a depth of 60–70 km. Still deeper crustal rocks cannot be excluded, but have been metamorphosed into eclogites, which owe their thickness to tectonic stacking of sedimentary and crystalline units, including mantle slices, as well as to magmatic additions to the crust. The hinterland, with its thin crust, overlies the underthrusted continental lithosphere of the foreland and oceanic crustal fragments of the central zone. In this case, there is no continuity between the Moho boundary of the foreland and the Moho boundary of the hinterland. The vertical offset between the two Moho-discontinuities may attain 10–30 km. In the closely spaced array of Mediterranean microplates (*see* **Europe:** Mediterranean Tectonics), the tectonic role of a crustal block may change across a short distance: the Adriatic Microplate (*see* **Europe:** The Alps) is the upper plate in the Alpine system, whereas it represents the lower plate in the Apennines. Three typical crustal cross-sections (Pyrenees, Western Alps, and Northern Apennines) highlight the main characteristics of the Moho topography in the Mediterranean belts (Figures 3, 9, 10, and 11).

Pyrenees

The Pyrenees are situated between the Iberian continental block and the European Plate. Extensional deformation was active from the Triassic through the Cretaceous. This extensional phase was followed by convergence between Iberia and Europe that occurred from the Late Cretaceous through the Middle Miocene, creating the Pyrenees. The present-day crustal structure across the Pyrenees displays a well-preserved crustal root created by the collisional process (Figure 9). The seismic reflection profile reveals how the Iberian crust was subducted below the European crust. A Moho offset of about 30 km has been detected beneath the North Pyrenean fault zone, and the reflecting lower crust of the Iberian crust can be traced down to a depth of 60 km.

Central Alps and Northern Apennines

The Alps are considered as a classical example of an orogen formed by the convergence between the continents of Europe and Africa, including the intervening Adriatic Microplate (*see* **Stratigraphical**

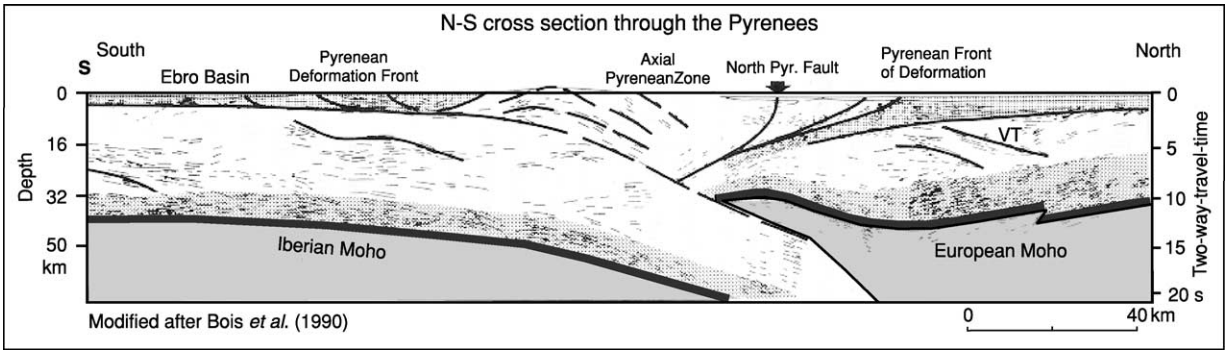


Figure 9 Line drawing and tectonic interpretation of a deep seismic profile across the Pyrenees. Beneath the central zone the offset of the Moho-discontinuity is clearly recognizable. Modified from Bois and ECORS (1994).

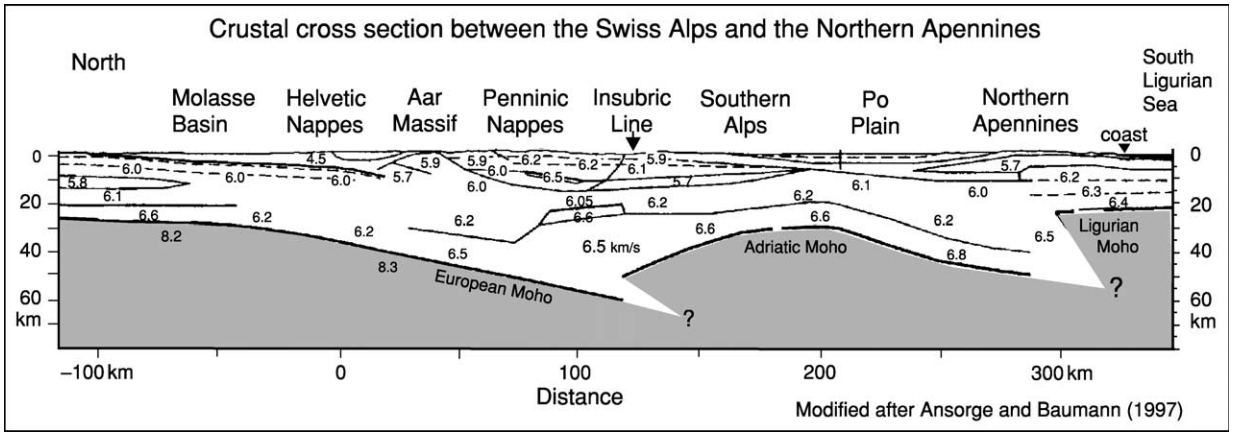


Figure 10 North- to south-directed crustal section between the Central (Swiss) Alps and the Ligurian Sea. This section shows two clearly expressed offsets of the Moho-discontinuity. Thus three separated Moho-discontinuities can be defined from north to south (the European, the Adriatic, and the Ligurian sea Moho). Modified with permission from Kissling (1997) in Pfiffner *et al.* (1997).

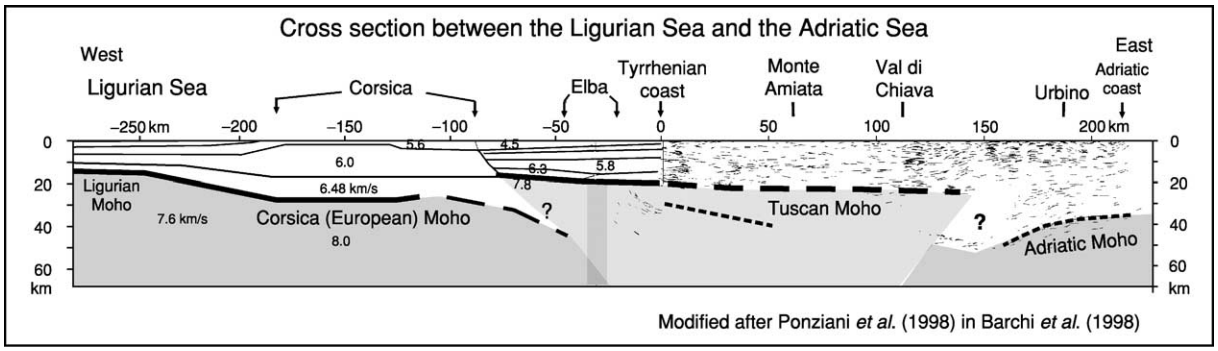


Figure 11 West- to east-directed crustal section across the Northern Apennines between the Ligurian Sea and the Adriatic Sea. This section is based on seismic refraction data in the western part and on seismic reflection observations in the eastern part. Note the different Moho-discontinuities from west to east: the thin oceanic crust of the Ligurian Sea, the European crust of Corsica (and Sardinia) with a normal Variscan crustal thickness of 30 km, the thinned crust in the Tuscany region, and the subducting crust of the Adriatic foreland. Reproduced from Ponziani *et al.* (1998).

Principles). After the Variscan Orogeny, a vast embayment in the eastern margin of Pangaea propagated westwards during the Mesozoic to produce the Tethys Ocean. In the Alps, convergence started in the mid-Cretaceous with the consumption of ocean crust and was followed by overthrusting of the marginal units of the Adriatic Microplate onto the European continental margin. In the Southern Alps (the Alpine hinterland), thrusting of the sedimentary cover was directed southwards towards the Po Plain. South of the Po Plain, along the southern and western margins of the Adriatic Microplate, the mountain belt of the Northern Apennines was formed in the Cenozoic. These repeated collisions have strongly affected the geometry and depth of the Moho-discontinuity. Figure 10 shows a north–south crustal cross-section between the Swiss Alps in the north and the northern Ligurian Sea in the south. This section traverses the Central and Southern Alps, the western Po Plain, and the Northern Apennines. The complex shape and geometry of the Moho-discontinuity are clearly imaged in the records of the seismic reflected and refracted waves. The seismic section shows the well-developed structural asymmetry of both mountain belts. From the Alpine foreland (the lower plate), the Moho-discontinuity dips southwards, from 30 km at the northern Alpine front to about 60 km beneath the Central Alps. Under the Southern Alps, the Adriatic Moho-discontinuity situated at a depth of about 40 km is clearly separated from the European Moho-discontinuity overlap of both Moho-discontinuities. The southern part of the section (Figure 10) traverses the Southern Alps and the Po Plain, the Northern Apennines, and the northern part of the Ligurian Sea. The continental crust of the Southern Alps and of the Po Plain must be related to the margin of the Adriatic Microplate. Beneath the Po Plain, the Adriatic Moho-discontinuity shows a minimum

depth of 35 km. Towards the north and south, the Adriatic Moho-discontinuity dips down to 40–50 km beneath the Southern Alps (as the upper plate) and also beneath the Northern Apennines (as the lower plate). Beneath the Northern Apennines, there is a clear offset of 20–30 km between the deep northern and the shallow southern Moho-discontinuity. The thinned crust of the Northern Apennines (about 20–25 km) must be regarded as an extended marginal part of the Adriatic Microplate. Southwards, the thin crust of the Northern Apennines passes into the thin oceanic crust of the Ligurian Sea.

Ligurian Sea, Tuscany, and Adriatic Sea

Figure 11 shows a WSW–ENE-trending cross-section between the Ligurian Sea in the west and the Adriatic Sea in the east, which passes by northern Corsica, Elba, Tuscany, and the Umbrian Apennines. This profile demonstrates the complex structure of this divergent orogenic belt, which resulted from repeated deformation with changing transport directions along the margin of the Adriatic Microplate. The counterclockwise rotation of the Corsica–Sardinia block has strongly affected the geometry and nature of the Moho-discontinuity west and east of Corsica. Along a south-west and north-east section, four different Moho-discontinuities can be identified (Figure 11): (1) a young oceanic Moho-discontinuity beneath the Ligurian Sea and Provençal Basin is situated at a depth of 14–18 km; (2) beneath the Variscan continental crust of Corsica, the European Moho-discontinuity has been detected at a depth of 25–30 km; (3) the young Moho-discontinuity beneath Elba and Tuscany is situated at a depth of only 20–25 km (the structural relations between the Corsican and the Tuscany Moho are a matter of debate: there is probably no direct link between the discontinuities); and (4) a well-expressed offset of 20–30 km separates the shallow Tuscan Moho-discontinuity from the deep Moho-discontinuity beneath the Northern Apenninic Arc (this offset defines the crustal root of the Apennines).

Along the section (Figure 11), there are three different zones in which the crustal thickness changes. In the westernmost part, the transition from the thin and young oceanic crust of the Ligurian Sea to the Variscan continental crust of Corsica can be interpreted as a passive continental margin. The situation is completely different in the contact zone between Corsica and Tuscany. Due to the counterclockwise rotation of Corsica, an eastward-dipping subduction zone must be postulated beneath Elba and western Tuscany. In the reflection profile CROP03 beneath Elba and western Tuscany, weak, eastward-dipping reflections indicate the existence of an interface that may be interpreted as the top of an eastward-dipping subduction zone. Very probably there is no direct link between the Moho-discontinuity beneath Corsica and the Moho-discontinuity beneath Tuscany. The Tuscany crust in the hinterland of the Northern Apennines is characterized by back-arc extension that has reduced the crust to a thickness of only 20 km. The third change in crustal thickness occurs beneath the Apenninic Arc. In this area, the Tuscany Moho-discontinuity terminates and another deeper Moho has been detected at a depth of 40–50 km. There is no link between the shallow Tuscany Moho and the deep Moho-discontinuity of the Adriatic Microplate. The offset between the two Moho-discontinuities amounts to about 20 km. This break can be interpreted as a continent–continent collision zone. A normal crustal thickness of 30–35 km has been measured beneath the Adriatic coast, the foreland of the Apenninic Arc.

The Subcrustal Lithosphere in Europe

The lithosphere is defined as the rigid outermost shell of Earth, comprising continental and oceanic crust and the uppermost mantle. The Moho-discontinuity is an intralithospheric interface. The lithosphere is a strong, elastic layer overlying the weak, ductile asthenosphere with a more or less gradual transition. The ductile asthenosphere is believed to represent the zone in the mantle where the melting point is most closely approached by the actual temperature curve. Because it transmits shear waves, this layer cannot be completely molten, but small amounts of melt may be present. The thickness of the lithosphere, like that of the crust, exhibits important variations. The thickness of the European lithosphere is depicted by the contour map in Figure 12. The cross-section reveals four main features. First, beneath the Baltic Shield and the Eastern European Platform, the thickness of the lithosphere is about 200 km. This feature correlates with the thick continental crust and low heat-flow values (40–50 $mW\,m^{-2}$). Second, a thin lithosphere of about only 50 km exists beneath the western region of Central Europe, especially along the Rhine and Rhône rivers. Here, also, the crust is relatively thin ($\leq$30 km) and the heat flow is distinctly higher than in Northern and Eastern Europe (80–90 $mW\,m^{-2}$). Third, in the Mediterranean region, the thickness of the lithosphere is very heterogeneous. A thick lithospheric root has been detected beneath the Alps and the Northern Apennines. This feature is at least partly explained by sinking delaminated lithospheric slabs, subsequent to continental collision. Under the young oceanic basins such as the

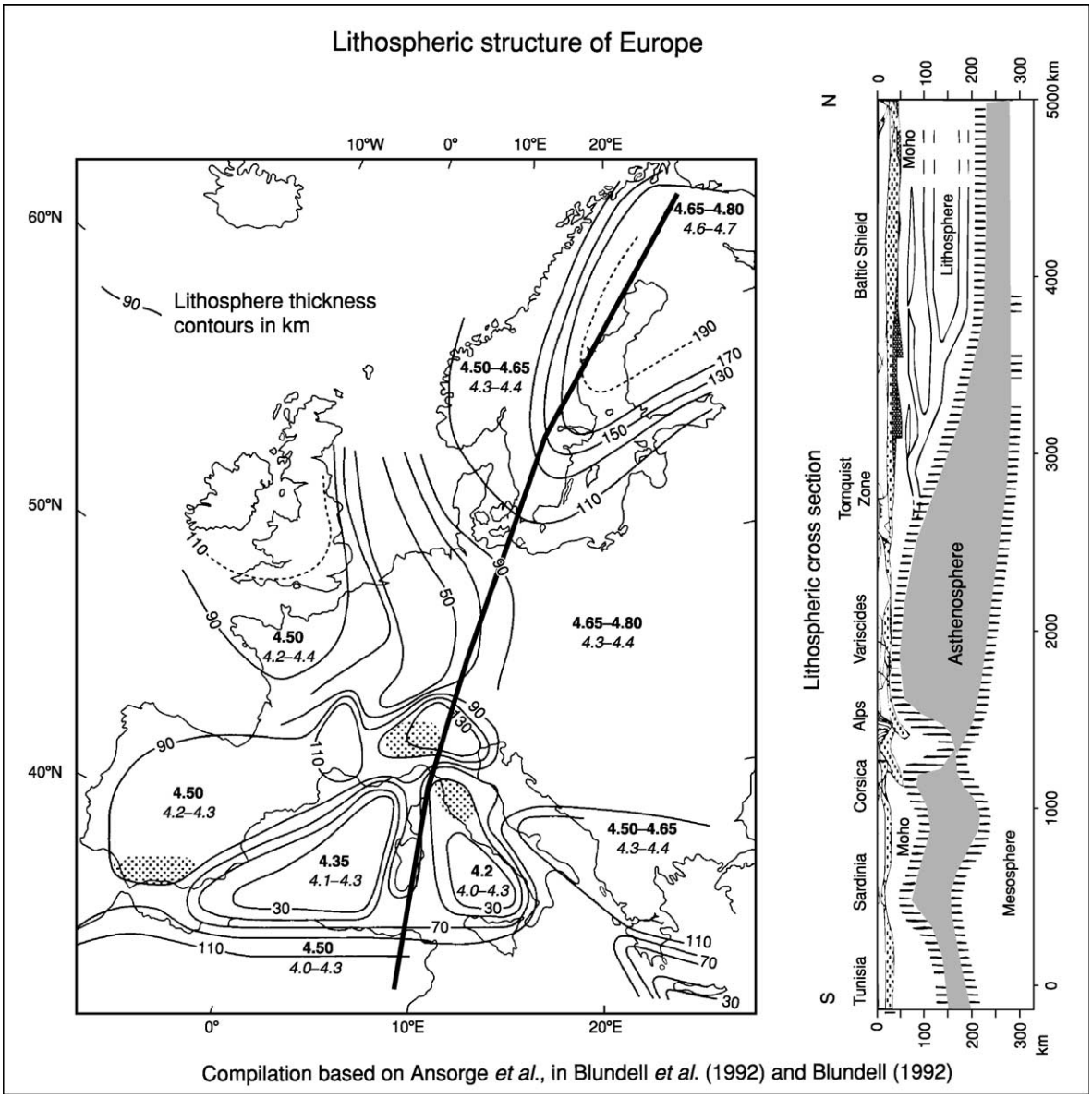


Figure 12 (Left) Contour map of the depth to the base of the lithosphere of Europe, derived from an interpretation of surface wave data. Numbers refer to S-wave velocities above and below the base of the lithosphere. Shaded areas indicate thickened lithosphere. Reproduced with permission from Panza (1992). (Right) North to south cross-section of Europe, demonstrating very clearly the basic differences of lithospheric thickness between the Baltic Shield in the north and the Variscan and Alpine part in Central and Southern Europe. Note the lithospheric root beneath the Alps. Modified from Blundell *et al.* (1992).

Tyrrhenian Sea, the asthenosphere is thin (10–20 km). Due to the very different tectonic situation, the heat flow varies strongly between 50 and 150 mW m^{-2}. Fourth, there is relatively thin lithosphere (about 50–80 km) in the Baltic Shield, but a thick one (about 200 km) exists beneath Central Europe.

Although recent tectonic and magmatic events in Europe are mainly concentrated in the Mediterranean region, there has been some recent activity in Central and Western Europe. Late Cretaceous to Holocene volcanism is found in an east- to west-trending belt that extends from western Poland to western Germany, continuing southwards across the French Massif Central to the Mediterranean. Recent seismological studies in the Eifel region and in the northern Massif Central have revealed low-velocity anomalies in the lithosphere and down to at least depths of 300–400 km. These findings support the concept that deep-seated mantle plumes can produce intraplate mantle volcanism.

Analysis

Studies of the European Moho-discontinuity offer the possibility to compare crustal structures of very different ages and evolutionary stages. The Moho-discontinuity is not a static feature in the lithosphere, but rather evolves geodynamically. In Europe, the Moho-discontinuity in each of the four main regions is distinguished by a specific nature and geometry. The Baltic Shield, including the East European Platform, is characterized by a 40- to 50-km-thick continental crust. In the shield regions, the average crustal velocity and density are relatively high due to a thick lower crust with mafic components. These shield regions have remained rather stable and undeformed in the Phanerozoic Eon, during which the crust of Central, Western, and Southern Europe has been repeatedly reworked. A relatively thin crust, only 30 km thick, characterizes the Variscan belts in Central and Western Europe. The Variscan Orogen was formed in Devonian and Carboniferous times by the sequential collision of continental fragments, producing a thickened crust such as that in the recently formed Alps. The Variscides subsequently collapsed and then re-equilibrated in Permian times with extensional overprinting. This re-equilibration was associated with widespread crustal melting, extensive volcanism, and the formation of late- to post-orogenic sedimentary basins. Consequently, the post-Variscan Moho image is mainly related to post-orogenic pervasive subhorizontal shearing associated with magmatic underplating. This layered structure of the lower crust is imaged by the highly reflective crust under sedimentary basins. An exception to this general evolution exists in the Palaeozoic orogen of the Urals, which have preserved their collisional structure with a crustal root. Apparently, this root was never destabilized by major fluid flow and heating.

The third main European region was formed by the young orogens in the Mediterranean regions of Southern Europe, which are characterized by a very heterogeneous crustal and Moho-discontinuity structure, morphology, and composition. Finally, spreading along the Mid-Atlantic Ridge since the Early Jurassic produced oceanic crust and an oceanic Moho-discontinuity.

The differences between the Baltic Shield and west-Central and Western Europe do not require different plate tectonic mechanisms, but can be explained by higher temperatures and rheological softening of the Variscan crust. This may ultimately be due to the repeated emplacement of melts rich in heat-producing radioactive elements. The Variscides have passed the three stages typical of an orogenic cycle: compression and crustal thickening, metamorphism and extensional collapse of the orogen. By comparison with the Variscides, the evolution of most of the young Mediterranean mountain belts has advanced only to the second stage or to an early part of the third stage. The mountain roots of these young orogenic belts are still preserved and are clearly recognizable in the seismic records. In all orogenic belts, it cannot be excluded that lower crustal rocks are present in eclogite facies and therefore escape detection by geophysical methods.

See Also

Earth: Crust. **Europe:** East European Craton; Caledonides of Britain and Ireland; Scandinavian Caledonides (with Greenland); Variscan Orogeny; The Urals; Permian to Recent Evolution; The Alps; Mediterranean Tectonics. **Minerals:** Olivines. **North America:** Continental Interior. **Plate Tectonics**. **Russia**. **Seismic Surveys**. **Shields**. **Tectonics:** Mountain Building and Orogeny. **Ultra High Pressure Metamorphism**.

Further Reading

Blundell D, Freeman R, and Mueller S (eds.) (1992) *A Continent Revealed: The European Geotraverse,* pp. 1–275. Cambridge: Cambridge University Press.

Bois C, Cazes J, Choukroune P, Gariel O, Hirn A, La Gall B, Lefort J P, Matte P, and Pinet B (1994) Seismic reflection image of the Pre-Meseozoic crust in France and adjacent areas. In: Keppi JD (ed.) *Pre-Mesozoic Geology in France and Related Areas,* pp. 3–47. Berlin Heidelberg: Springer-Verlag.

Echtler HP, Stiller M, Steinhoff F, Krawczyk CM, Suleimanov A, Spiridonov V, Knapp JH, Menshikov Y, Alvarez-Marron J, and Yunusov N (1996) Preserved collisional crustal architecture of the southern Urals revealed by Vibroseis CMP profiling. *Science* 274: 224–226.

Gee DG and Zeyen HJ (eds.) (1996) *Europrobe Lithosphere Dynamics: Origin and Evolution of Continents,* pp. 1–138. Uppsala, Sweden: EUROPROBE Secretariat.

Giese P and Pavlenkova N (1988) *The Structural Maps of the Earth's Crust of Europe* (In Russian). Moscow: Nauka.

Kaban MK (2001) Gravity model of the North Eurasia crust and upper mantle: 1. Mantle and static residual gravity anomalies *Russian Journal of Earth Sciences* 01.3. No.2.

Meissner R, Wever T, and Flüh ER (1987) The Moho in Europe: Implications for crustal development. *Ann. Geophys.* 3: 357–364.

Morelli C (1998) Lithospheric structure and geodynamics of the Italian Peninsula derived from geophysical data: a review. *Memorie della Societä Geologica Italiana* 52: 113–122.

Pfiffner OA, Lehner P, Heitzmann P, Mueller S, and Steck A (eds.) (1997) *Deep Structure of the Alps, results from NRP 20,* pp. 1–380. Basel Switzerland: Birkhäuser.

Ponziani F, de Franco R, and Biella G (1998) Geophysical reinterpretation of 1974 and 1978 DSS experiments along CROP03 profile. *Memorie della Società Geologica Italiana* Vol. LII: 215–224.

Tittgemeyer M, Wenzel F, Ryberg, T, and Fuchs K (1999) Scales of heterogeneities in the continental crust and upper mantle. *Pure Appl. Geophys.* 156: 29–52.

MOON

See **SOLAR SYSTEM: Moon**