Roger Narayan

Editor

Biomedical Materials

Biomedical Materials

Roger Narayan
Editor

Biomedical Materials

 Springer

Editor
Roger Narayan
Department of Biomedical Engineering
University of North Carolina, Chapel Hill
152 MacNider Hall
Chapel Hill, NC 27599-1175
USA
roger_narayan@unc.edu

Printed on acid-free paper

springer.com

A Historical Perspective on the Development of Biomedical Materials

There have been enormous strides in the development of novel biomedical materials over the past three decades. A biomedical material (also known as a biomaterial) is a polymer, metal, ceramic, or natural material that provides structure and/or function to an implantable medical device. In one generation, a large number of biodegradable polymers, bioactive ceramics, and wear-resistant metal alloys have made their way from research laboratories into widely-used medical devices. This heavy flurry of recent progress by materials scientists has partially overshadowed the efforts of surgeons, who previously led research efforts to develop biomedical materials. Until the 1960's, surgeons were at the forefront of efforts to find new materials for use in medical prostheses. Surgeons were driven by their clinical duties to improve the treatment of those suffering from congenital malformations, trauma, or disease. These surgeon-scientists attempted to alleviate patient suffering using "off-the-shelf" materials, which were developed for nonmedical applications. This brief historical perspective describes some initial efforts to develop novel biomedical materials.

Bronze or copper have been used for thousands of years to repair fractured bones. However, use of bronze or copper was limited due to copper accumulation in the eyes, liver, brain, and other body tissues. Two developments in the late nineteenth century accelerated the use of synthetic materials in the human body. The development of the X-ray revealed that conventional external treatments were insufficient and stimulated the development of internal fixation procedures. In addition, broad acceptance of Lister's antiseptic procedures allowed for internal treatment of medical conditions with minimal risk of infection. Lister himself used antiseptic procedure to successfully suture fractured patellae with silver wire in 1885. Themistocles Gluck described replacement of both the acetabulum (pelvis) and femur (hip bone) using carved ivory at the 10th International Medical Congress in 1890; however, bone resorption and concomitant infection eventually caused these prostheses to fail. Neither acceptable materials nor designs were available at the time to fabricate medical devices.

Surgeons made several advances in biomedical materials development in the early- to mid- twentieth century. At the turn of the century, many European surgeons were experimenting with celluloid, rubber, magnesium, zinc, and other materials. In 1924, A. A. Zierold described an animal study on the interaction between bone

and several metals, including aluminum, aluminum alloy, copper, low carbon steel, cobalt-chromium alloy, gold, iron, lead, magnesium, nickel, silver, and zinc. X-ray and histological sections of canine bone-material interface revealed that gold, silver, stellite, lead, and aluminum were encapsulated by tissue and were well-tolerated by the animals. Venable, Stuck, and Beach demonstrated the electrolysis of implanted biomedical materials in a 1937 study. They placed aluminum, brass, carbon, cobalt-chromium alloy, copper, galvanized iron, gold, lead, magnesium, nickel, silver, stainless steel, vanadium steel, and zinc in bones of experimental animals. Their bio-chemical, radiographic and clinical findings demonstrated that ion transfer between different metals in the body occurred in accordance with electromotive force. Their work also demonstrated that cobalt-chromium alloy was essentially nonelectrolytic; this material has remained a mainstay medical alloy to the present day.

The orthopedic surgeon Sir John Charnley, made several significant advances in the field of biomedical materials, including (a) the introduction of the metal-ultra high molecular weight polyethylene bearing couple; (b) the use of poly (methyl methacrylate) for fixation; (c) the reduction of postoperative sepsis due to the use of laminar flow air-handling systems and prophylactic antibiotics; and (d) the place-ment of antibiotics in bone cement. In the 1950's, Charnley discovered that natural joints exhibit boundary lubrication. He then attempted to find a synthetic material with similarly low frictional properties. His first choice for an acetabular cup ma-terial, polytetrafluoroethylene, demonstrated poor wear properties. Unfortunately, this issue only became evident after a large clinical trial involving polytetrafluo-roethylene implants had begun. More than three hundred polytetrafluoroethylene implant surgeries had to be revised due to poor wear properties, necrosis, and im-plant loosening. Charnley's laboratory assistant subsequently examined ultra high molecular weight polyethylene, which had found use in mechanical looms. Charn-ley noted that: (a) polyethylene had better wear characteristics than polytetrafluo-roethylene; and (b) polyethylene was capable of being lubricated by synovial fluid. The use of novel materials and surgical procedures revolutionized the practice of joint replacement surgery and raised the success rate of this procedure to an ex-ceptionally high level (>90%). Charnley's total hip replacement is considered the gold standard for joint replacement; few changes have been made to this pros-thesis design in the past forty years. However, the Charnley prosthesis does have several disadvantages. One is an unacceptable rate of wear (about 200 μm/year). Metallic and, more commonly, polymeric wear particles cause a severe foreign-body reaction in the tissues that surround the prosthesis, which can lead to implant loosening.

The modern field of biomedical materials science owes a great deal to these pi-oneering individuals, who utilized existing knowledge at the interface of materials science and biology in order to improve the quality of life for others. In a similar manner, current-day biomedical materials researchers who work on porous coatings, bulk metallic glasses, artificial tissues, and nanostructured biomedical materials are utilizing modern concepts at the interface of materials science and biology in order to further knowledge in this rapidly developing area.

The goal of this book is to address several core topics in biomedical materials, including the fundamental properties of the materials used in medicine and dentistry; the interaction between materials and living tissues; leading applications of polymers, metals, and ceramics in medicine; and novel developments in biomedical materials. Homework problems and other material for each chapter can be found at the website http://springer.com/978-0-387-84871-6. We hope that this work will spur productive discussions and interactions among the many groups involved in the development and use of biomedical materials, including biomedical materials researchers, biologists, medical device manufacturers, and medical professionals. Finally, we would like to thank Elaine Tham, Lauren Danahy, and the staff at Springer Science+Business for making this book possible.

Chapel Hill, North Carolina   Roger Narayan

Contents

**Part I   The Fundamental Properties of the Materials Used
in Medicine and Dentistry**

Contributors

Sarit B. Bhaduri Department of Mechanical, Industrial and Manufacturing Engineering and Department of Surgery, University of Toledo, Toledo, OH 43606, USA, sarit.bhaduri@eng.utoledo.edu

Sutapa Bhaduri Department of Mechanical, Industrial and Manufacturing Engineering and Department of Surgery, University of Toledo, Toledo, OH 43606, USA, sarit.bhaduri@eng.utoledo.edu

C.K. Chua School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, mckchua@ntu.edu.sg

Florent Cros Wireless Sensor Division, CardioMEMS, Inc., Atlanta, GA 30308, USA, fcros@cardiomems.com

Patrick Doherty Centre for Lifelong Learning, University of Liverpool, Liverpool L69 3GW, UK, p.j.doherty@liverpool.ac.uk

Abraham J. Domb Division of Identification and Forensic Sciences (DIFS), HQ Israel Police, Department of Medicinal Chemistry and Natural Products, School of Pharmacy-Faculty of Medicine, The Hebrew University, Jerusalem, Israel, avid@ekmd.huji.ac.il

John P. Fisher Fischell Department of Bioengineering, University of Maryland, College Park, MD 20742, USA, jpfisher@umd.edu

Lauren Flynn Department of Chemical Engineering, Queen's University, Kingston, ON, Canada K7L 3N6, lauren.flynn@chee.queensu.ca

David Hill Rocket Medical plc, Washington, Tyne & Wear, NE38 9BZ, England, Dave@rocketmedical.com

Chunming Jin Department of Biomedical Engineering, North Carolina State University, Raleigh, NC 27695, USA, cjin@ncsu.edu

C. James Kirkpatrick Institute of Pathology, Johannes Gutenberg University, Langenbeckstrasse 1, D-55101 Mainz, Germany, kirkpatrick@repair-lab.org; kirkpatrick@ukmainz.de

Damien Lacroix Institute of Bioengineering of Catalonia, C. Baldiri Reixas, 13, 08028 Barcelona, Spain, dlacroix@ibec.pcb.ub.es

K.F. Leong School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, MKFLEONG@ntu.edu.sg

Miroslav Marek School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0245, USA, mirmarek@earthlink.net

Janice L. McKenzie Nanovis, Inc., West Lafayette, IN 47906-1075, USA, Thomas_Webster@brown.edu

Showan N. Nazhat Department of Mining and Materials Engineering, McGill University, Montreal, Quebec, Canada H3A 2B2, showan.nazhat@mcgill.ca

Kirsten Peters Department of Cell Biology/Junior Research Group, Biomedical Research Center (BMFZ), Schillingallee 69, 18057 Rostock, Germany, kirsten.peters@med.uni-rostock.de

Robert M. Pilliar Faculty of Dentistry, University of Toronto, Toronto, Ontario, Canada M5G 1G6, bob.pilliar@utoronto.ca

Josep A. Planell Department of Materials Science and Metallurgy, Technological University of Catalonia, 08028 Barcelona, Spain, josep.a.planell@upc.es

Jonathan Pratten Division of Microbial Diseases, Eastman Dental Institute, University College London, London, UK, j.pratten@eastman.ucl.ac.uk

Afsaneh Rabiei Mechanical and Aerospace Engineering, Biomedical Engineering, North Carolina State University, Raleigh, NC 27695-7910, arabiei@eos.ncsu.edu

Raju V. Ramanujan School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, ramanujan@ntu.edu.sg

Teerapol Srichana Department of Pharmaceutical Technology, Faculty of Pharmaceutical Sciences, Prince of Songkla University, Hat Yai 90112, Thailand, srteerap@ratree.psu.ac.th

K.H. Tan School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, kwanghui_tan@pmail.ntu.edu.sg

Irene G. Turner Department of Engineering and Applied Science, University of Bath, Bath BA2 7AY UK, I.G.Turner@bath.ac.uk

Ronald E. Unger Johannes Gutenberg University, Institute of Pathology, REPAIR Lab, Langenbeckstr 1, D-55101 Mainz, Germany, runger@uni-mainz.de

Thomas J. Webster Division of Engineering, Brown University, Providence RI 02912 USA, Thomas_Webster@brown.edu

Wei Wei Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC 27695, USA, wwei@ncsu.edu

Rachel L. Williams Clinical Engineering and Ophthalmology, School of Clinical Sciences, University of Liverpool, Liverpool L69 3GA UK, rlw@liverpool.ac.uk

David Wong St Paul's Eye Unit, Royal Liverpool University Hospital, Liverpool L7 8XP, UK; The LKS Faculty of Medicine, Eye Institute, HBHA Centre, University of Hong Kong, Hong Kong, rlw@liverpool.ac.uk

Kimberly A. Woodhouse Department of Chemical Engineering and Applied Chemistry, Institute of Biomaterials and Biomedical Engineering, Sunnybrook and Women's College Health Sciences Centre, University of Toronto, Toronto, ON, Canada, kas@chem-eng.utoronto.ca

Diana M. Yoon Department of Chemical and Biomolecular Engineering, University of Maryland, College Park, MD 20742, USA, dmyoon@umd.edu

Anne M. Young Biomaterials and Tissue Engineering Division, UCL Eastman Dental Institute, University College London, 256 Grays Inn Road, London, WC1X 8LD, UK, A.Young@eastman.ucl.ac.uk

Chapter 1
Ceramics and Glasses

Irene G. Turner

1.1 Introduction

Ceramics are renowned for their excellent wear properties, good resistance to
degradation in corrosive environments, high modulus and hardness values and high
melting points. Equally well known are their poor thermal and electrical conductiv-
ities and their reputation for being notch sensitive with low values of fracture tough-
ness. Historically, as a consequence of low impact and tensile strengths combined
with inherent brittleness, their use was limited. However, more recently, advances
in manufacturing technology have meant that a group of what can be termed high
performance engineering ceramics have emerged that can be used for a wide range
of applications. Their high melting point has led to their use in engines and turbines
at elevated temperatures. As a consequence of their improved toughness, they have
been incorporated in the design of body armour. Their excellent wear resistance,
high compressive strength properties, pleasing aesthetic appearance and proven bio-
compatibility have led to the development of a specific range of what are referred to
bioceramics which are now used extensively in many different areas of medicine to
augment or replace parts of the body. Alumina and zirconia are used to manufacture
components of hip joint replacements; hydroxyapatite (HA) and glass ceramics are
used as coatings on prosthetic stems; calcium phosphate based materials are used as
porous scaffolds, spinal implants and bone grafts; composites of HA combined with
a polymer are used to manufacture replacements for the bones of the inner ear –
these are only a few examples of the wide range of applications of ceramics in med-
ical engineering.

This chapter is divided into two sections. The first defines ceramics, glasses and
glass ceramics in terms of their structure, processing and properties. The differ-
ent manufacturing routes involved in pressureless sintering, cold and hot isostatic
pressing, tape casting and liquid phase sintering are outlined. Production methods
for porous ceramics using burnt out polymer spheres (BurPS), foamed slips and
reticulated foams are described, as are the surface engineering techniques of ion
implantation and thermal spray coating. The effect of processing method and poros-
ity on the mechanical properties of ceramics are also considered. Following a brief
section on the structure and properties of bone, specific categories of bioceramics
are then introduced. These include dense forms of alumina and zirconia for weight

bearing, articulating joint replacement components, porous forms of calcium phosphate based ceramics for replacement and repair of natural bone, HA coatings for prosthetic stems and bioactive glasses for dental and surgical use.

1.2 What Is a Ceramic?

The term ceramic comes from the Greek word *keramos* which literally means 'burnt stuff' alluding to the fact that the raw materials must be fired by heating to an elevated temperature in order to form a material with a useful range of engineering properties. Ceramics are generally composed of two or more metallic and non-metallic elements resulting in more complex crystal structures than in, for example, metals. The nature of the atomic bonding in ceramics varies from primarily ionic (strong, primary bonding between oppositely charged ions) to primarily covalent (strong, primary bonds formed by sharing of electrons) with many ceramics having a combination of the two types of bond. The strength of the bonds imparts the characteristic properties of high hardness, and thermal and electrical resistance to the ceramics [1, 2].

Ceramics, which are predominantly ionic in nature, combine a metal and a non-metal with dissimilar electrical charges. The anions often form a regular three dimensional array in space providing interstitial sites into which the cations will have to fit to result in the correct coordination. The arrangement of the ions in the unit cell must satisfy ionic radius ratio requirements and ensure the proper charge balance is maintained. A good example of this is alumina, Al_2O_3, in which the oxygen ions (O^{2-}) form a close packed hexagonal (c.p.h.) arrangement which has one octahedral and two tetrahedral holes per atom as shown in Fig. 1.1a. The Al ions (Al^{3+}) occupy the octahedral sites such that each is surrounded by six O^{2-} ions. In order to balance the charges overall, the Al^{3+} ions must only fill two thirds of the sites, leaving one third of the sites vacant.
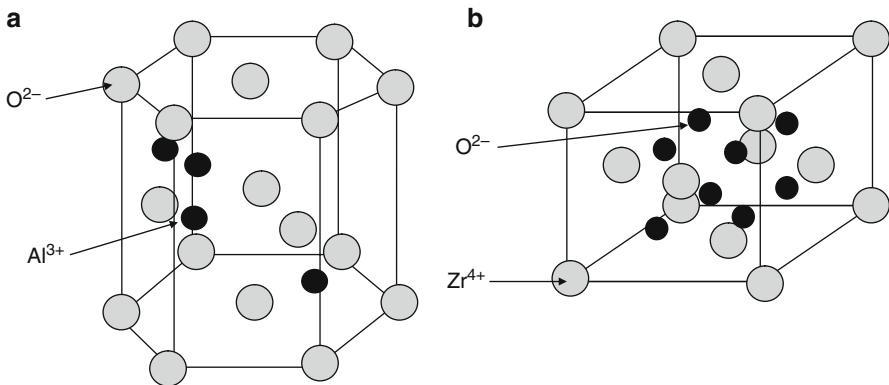
Fig. 1.1 (a) Hexagonal alumina, Al_2O_3 (b) Cubic Zirconia, ZrO_2

In contrast, zirconia, ZrO_2, can be described as a face centered cubic (f.c.c.) arrangement of zirconium with the O^{2-} ions occupying the tetrahedral sites, of which there are two for each atom in the f.c.c. structure (Fig. 1.1b). In both cases the electrostatic attraction between the unlike charges on the ions ensures that the ions pack densely together in a regular pattern, resulting in strong ionic bonding throughout the structure.

Ceramics which are predominantly covalent can be single elements, for example, diamond, C, or can be a combination of two non-metallic elements, for example, silica, SiO_2. Covalent bonding differs from ionic bonding in that adjacent atoms share electrons which results in a fixed number of strong directional bonds throughout the structure. In some cases this can result in the formation of extended chains, sheets or three dimensional networks. Ceramics can therefore exist in the form of single crystals, regular polycrystalline solids or irregularly arranged amorphous materials. They are a diverse group of materials whose biological response and mechanical properties such as strength and toughness will vary with phase composition and grain size.

1.3 Ceramic Processing

In order to process high performance engineering ceramics it is necessary to produce a powder. In the case of alumina, the starting point is bauxite, a naturally occurring form of hydrated aluminum oxide which can be found in various parts of the world. The bauxite is mined, crushed and purified, followed by heating to 1150 °C, the temperature at which it decomposes to form alumina:

$$2Al(OH)_3 = Al_2O_3 + 3H_2O.$$

The alumina is then milled and sieved to produce a high purity, fine starting powder ($<1\,\mu m$ particle size) which can then be processed in a number of different ways.

1.4 Powder Processing

The high melting point of ceramics means that, unlike most metals and polymers, they cannot easily be heated to their melting temperature to facilitate processing – alternative routes have had to be developed. The processing of powders to form solid products necessarily involves the packing together of fine powders, their consolidation and heating to form bonds. This process of diffusion and bonding is referred to as solid state sintering. Fine powders will have a large surface area and therefore a large surface energy associated with them. This energy can act as a driver for the sintering process.
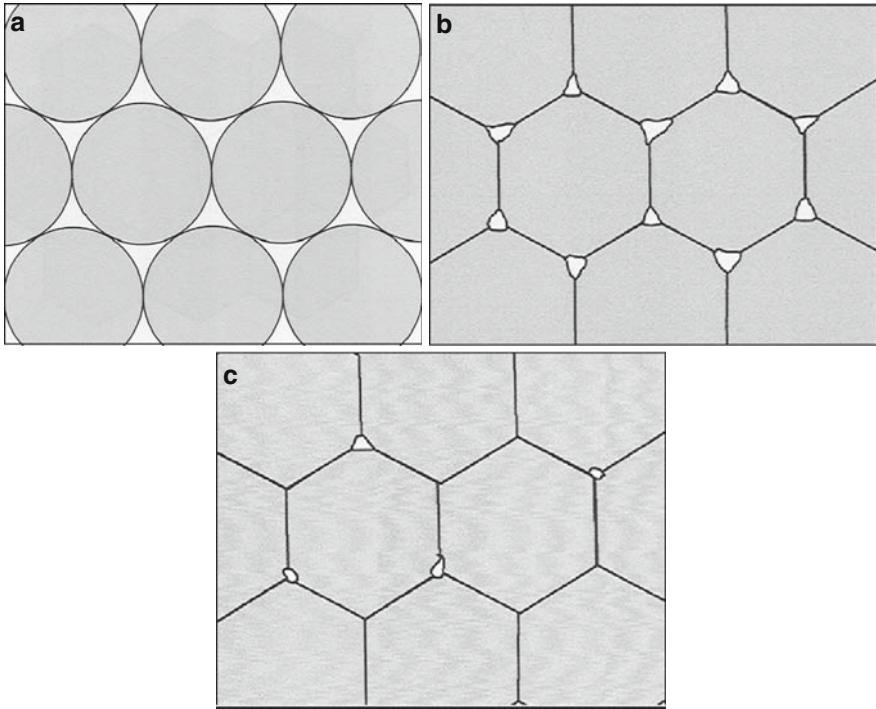
Fig. 1.2 (**a**) Powder particles compacted together (**b**) Particles beginning to bond together (**c**) Fully sintered ceramic

The process of sintering takes place in thee identifiable stages as shown in Fig. 1.2. In the first stage the powder is compacted; the particles are in contact with one another but are not physically bonded in any way (Fig. 1.2a). The compacted powder is heated to a temperature which is generally about 2/3 of T_m, the melting point. At this stage 'necks' begin to form between the particles, bonding them together (Fig. 1.2b). The small contact areas between the particles expand, and at the same time the density of the compact increases and the total void volume decreases. Small diameter particles will have a high surface area and a high surface free energy. The high surface free energy of the particles is the driving force for the sintering process. Therefore, there is a strong thermodynamic drive to decrease the surface area by bonding particles together. As the number of bonds grow, the surface area, and thus energy, is reduced. In the final stage (Fig. 1.2c) individual particles can no longer be seen – they are fully bonded together leaving residual porosity in the form of closed off pores which are of sufficiently small diameter so as not to have a detrimental effect on the mechanical properties of the final material.

The original powder particle size will control the final pore size and distribution: the smaller the particle size the smaller the pores and the better the mechanical properties will be. As the powder is sintered, the grains will grow such that the

final grain size will often exceed the initial powder particle size. Ideally, to optimize strength, the powder needs to be densified quickly to allow minimal grain growth. For dense ceramics the strength is a function of the grain size. Finer grain size materials will have smaller flaws at the grain boundaries and thus be stronger than ceramics with larger grain sizes. The relationship between grain size, D_g, and strength, σ_0, is defined by the Hall-Petch equation:

$$\sigma_0 = \sigma_s + BD_g^{-1/2}$$

where σ_s is the yield strength of a single crystal and B is a constant.

The sintering mechanism is controlled by diffusion at the grain boundaries. A combination of rapid grain boundary diffusion with slower lattice diffusion allows the atoms to diffuse towards the pores. Vacancies tend to flow away from the surface of the sharply curved neck; this is the equivalent of flow of material towards the neck which grows as the void shrinks. The flow is always from a source to a sink. The source can be the sharply curved neck; the sink can be a grain boundary, a dislocation or the surface of the particle. Vacancies can follow different paths resulting in different diffusion mechanisms. The path can be through the lattice, along the surface, along the grain boundaries or via dislocations resulting in volume, surface, grain boundary or pipe diffusion respectively. The flow of vacancies to any of the sinks is equivalent to the flow of material in the opposite direction. Although one mechanism will usually dominate, the rate of sintering will depend on the sum of all the available mechanisms. The rate of densification is given by

$$d\rho/dt = C/a^n \exp(-Q/RT)$$

where ρ is the density, a is the particle size, C and n are constants, Q is the activation energy for sintering, R is the gas constant and T the absolute temperature. Q is usually the activation energy for grain boundary diffusion and n typically has a value of about 3. As sintering progresses, the dominant mechanism will change with the increase in particle radius [3].

As the powder sinters and permanent bonds are formed between the particles, the mechanical properties change significantly as indicated in Fig. 1.3 – the stiffness, strength and density all increase. When processing powders, the shape and particle size distribution will affect the packing and thus the mechanical properties; for example, if a single particle size is used, the initial void space could be as high as 30%. Thus a range of particle sizes are desirable to maximize the packing density and uniformity whilst minimizing shrinkage and porosity.

1.5 Deformation and Fracture

Ceramics, when they fail, tend to do so in a dramatic manner. As has been discussed earlier, a characteristic of ceramics is their high strength ionic bonds which have to be broken in order for the material to fail. The brittle nature of ceramics
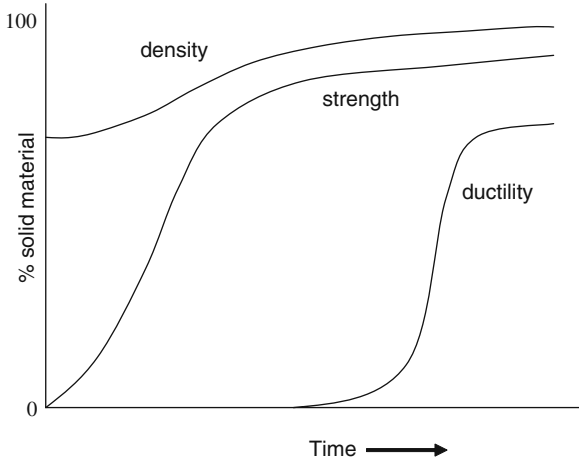
Fig. 1.3 Density, strength and ductility in relation to % solid material

is therefore attributed to this strong bonding in that it is not possible for plastic deformation to occur in a ceramic prior to failure as slip cannot occur. Consequently, if a crack is initiated, its progress will not be hindered by the deformation of material ahead of the crack as would be the case in a ductile material such as a metal. The crack will continue to propagate, rapidly resulting in catastrophic failure.

In a ceramic, therefore, the presence of a crack can concentrate the applied stress, reducing its ability to withstand tensile forces. This effect can be explained using the Griffith crack theory. The actual stress that will act on the crack tip when a tensile stress σ is applied is given by the equation

$$\sigma_{actual} \cong 2\sigma\sqrt{\frac{a}{r}}$$

where a is the length of the crack and r the radius at the crack tip.

From this equation it can be seen that a long crack with a high value for a, or a thin crack with a large value for r will result in a high concentration of stress at the crack tip. Unless there is something present in the structure to impede the growth and progression of the crack, as the length of the crack increases the stress concentration will also increase. As a consequence, ceramics tend to have low fracture toughness values.

The response to an applied stress is an elastic strain which can be related to the modulus of elasticity, E, of the material. When this strain energy is released, the total overall energy will be reduced as the crack propagates. However as the crack grows, two new surfaces are produced which will have energy associated with their

surfaces. The critical stress required to propagate a crack is given by the Griffith equation:

$$\sigma_{critical} = \sqrt{\frac{2E\gamma}{\pi a}}$$

where a, is the length of the crack and γ is the surface energy per unit area.

So, as before, it can be seen that even small flaws can have a profound influence on the strength properties of a ceramic. The effect will be significantly reduced if a compressive load is applied, in which case the crack will be closed up. Therefore ceramics generally have much better compressive properties than tensile properties. Also, if the crack length is less than the critical length, the crack will not propagate, however high the stress is at the tip [4]. Hence the relevance of critical flaw size when considering the structure and mechanical properties of ceramics.

1.6 Transformation Toughening

Transformation toughening is a mechanism that can be used to improve the toughness properties of some ceramics, zirconia being a good example. Zirconia can exist in three different phases – monoclinic, tetragonal and cubic – depending on the temperature and environment. The transformation temperatures are 900–1170 °C for monoclinic to tetragonal and 2370 °C for tetragonal to cubic. The transformation from tetragonal to monoclinic is associated with an increase in the volume of the lattice cell of 3–5%. Consequently, when cooling from sintering temperature, a pure zirconia block can crack due to the significant expansion in volume.

However, in the 1970s, it was realized that the phase transformations and their associated volume increase could be used to advantage to improve both the toughness and strength of zirconia by controlling the transformation process in the stress field ahead of the crack tip, as shown in Fig. 1.4. The volume expansion associated with the transformation from the tetragonal to monoclinic phase acts on the crack in such a way as to reduce its potential to propagate. The metastable tetragonal particles transform into the more stable monoclinic phase, at the same time helping to close the crack by shielding of the crack by the compressive stresses associated with the transformed particles as illustrated in Fig. 1.4b. The crack energy is effectively absorbed, thus arresting further crack growth and toughening the material. The toughness has been found to increase in proportion to the quantity of tetragonal particles retained in the structure [4–6].

Carefully controlled processing is the key to the success of the transformation toughening process. If zirconia is heat treated for too long the toughening effect is reduced as the tetragonal particles become too coarse. If the particles are too large the metastability of the tetragonal phase is reduced, the tetragonal particles become stable, and as a consequence the transformation and toughening effects disappear. By utilizing homogeneous starting powders, keeping the sintering temperature to
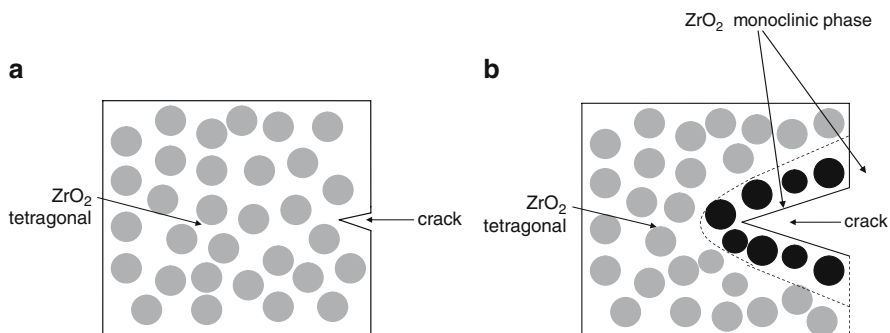
Fig. 1.4 (**a**) ZrO_2 tetragonal phase prior to transformation toughening (**b**) ZrO_2 transforming from tetragonal to monoclinic phase ahead of the crack tip

below 1400 °C and maintaining the grain size below 0.7 μm, tough, strong ceramics with excellent mechanical properties can be produced.

Orthopaedic grade zirconia has a complex chemistry with small quantities of stabilizing agents, 5.2 wt%Y_2O_3, 2 wt% HfO_2 and 0.5 wt% Al_2O_3 added to stabilize the phases. More than 95% of the zirconia is in the metastable tetragonal phase, the remainder being either cubic or monoclinic zirconia. The grain size is kept below 0.5 μm to ensure the tetragonal phase remains stable in physiological conditions [7, 8].

The operation of this mechanism is not confined to zirconia itself but can be utilized in other types of ceramic. Small particles of partially stabilized zirconia can be dispersed within the matrix of, for example, alumina where the same beneficial effect of arresting crack growth can be observed.

1.7 Pressureless Sintering

In the majority of cases the ceramic powder will be processed by simply pressing in a die to form an initial compacted shape as shown in Fig. 1.5. At this stage the starting powder may have been mixed with a binder to impart strength in the early stages of processing to enable it to be handled in what is referred to as the 'green state' before sintering. In addition, blending may be used to alloy the powders, lubricants may be added to reduce friction between particles, binders can be added and sintering aids may be introduced to help densification. Significant shrinkage will occur on firing as the particles coalesce but controlling the powder size distribution can maximize the packing density and minimize dimensional changes.

1.8 Isostatic Pressing

Cold isostatic pressing is a method applied to 'green' compacts to achieve a more uniform compaction prior to sintering. Powder is placed in a flexible mould made from neoprene, polyurethane or polyvinyl chloride. This is then placed in a
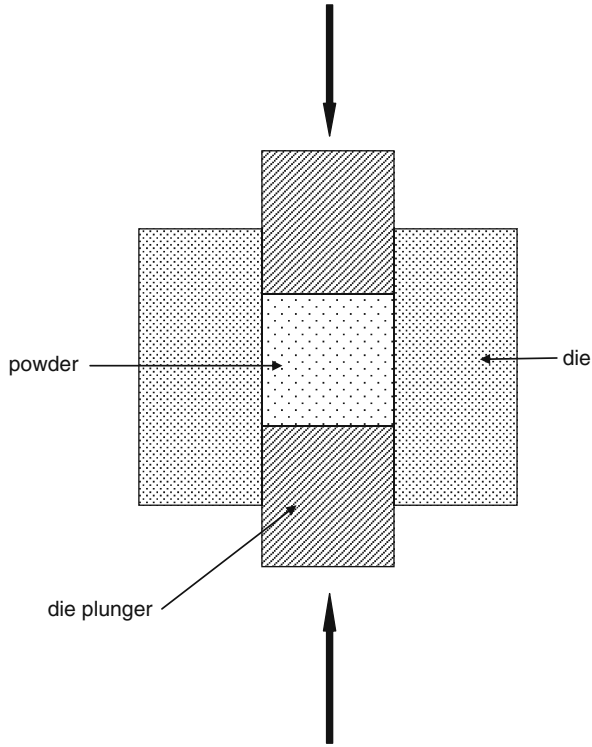
Fig. 1.5 Cold pressing of ceramic powder to form a compact
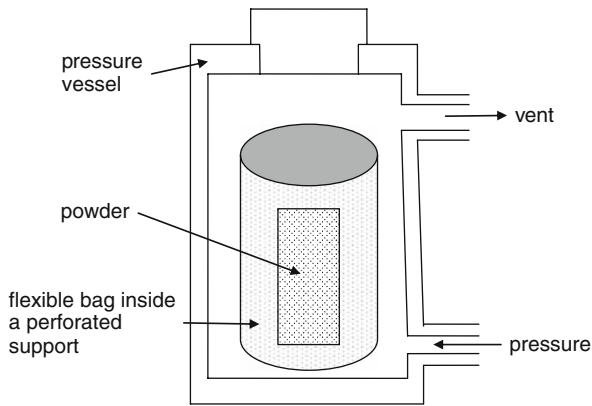
Fig. 1.6 Cold isostatic pressing
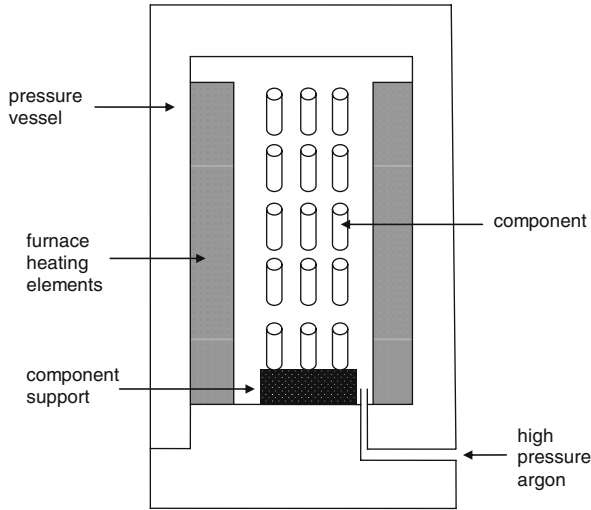
Fig. 1.7 Hot isostatic pressing process

chamber and subjected to a uniform hydrostatic pressure, usually with water, as shown in Fig. 1.6. This results in more even compaction of the powder. Hot isostatic pressing (HIP) is the simultaneous application of pressure and temperature to the powder as illustrated in Fig. 1.7. The components are placed in a furnace which is in turn placed in a pressure vessel. The sintering times are reduced significantly as the pressure adds to the surface energy driving force; the resultant components have both high densities and small grain sizes as the reduction in sintering time restricts the opportunity for grain growth. Hot isostatic pressing is more expensive than cold but gives a better quality product with superior mechanical properties. Often components can be produced with such precision that no further finishing processes are necessary [8].

1.9 Liquid Phase Sintering

An alternative way to increase the density of a ceramic is to use liquid phase sintering. This involves the addition of small quantities of a selected compound to the starting powder that will form a low-melting glass phase which will flow between the powder particles during firing, filling the pore spaces. On cooling, the liquid glass phase solidifies to form a glassy matrix bonding the unmelted particles together. The presence of the liquid phase allows both pore and particle mobility prompting densification. As little as 1% of the glass phase can be present but as it distributes itself along the grain boundaries, it reduces the high temperature strength of the ceramic [3, 9].

1.10 Tape Casting

A route that is used successfully for producing high volumes of thin, flexible ceramic sheets in the 'green' state is tape casting. A concentrated slurry containing deflocculated powders with a high concentration of binders and plasticizers is prepared. A tape is then produced by allowing this slurry to flow beneath a blade, forming a film on a moving carrier substrate; the tape is then dried. Alternatively, the slurry can be poured onto a flat surface and a blade moved over the surface to form a thin film. The dried tape or film, which remains flexible with a smooth surface, can be removed from the substrate and sectioned [6, 10].

1.11 Costs of Powder Processing

Generally speaking, the costs associated with producing powder processed products are significantly higher than those associated with the more traditional manufacturing methods of casting or forming as applied to other classes of material. However, there is much less wastage associated with the powder processes in that the overall material losses are less than 5%. This compares with a 10–30% loss for precision casting, 30–50% for die forging and up to 85% for machining. The predominant cost in powder processing is the initial capital costs; thereafter the processing costs are lower than other manufacturing routes where the running costs dominate. Powder processing can be efficient and economic. It can be easier to handle powders than, for example, molten metals and the resultant components can be of high quality with controlled microstructures and optimized mechanical properties [9].

1.12 Porous Ceramics

Controlled levels of porosity can be engineered into ceramics as part of the manufacturing process. Porous forms of ceramics can be produced using a variety of methods: these include the burn out of polymer spheres – known as BurPS, foamed slip techniques, and the use of reticulated foams.

1.12.1 BurPS

Controlled large scale porosity can be introduced into polymers using the BurPS technique [11]. An organic filler of an appropriate size is combined with the ceramic powder. The 'green' bodies are usually fired at around $500\,°C$ to burn out the polymeric material before sintering at an elevated temperature in excess of $1200\,°C$. The organic particles are burnt out to leave closed pores throughout the sintered ceramic body. Various types of polymer materials have been used for this purpose, including polyvinyl butadiene, naphthalene and polyethylene glycol. The resulting porosity

Fig. 1.8 Scanning electron micrograph of porous hydroxyapatite
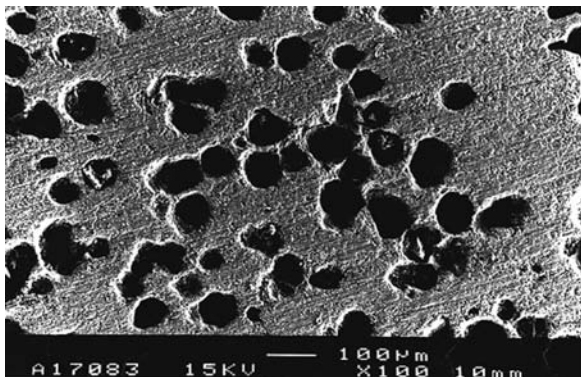
is often spherical – reflecting the original morphology of the polymer powders – and isolated, i.e., not interconnected, as shown in Fig. 1.8, an example of porous hydroxyapatite manufactured using BurPS.

1.12.2 Foamed Slips

Porous forms of ceramic can be produced using foamed slip techniques. Ceramic slips can be foamed in a number of ways which include bubbling air through the slip or the use of high speed rotating blades or whisks to introduce air into the slip. Stabilising agents such as agar or polymer based materials may be added to the slip to enable it to be handled in the 'green' state and set whilst maintaining the porous structure. A porous green body can be produced by casting the foamed slip and allowing it to dry before firing. On sintering the organic stabilizing agents are burnt off to leave a porous ceramic structure [12]. It has been possible to produce ceramics with interconnected porosity using stabilized foam slip methods.

1.12.3 Reticulated Foams

Polymeric foam substrates can also be used to produce porous ceramic materials [13, 14]. The foam is either impregnated or coated with the ceramic slip and allowed to dry prior to firing. During the firing process the polymer substrate is burnt out and the ceramic sintered to give a crystalline structure with interconnected porosity. The density of the ceramic and the size, shape and distribution of the porosity will depend on the method used. Open porous structures with a porosity of up to 70% can be produced by making a positive image of the foam. The slip can be coated onto the foam struts and, on firing, the foam is burnt out to leave an open porous structure, which is a positive image of the porous foam, as seen in Fig. 1.9a. Alternatively, the foam can be completely impregnated such that all the space available is filled
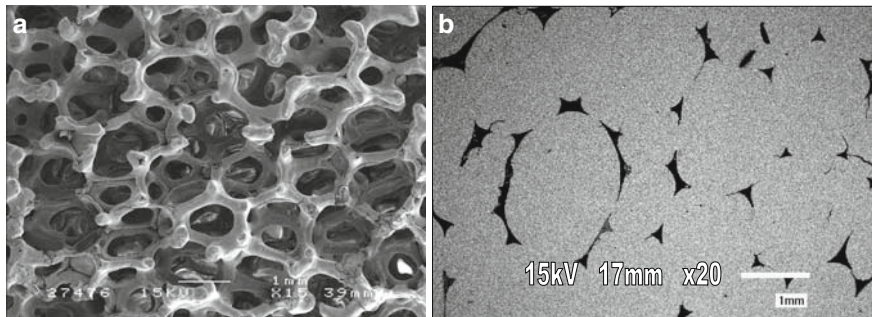
Fig. 1.9 (**a**) Open porous HA structure (**b**) Dense HA ceramic with interconnected porosity

with slip. On firing, a negative image of the foam will be produced as shown in Fig. 1.9b. This material has approximately 5–10% porosity. By tailoring the starting foam, samples with different porosity size and distribution can be manufactured [15, 16]. As before, the porosity will be interconnected throughout the ceramic body, replicating the structure of the foam that has been burnt out.

1.13 Measurement of Porosity in Porous Ceramics

There are a number of established ways to measure porosity in terms of its size and distribution in porous ceramic materials. Following sintering, the pores in a ceramic can be either isolated (closed porosity) or interconnected (open porosity) throughout the body of the ceramic. The accuracy of the measurements will depend on the nature of the porosity and whether it is connected or isolated.

Total porosity can be measured using gravimetric methods where

$$\prod = \frac{1 - \rho_{scaffold}}{\rho_{material}}$$

where $\prod$ is the total porosity, $\rho_{scaffold}$ is the density of the scaffold found by dividing the weight by the volume of the scaffold, and $\rho_{material}$ is the density of the material from which the scaffold is made.

Mercury porosimetry is a very accurate method for measuring both porosity and pore size and distribution. Mercury is forced under pressure into the porous structure. This technique will obviously measure the accessible porosity, i.e., that which is open and available for mercury intrusion. The level of open porosity, π, is given by

$$\pi = V_{intrusion}/V_{scaffold}$$

where $V_{intrusion}$ is the total intrusion volume of mercury, and $V_{scaffold}$ is the volume of the scaffold.

The closed porosity, which will not be accessible to the mercury, is equal to the difference between the two values, II– π.

Open porosity can also be calculated using a liquid displacement method to give a value for the apparent porosity which is a measure of the total volume of the open pores in a porous body. This gives a measure of the interconnectivity and thus of the potential for fluid ingress or in the case of bioceramic materials, the bone ingrowth, into the material. Apparent porosity is calculated as below:

$$\text{Apparent porosity} = \frac{Ww - Wd}{Ww - Ws} \times 100$$

where W_d is the dry weight of the ceramic, W_s is the weight of the ceramic suspended in water, and W_w is the weight of the ceramic after removal from the water.

B, the bulk density, is given by

$$B = \frac{Wd}{Ww - Ws}$$

The true porosity is given by

$$\text{True porosity} = \frac{\rho - B}{\rho}$$

where ρ is the true density of the material.

For a porous ceramic, the fraction of closed pores can then be calculated by subtracting the apparent porosity from the true porosity [4].

Scanning electron microscopy combined with image analysis software can also be used to estimate porosity size and distribution although it must be borne in mind that this technique uses two dimensional images of three dimensional features.

More recently, technological advances have meant that porosity size and distribution can be measured using the techniques of microcomputed tomography (micro-CT) imaging and analysis. Data collected from isotropic slices of three dimensional porous ceramic scaffolds can be processed to generate three dimensional images yielding quantitative information on the scaffold morphologies.

1.14 Surface Engineering

Surface engineering processes can be used successfully to either modify existing surfaces or to apply coatings. Their use is extensive across a range of materials and will be dealt with in a separate chapter. However the processes relevant to ceramics will be reviewed in this section. Coatings can be applied for a diversity of reasons. In the field of bioceramics they can be applied to increase hardness, improve wear and corrosion resistance or to increase biocompatibility and promote osseointegration of a ceramic component. One technique that has been used successfully to increase the wear resistance of the bearing surface of hip joint replacements has been the ion

implantation of nitride into the surface of titanium. Another widely use technique is the application of thermally sprayed HA coatings onto metallic prosthetic stems to promote bone ingrowth into the surface of cementless components.

1.14.1 Ion Implantation

Ion implantation is a low temperature process in which ions penetrate the surface of the component to modify rather than coat it. In titanium implants, nitrogen ions penetrate approximately $0.1 \, \mu m$ below the surface without causing any distortion. Consequently there will be no measurable dimensional changes to the component. The mechanism is such that implanted ions are scattered beneath the surface of the substrate. Either they become trapped in dislocations and other defects pinning them down or they form metastable nitrides which decompose to give martensitic hardening. The result is that compressive forces are generated on the surface which improve both fatigue and corrosion resistance. Interestingly, the enhanced properties are retained to depths of up to 100 times more than the original implanted depth. The mechanism by which this occurs is not fully understood but it has been suggested that the nitrogen ions diffuse ahead of the wear front. The process is expensive and can account for 10–30% of the original cost of the component but a potential increase of several hundred times in the 'in-service' life of titanium alloys for prosthetic use has been reported.

1.14.2 Thermal Spray Coatings

Hydroxyapatite is used in the form of a coating to encourage direct apposition with bone to enhance the early fixation of cementless stems in total joint replacement. The coatings can be applied to the substrates using a variety of processes but commercially the usual route is via a thermal spray process which can take one of several forms – air plasma spraying (APS), vacuum plasma spraying (VPS), high velocity oxyfuel spraying (HVOF) or detonation spraying (DGUN). The first three of these processes involves heating the hydroxyapatite powder in an induced plasma of ionized gases and accelerating the partially melted particles towards a metal substrate where they impact and freeze to form the coating [17, 18]. Thermally sprayed coatings have a characteristic layered structure which is illustrated in Fig. 1.10. Individual powder particles assume splat shapes on the substrate oriented parallel to the substrate surface producing high levels of anisotropy and leaving isolated pores. The result is a coating with a porosity that varies from 2% to 20%, which contains both amorphous and crystalline phases. Small particles have a tendency to melt rapidly and evaporate. Large particles may only partially melt leaving an amorphous splat with a crystalline core. The original particle size and morphology, the plasma gas mixture, working distance, dwell time of the particles, and substrate temperature

In coming particle with unmolten core

Splats, voids and
unmelted particles
making up coating

Substrate
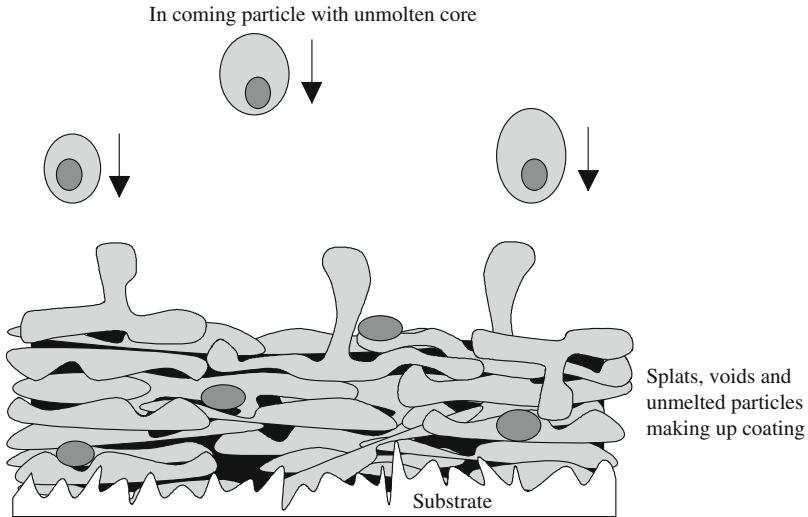
Fig. 1.10 Schematic diagram showing a cross-section through a thermally sprayed coating

will all influence the final structure of the ceramic coating. They must therefore be carefully controlled to optimise the performance of the coating in vivo.

Vacuum plasma spraying is carried out in an argon atmosphere at a temperature of $>10,000\,^\circ$C, as illustrated in Fig. 1.11. The dwell time is a fraction of a second and the substrate temperature typically $300\,^\circ$C. The starting powders for the process are angular with a diameter of approximately $100\,\mu$ as shown in Fig. 1.12a. They have been calcined and crushed. X-ray diffraction analysis shows that they are 80% crystalline in nature prior to spraying. The resultant coatings have a crystallinity of 60%, a porosity of 6%, and are in the order of $40\,\mu$m in thickness [19].

By way of contrast, DGUN spraying is carried out in an oxygen and acetylene atmosphere; it involves an explosive spark discharge at temperatures in excess of $3,000\,^\circ$C and particle speeds of 600 m/s. The coatings produced have a high density and superior adherence to the substrate. The initial starting powders are shown in

Anode

Plasma
Flame

Gas

Substrate

Cathode
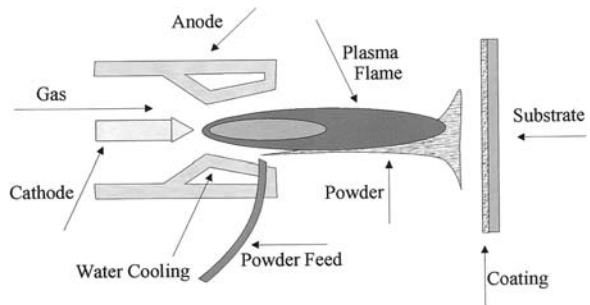
Powder

Fig. 1.11 Schematic diagram
showing the vacuum plasma
spraying technique

Water Cooling

Powder Feed

Coating
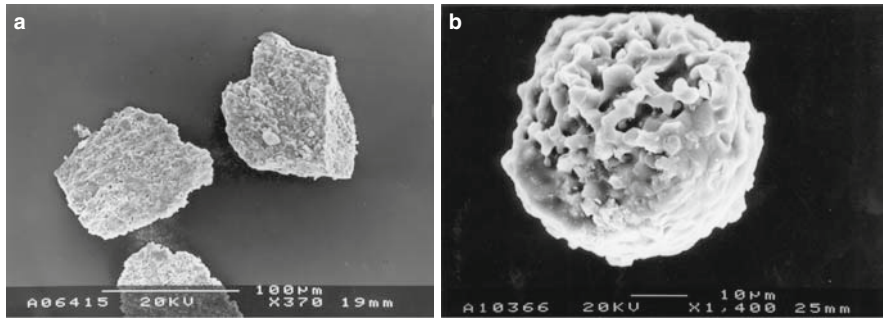
Fig. 1.12 (**a**) SEM micrograph showing the angular VPS HA spray powder (**b**) SEM micrograph showing DGUN HA spray powder

Fig. 1.12b. They have been produced by freeze drying and are characteristically spherical in shape. They have a crystallinity of 80% prior to spraying; this is reduced to 34% after spraying and the coatings typically have a porosity of 1.5% and a thickness of 70 μm [19].

There is much debate in the literature with regard to the ideal morphology of hydroxyapatite coating. Different phases are known to have different solubilities in the body and therefore different dissolution behavior. Amorphous HA is known to be more soluble in aqueous solutions than crystalline HA. The biological performance of the coating system will depend on a number of factors – the size and distribution of the porosity, microcrack formation, residual stresses, the oxide content and the amorphous: crystalline ratio.

1.15 Glasses and Glass-Ceramics

1.15.1 Glasses

Ceramics can also exist in an amorphous or glassy state; a typical example is silica which can exist in both crystalline and amorphous forms. The relationship between the liquid, crystalline and glassy states is shown in Fig. 1.13. A glass is a material that has solidified and become rigid without forming crystals. It is sometimes referred to as a supercooled liquid. All materials have a specific melting or freezing point, T_m. From Fig. 1.13 it can be seen that as the liquid cools from an elevated temperature to T_m there is a linear relationship between volume contraction and temperature. If the material is cooled slowly, on reaching T_m, where the liquid forms a solid, there is a significant contraction in volume indicating that the material has organized itself internally to form a regular crystalline structure which occupies less space than the disorganized liquid structure. In contrast, if the liquid is cooled rapidly, there is no time for this reorganization to take place. In this situation, on reaching T_m, the slope of the curve remains the same, indicating that there has been

Fig. 1.13 Density:
Temperature relationships
between amorphous glass
and crystalline solid
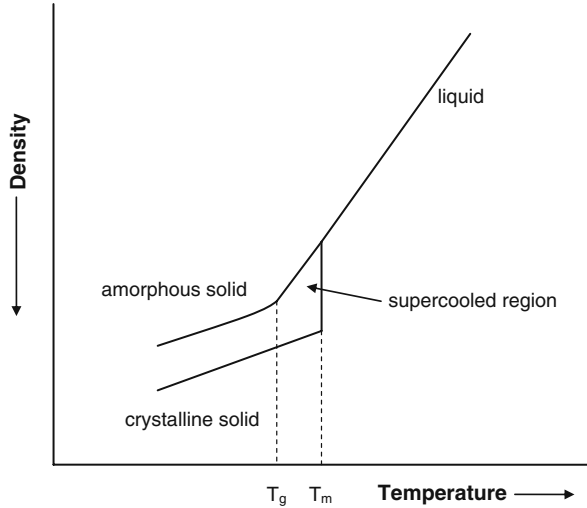
no change in the structure, until a point T_g is reached. T_g is the glass transition tem-
perature of the material, where the glass is deemed to have formed a solid. At T_g,
there is a change in the slope but no sudden marked drop in volume as observed with
the cooling curve for the crystalline material. Unlike the melting or freezing point,
T_g is not a well defined temperature and depends on the cooling rate. The faster the
cooling rate, the higher the measured value for T_g. Between T_m and T_g the material
is described as being a supercooled liquid. Below T_g, the material is solid and in a
what is termed a glassy or amorphous state which is very similar, structurally, to the
liquid state [1, 2].

Glasses can be described as having a metastable structure which, with time, will
very slowly change to a crystalline state. They are also referred to as vitreous solids
and if they do transform to crystalline materials, they are said to be devitrified.
However, at room temperature this process is infinitely slow and would take thou-
sands of years. Raising the temperature of the glass increases the molecular mobility
within the structure and devitrification can then take place at a faster rate. If there
are foreign particles present they will enhance the devitrification process by acting
as nucleation sites for crystallization. This can be a disadvantage during the pro-
duction of glass as the strength may then be reduced and the glass may become
opaque. However it is a useful attribute when manufacturing glass-ceramics as will
be described in detail in a later section.

Glassy structures can be formed by joining together ionic groups to form a solid,
non-crystalline structure with an open framework. A common example is silica glass
which is made up of silica tetrahedra joined together in a relatively random arrange-
ment to form a continuous network with short range order as shown in Fig. 1.14a.
Pure silica has a high melting point and is therefore difficult to fabricate due to its
high viscosity. Adding soda lime, Na_2O, to silica reduces its viscosity significantly,
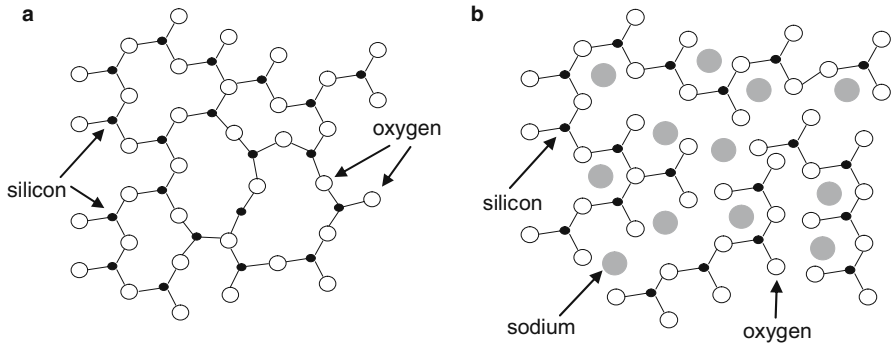
Fig. 1.14 (**a**) Network structure of glass (**b**) Open network structure of silica glass

making it easier and cheaper to process. A structural feature of silica is the open nature of the network which can accommodate atoms of different species. In, for example, soda-silica glass, up to 15% Na_2O is added as a network modifier. The effect is shown in Fig. 1.14b. If sodium is added to pure silica the continuity of the network is disrupted. In pure silica each oxygen atom is bonded to two silicon atoms. The valency of silicon is four (Si^{4+}) whereas that of sodium is one (Na^{1+}). Therefore, when sodium is added to silica, in order to maintain electrical neutrality in the structure each silicon atom must be replaced by four sodium atoms. As a consequence of adding the sodium, the network is disrupted and some oxygen atoms are only bonded to one silicon atom and become 'non-bridging'. Sodium is therefore described as a network modifier. The effects are beneficial in terms of processing as the result of adding the sodium is that the viscosity is markedly reduced. However there is a limit to the extent to which the network can be modified, as above a certain level there will be so much disruption to the network and so many non-bridging oxygens present, that the network structure will no longer be able to form and the glass will crystallize [1].

Alternatives to adding network modifiers to open up the network structure exist in that glass can be modified by the addition of network or glass formers. As the name suggests, these encourage network formation and control crystallization in the glass to give a type of material which is described as being a glass-ceramic.

1.15.2 Glass-Ceramics

Glass ceramics can be defined as materials that start off as a glass and are processed in a two step operation to become devitrified forming fine-grained polycrystalline materials. In the first step the glass is heat treated to a temperature in the range 500–700 °C. The objective is to produce a significantly large number of well dispersed nuclei from which crystals can subsequently grow to form the fine grained structure. A nucleating agent, such as titanium dioxide, is commonly added to induce the

Fig. 1.15 Nucleation and growth of crystals in glass ceramics with time
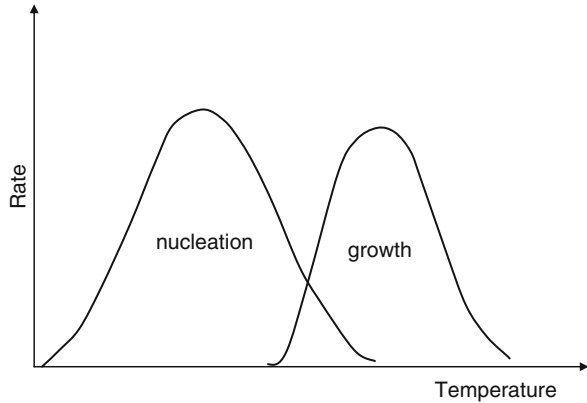
crystallization process. Once the nuclei have been formed, crystal growth is promoted by raising the temperature to 600–900 °C where the crystal growth rate is at a maximum, as shown in Fig. 1.15. Crystals grow from the large number of nuclei; the greater the number of nuclei present originally, the finer the crystal structure will be in the resulting glass-ceramic. Some 10^{12}–10^{15} nuclei per cubic centimeter may be used to produce the very small fine-grained crystals desired. Nucleation is complete when all the small crystals, which have grown with a random orientation, impinge on one another forming a fine-grained, pore free, solid, crystalline structure. The grain size range would typically be 0.1–1.0 μm with at least 90% crystalline material present. Glass-ceramics, like glass, have negligible porosity and mechanical properties that are intermediate between those of a glass and those of engineering ceramics such as alumina and zirconia [2, 4].

The advantages of glass ceramics include a low coefficient of expansion – giving good resistance to thermal shock – combined with relatively high mechanical strength properties and thermal conductivity. They can be either transparent or opaque and produced using conventional glass-forming processes which enables mass production. Commercially the first glass-ceramics were manufactured under common trade names such as Pyrex which contains ~15% B_2O_3 as a glass former. Such additives have been used to produce oven-to-table ware and laboratory glassware due to their superior thermal shock resistance and high thermal conductivity. Subsequently they have been adapted for applications which include electric insulators, substrates for printed circuit boards, cladding, ceramic hobs and more recently in the field of medical materials as bioactive implants.

1.15.3 Bioceramics

Many different types of bioceramic have been developed to replace hard tissues in the human body. This can be for repair or replacement necessary as a consequence

of disease, trauma or the effects of the ageing process. Alternatively, as in the case of coatings, they can be used as a means of promoting bone ingrowth and direct bone apposition at the prosthetic interface. Consequently there are many different types of bioceramic currently in clinical use in a diversity of applications and devices. This next section will begin with a short review of bone as a material, before considering the range of ceramic materials that are currently employed as bone replacement materials.

1.15.4 Bone

Bone is a complex material that comprises an inorganic component, hydroxya-patite, $Ca_{10}(PO_4)_6OH_2$, combined with an organic matrix of collagen. The mineral accounts for almost two thirds of the weight of bone and occupies 40% of the volume.

Bone has several levels of structural organization. On the macroscale there are two forms of bone – compact and cancellous. Compact bone has the appearance of a solid continuous mass in which the structural elements can only be seen with the aid of a microscope. Cancellous, also known as trabecular or spongy bone, is a three dimensional network of bony spicules which combine to give an overall 'spongy' appearance. In a typical long bone, the shaft consists of a thick layer of compact bone, the ends of which are primarily trabecular bone covered in a thin layer of compact bone. The relative quantities of each type of bone and the architecture vary characteristically for each bone in a manner which reflects its overall shape, position and functional role [20].

On a microstructural level bone consists mainly of a number of irregular cylindrical units known as Haversian systems or osteons which run parallel to the long axis of the bone. A scanning electron micrograph of a polished transverse section is shown in Fig. 1.16a. It can be seen that each system has a central canal which is surrounded by concentric lamellae of bony tissue. The number of lamellae vary from 4 to 20. The central cavity contains one or two blood vessels and is 15–20 μm in diameter. The outer limits of each Haversian system are demarcated by cement lines which appear as bands approximately 1 μm in width. The overall diameter of an Haversian system varies between 100 and 200 μm. Other distinguishing features are the irregular shaped sections between the Haversian systems known as interstitial areas and the circumferential lamellae which extend uninterrupted around the inner an outer surfaces of bone. The longitudinally oriented Haversian systems are connected with the free surface of the long bone and one another along transverse channels known as Volkmann's canals as seen in the longitudinal section shown in Fig. 1.16b. Within the lamellae surrounding the Haversian canal there are a number of small lenticular shaped pores known as lacunae, each of which is filled with a bone cell or osteocyte. These lacunae are approximately 10 μm in length and are connected to one another by small radiating channels ∼0.2 μm in diameter. These small canalicular processes provide an interconnecting system of channels
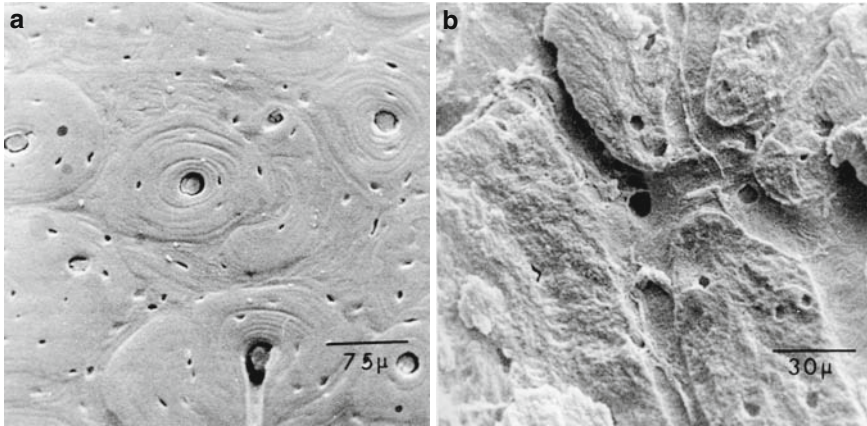
Fig. 1.16 (a) SEM of polished transverse section of compact bone (**b**) SEM of longitudinal section through compact bone demonstrating interconnectivity of channels

throughout the bone necessary for the circulation of nutrients and removal of waste products. Hence any porous ceramic material which is to act as a scaffold for bone ingrowth must allow for this very specific geometry.

The mineral component in bone has long been recognized as being closely related crystallographically to the group of mineral apatites. The unit cell of hydroxyapatite is hexagonal with dimensions of a = 9.42 Å and c = 6.88 Å. Bone mineral has been identified as a non-stoichiometric, poorly crystallized, highly defective form of mineral apatite, $Ca_{10}(PO_4)_6OH_2$ which contains small quantities of cations and complex anion groups such as Ca^{2+}, PO_4^{3-}, and CO_3^{2-}, in addition to even smaller quantities of Mg^{2+}, Fe^{2+}, F^-,Cl^-, Na^+, and K^+. In some cases this is perceived as being advantageous, for example the substitution of OH by F results in a contraction along the direction of the *a* axis, which reduces the solubility and strengthens the structure such that teeth become more resistant to caries. Biological apatites therefore differ from naturally occurring mineral apatite in terms of their composition and crystallinity and, as a consequence, their mechanical properties. Substitutions of one type of atom or ion for another are common but must occur in an orderly fashion so that overall electrical neutrality is maintained.

Mechanically, bone is an interesting material with a unique combination of properties. It can be regarded as a composite material on a number of levels. The main components are the bone mineral which accounts for in excess of 60 wt%, the organic collagen which accounts for over 25 wt% and liquids which make up the balance. Table 1.1 compares the weight and volume contributions of each of these main components in bone. Both the collagen and the bone mineral form continuous networks throughout bone and as such can be isolated as seen in Fig. 1.17. The organic component can be removed by heat treatment to leave the mineral in the form of a continuous network throughout the osteon, as seen in Fig. 1.17a. Similarly, the continuous organic network around the central canal of the osteon is shown

Table 1.1 Relative weight and volume percentages of the main constituents in bone

constituent	weight %	volume %
mineral	61.2	38–42
collagen	26.6	38–42
liquid	12	20–24
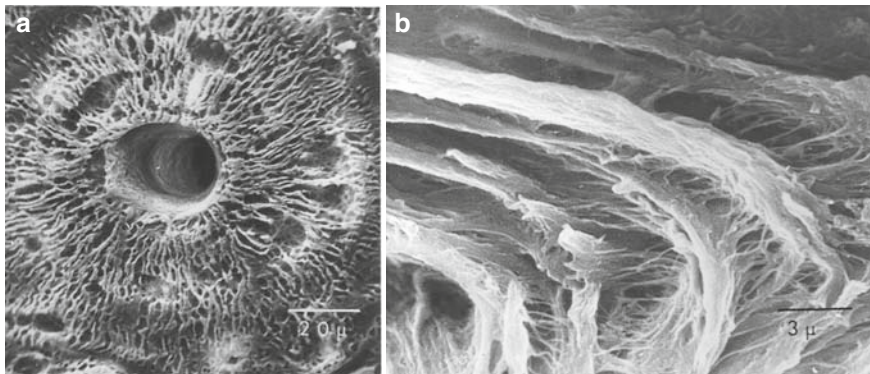
Fig. 1.17 (**a**) SEM of the mineral component isolated in an osteon (**b**) SEM of the organic component within the lamellar structure of an osteon

in Fig. 1.17b where the mineral has been chemically removed. Combining the relatively brittle mineral with the flexible collagen results in a composite material that has a modulus of elasticity in the order of 17–20 GPa, a tensile strength of 120–150 MPa and compressive strength of 100–160 MPa. In addition, bone can have a strain to failure value of up to 4% depending on the rate of application of load. Finding a successful bone replacement material that can in any way match these properties therefore presents an interesting challenge.

1.15.5 Medical Ceramics

Clinically, ceramics are currently used in a variety of forms for a diversity of applications. Dense zirconia and alumina ceramics are used as both femoral head and acetabular cup components in total hip joint replacements where they both articulate and load bear. Porous forms of calcium phosphate based ceramics are used primarily in non-load bearing positions in reconstructive or bone replacement surgery. They can be in the form of small blocks or granules of varying composition with a range of mechanical properties. Bioglass ceramics are also used for a variety of non-loadbearing applications. Hydroxyapatite is used as a bioactive coating on prosthetic stems to promote ingrowth in cementless fixation. It can also be combined with a polymer to give a composite material used commercially to produce cochlear implants.

Ceramics that are used for implant purposes can be classified according to the response they elicit in the body [21]. Accordingly, they can be described as being nearly inert, resorbable or bioactive. The last of these classifications is further divided into osteoconductive or osteoinductive materials. In nearly inert materials there is no chemical or biological bond at the interface between the implant material and the host tissue; there is relative movement and there can be development of a fibrous capsule around the implant both in soft and hard tissues. Examples include alumina and zirconia hip components. Resorbable implant materials are designed to resorb over a period of time and be replaced by natural tissue without the development of a fibrous capsule and with no adverse biological response to the resorbing material; examples include bone graft substitute materials, typically based on calcium phosphate. Bioactive implants are designed to be intermediate between resorbable and nearly inert. They will elicit a specific biological response at the interface between the implant and the surrounding tissues promoting the formation of a physical bond. Bioactive glasses and glass-ceramics for dental and orthopedic applications fall into this category. Osteoconductive implants have the ability to promote the growth of new bone from adjacent bone tissue onto the material's surface. The implant acts as a stable scaffold onto which the tissue can grow and can often provide some additional stimulus to promote this growth. Osteoinductive implants have the ability to initiate and promote the growth of new bone without the need for intimate contact with the bone.

1.15.6 Biomedical Use of Bioceramics

Ceramics with their known low wear rates could be seen as having distinct advantages in load bearing orthopedic applications. However, historically, there have been difficulties with components failing due to poor quality of materials, suboptimal design and surgical technique. The reasons for these early failures are now fully understood and as a consequence improvements in both materials properties and processing have been introduced. These, combined with the use of modular cups and improved surgical experience, have resolved the problems initially associated with the early failures The two ceramic materials most widely used in solid form for implant purposes are alumina and zirconia [21–27].

1.15.7 Alumina

Alumina, Al_2O_3, or aluminum oxide, also known as corundum, can take several common forms. The substitution of small quantities of chromium or iron in the crystal structure results in the gemstones ruby and sapphire respectively. As is characteristic of ceramics, alumina is well known for its good electrical and thermal insulation properties combined with high temperature chemical stability. As such it has been used in a diversity of areas including cutting tools, acid resistant valves, crucibles and space shuttle tiles – as well as for medical applications.

Alumina was first introduced for medical use in 1969 and since this time over 200 million alumina heads and 300,000 acetabular cup liners have been used in total hip joint replacement (THR) operations. Medical grade alumina has since gone through almost four decades of evolution. Originally there were incidences of failure related to implant design, fixation problems and materials issues. During the 1970 s the failure rate was as high as 10%. The quality of the material had not been optimized for implant purposes and as a consequence both variable grain size and residual porosity were sources of weakness. By eliminating flaws that could act as crack initiation sites and refining the grain size and distribution, the properties of the materials were significantly improved. Process changes, included the introduction of hot isostatic pressing which resulted in a lower porosity, higher density material. In addition, the HIPing process resulted in a significant reduction in the grain size of the alumina as the shorter dwell time meant that grain growth during the sintering cycle was reduced. This finer grain size and uniform grain size distribution resulted in improved materials properties and therefore more reliable alumina components.

The introduction of laser engraving, surface polishing, and 100% proof stressing have also had an impact on implant reliability. Non-destructive testing has now been incorporated as part of the final inspection process for alumina components. As part of this process the finished components are subjected to high levels of stress in excess of those which they are expected to experience in vivo. In this way, any components that contain flaws above the critical size for crack propagation will fail by fracturing. Thus any components likely to fail are removed, eliminating the risk of failure in vivo. Coupled with design improvements incorporating modular systems, all these improvements have resulted in a more reliable components with a predicted failure rate of 4 in 100,000 or 0.004%.

Medical grade alumina is a single phase material comprising alpha-alumina or corundum. It is a chemically pure material with less than 0.1% grain boundary impurities (Table 1.2) and must comply with ISO 64749 and ASTM F 60310 specified minimum chemical, mechanical and physical properties for the material. The cooling cycle after sintering is carefully controlled to avoid excess formation of a glassy phase at the grain boundaries. Alumina is thermodynamically stable up to a

Table 1.2 A comparison of the main properties of alumina and zirconia

Property	Alumina	Zirconia
Chemical composition	99.9% Al_2O_3, MgO doping	97% ZrO_2, 3%Y_2O_3
Density	$\geq$ 3.97 g/cm^3	$\geq$ 6.08 g/cm^3
Porosity	0.1%	0.1%
Bending strength	500 MPa	500–1000 MPa
Compression strength	4100 MPA	2000 MPa
Young's modulus	380 GPa	210 GPa
Poisson's ratio	0.23	0.3
Coeff. thermal expansion	$8 \times 10^{-6} K^{-1}$	$11 \times 10^{-6} K^{-1}$
Thermal conductivity	30 W/mK	2 W/mK
Hardness	> 2000 HV	1200 HV

temperature of 2000 °C. It is a hard material, 1900 Hv (second only to diamond) that can be highly polished (surface roughness, Ra, 0.01). It has extremely good wear properties with reported wear rates, when tested in a simulator, of less than $0.1 \text{ mm}^3/10^6$ cycles in vitro and $1–5 \text{ mm}^3$/year in vivo. Table 1.3 shows its average wear rate in comparison to other combinations of orthopedic bearing materials. Simulator studies have shown that the amount of wear debris produced by ceramic couples is in the order of ten times less than that for metal:metal couples under in vitro hip joint simulation [27]. The debris produced in vivo is well tolerated by the body with a relatively benign tissue response.

In 1988, the FDA designated alumina femoral bearings as Class II and as such they are required, mechanically, to comply with minimum strength criteria which include a minimum average burst strength of 46 kN when axially loaded on a hip stem trunion. This is approximately 60 times the average body weight and four times the breaking strength of healthy human femur. In addition, the components must be able to endure 107 cycles under a maximum sinusoidal compressive load of 14 kN – approximately 18.5 times the average body weight. This far exceeds the test requirements for metallic stems which have a specified fatigue load of 3.3–4.9 kN or six times the average human body weight.

Despite the improvements that have been made and the standards that have been specified, some inherent limitations remain with the use of alumina:alumina bearings. The properties of alumina are such that it can successfully be combined with titanium or stainless steel femoral stems but it is not sufficiently strong to pass the FDA strength requirements for use with cobalt chrome stems. Also, the size of heads available is restricted to the 28–36 mm range, precluding its use where a smaller diameter head may be required.

Alumina has also been used successfully to make artificial teeth for implantation. The reported success rate is 90% after 5 years. Different tissue responses have been observed in relation to the alumina implants – both connective tissue and mineralized bone have been found depending on the relative motion between the implant and the surrounding tissues.

Table 1.3 Relative wear rates of different bearing combinations	**Combination** **Wear rate/annum in microns**

Combination	Wear rate/annum in microns
CoCr/HDPE	20–200
metal/metal	<10
Al_2O_3/HDPE	<10
Al_2O_3/Al_2O_3	<1

1.15.8 Zirconia

Zirconia femoral heads were first implanted in 1985; since then, over 300,000 surgical implantations have been recorded. The chemistry of zirconia is more complex than alumina. Zirconia can exist in three different phases – monoclinic, tetragonal

and cubic – depending on the temperature and environment. The monoclinic phase transforms to a tetragonal crystal structure at 900–1220 °C. If the temperature is increased further, the tetragonal phase then transforms to cubic at 2370 °C. There is a 3–5% volume increase associated with the tetragonal to monoclinic transformation. As a result of this expansion in volume, as pure zirconia cools down from its sintering temperature it can be prone to cracking.

However, the existence of this phase transformation and the associated volume increase can be used to advantage to improve both the toughness and strength of zirconia. Orthopaedic grade zirconia contains small quantities of stabilizing agents, 5.2 wt% Y_2O_3, 2 wt% HfO_2, and 0.5 wt% Al_2O_3 added to stabilize the phases, thus avoiding the cracking associated with the expansion in volume. More than 95% of the zirconia is in the metastable tetragonal phase; the remainder is either cubic or monoclinic zirconia. The grain size is kept below 0.5 μm to ensure the tetragonal phase remains stable in physiological conditions. Wear rates reported for zirconia on polyethylene in a joint simulator are in the order of 10.9 $mm^3/10^6$ cycles. This compares well with the 15.7 $mm^3/10^6$ cycles reported for cobalt chrome heads articulating against polyethylene in comparable circumstances.

A concern with the use of zirconia as a bearing material is the possibility of transformation aging – a spontaneous transformation from the tetragonal to the monoclinic phase – which has been observed on the surface of ceramic femoral heads. The indications are that this transformation can take place in the presence of aqueous fluids in the temperature range 65–500 °C. The rate at which the aging occurs will be affected by the grain size, the composition of the zirconia and the stabilizer content. It is suggested the optimum grain size should be 0.3–0.6 μm to retard this aging. The consensus seems to be that the transformations observed do not have a significant effect on the strength and wear properties of the heads but they remain a concern.

Overall, zirconia is a tougher and more fracture resistant material than alumina. It also has the advantage that it can be used to produce small diameter heads down to 22 mm and it is suitable for any type of spigot material – titanium, stainless steel and cobalt chrome. The relative properties of the two ceramics are compared in Table 1.2.

1.15.9 Hydroxyapatite

Hydroxyapatite can be produced in both dense and porous forms for implant purposes. Dense hydroxyapatite products can been manufactured by one of two powder processing routes. The first is cold pressing followed by sintering, the second hot pressing where heat and temperature are applied simultaneously. The final product typically has a pore volume of <5%, a pore size of <1 μm. Uses of such material include the filling of bony defects following dental or craniofacial surgery.

Porous forms of hydroxyapatite are currently generating much interest as potential bone replacement materials, where they can be used for augmentation,

replacement and repair of diseased or damaged bone. The size range and distribution of porosity throughout a solid are crucial to its success as an implant material. The porosity can be defined on two levels – macroporosity which relates to pores having a diameter of $>100\,\mu m$ and microporosity which relates to pores with a diameter of $<5\,\mu m$.

The biological response of tissues to the scale and distribution of the porosity has been the subject of several studies. It is now known that the tissue type and rate of ingrowth are affected by the nature of the porosity. A porous surface can offer two possibilities with respect to fixation mechanisms. In the first instance a rough or porous surface can offer the opportunity for mechanical interlocking at the interface. In this a way a physical connection between the implant and the adjacent bone may be formed, enhancing stability at the interface. Alternatively a material with porosity distributed throughout its mass offers the potential for bone or tissue ingrowth. In some cases the porous bioceramic may act as a scaffold, resorbing as the new bone is formed, ultimately resulting in its complete replacement with new bone.

Studies have shown that the minimum pore size needed to ensure that mineralized bone is regenerated is in the order of $100\,\mu m$. Pores less than $10\,\mu m$ in diameter have been shown to inhibit the ingrowth of cells; pore sizes of $10–75\,\mu m$ encourage fibrovascular ingrowth; pore sizes of $75–100\,\mu m$ result in ingrowth of unmineralized osteoid tissue; pore sizes of greater than $150\,\mu m$ facilitate the ingrowth of substantial quantities of mineralized bone. It is now widely recognized that a pore size of $150–300\,\mu m$ is appropriate to allow bone ingrowth as this relates directly to the bone architecture, and in particular the dimensions of the Haversian systems with an average diameter of $100–200\,\mu m$ as seen in Fig. 1.15. In addition to the pore size and distribution, the degree of interconnectivity of the pores has also been shown to have a strong influence on the biological response – maximum interconnectivity promotes ingrowth as there are no 'blind alleys' to prevent tissue ingress [28].

1.15.10 Porous Bioceramics

A range of porous ceramics are commercially available for surgical use. There are two main types – those made from coral or animal bone and those produced synthetically based on hydroxyapatite, tri-calcium phosphate or combinations of the two.

Converting coral for use as a bone graft material was first proposed over 20 years ago [29]. Coralline bone graft is derived from marine corals as the name suggests and is commercially produced under the tradename Pro Osteon™. Essentially, the production process involves converting the calcium carbonate skeleton of the marine coral to hydroxyapatite via a hydrothermal exchange reaction using heat, water, and a source of phosphate ions. The original interconnected porosity of the coral is preserved and the final product can be tailored to have optimal pore diameters for bone ingrowth. Typical pore sizes in the final product are $200–500\,\mu m$ with an overall

porosity of 50–65%. In contrast to this, the commercial product EndoBon is sourced from bovine bone which has been heated to a temperature of >1200 °C. The organic component is burnt off to leave a material which comprises hydroxyapatite. As it is derived from cancellous bone, it has an open porous structure of an appropriate size for new bone ingrowth and can be produced in a variety of geometries. However, concerns over the use of these two sources of material exist. For the coralline materials there are obvious environmental concerns, whilst for the bovine based product general concerns remain over the use of xenografts.

The ideal bone graft material is an autograft from a donor site in the patient as the risk of transmitted infection is minimized. However, there are problems associated

(a) 1100°C

(b) 1150°C

(c) 1200°C

(d) 1250°C

(e) 1300°C

(f) 1400°C
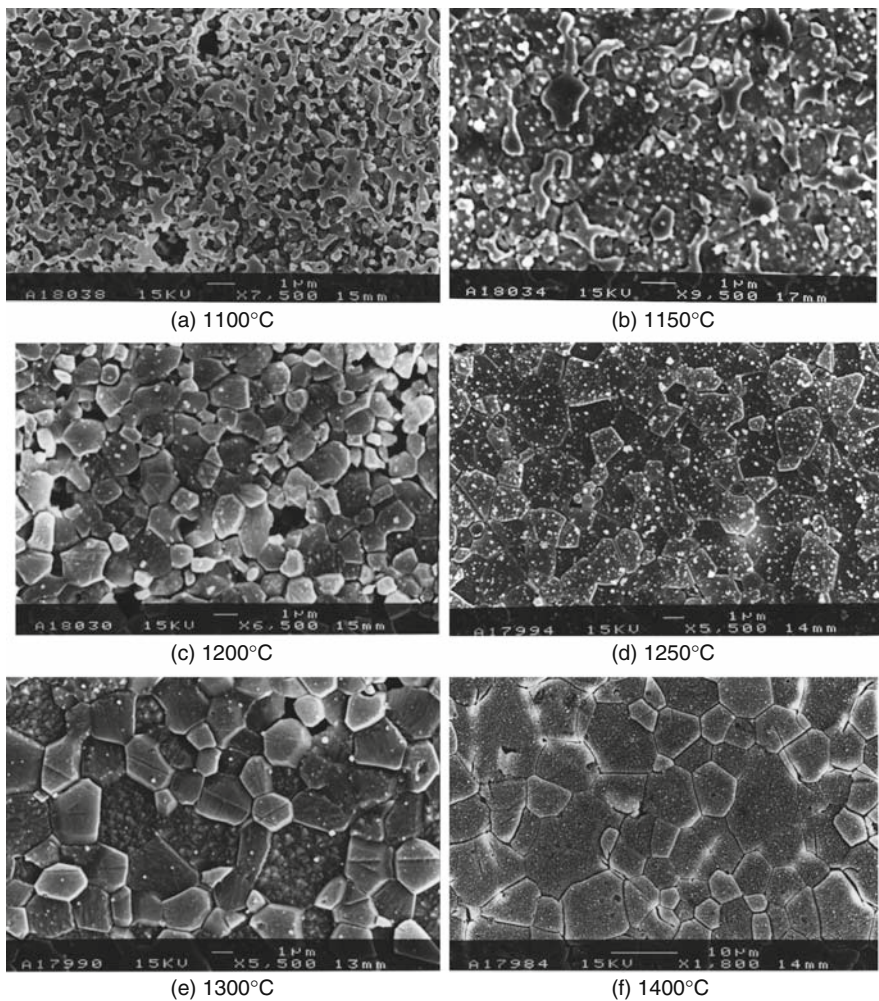
Fig. 1.18 SEM micrographs of porous hydroxyapatite sintered in air for 4 h

with the harvesting of autograft in that the donor site can remain painful for an extended period and there are obvious limits to the supply. The alternative is to use allograft – donated from other patients or sourced from cadavers. Allograft is treated and then stored frozen or freezedried. It has the advantage of being osteogenic but the risks of disease transmission remain and, again, it is in limited supply.

As a consequence, much research has focused on the development of a range of synthetic materials that can be used as bone replacement materials. These are porous ceramics, primarily based on hydroxyapatite, tri-calcium phosphate or combinations of the two. Their success has been attributed to the ability to control both pore size and shape and pore interconnectivity as these are crucial for ingrowth. By tailoring the composition of the bioceramic, the rate of resorption can also be controlled. The sintering temperature is important in this respect as phase transformations can occur at specified temperatures. Calcium deficient HA sintered at 800–900 °C begins to form β-TCP as densification occurs. At 1180 °C this transforms to α-TCP; above 1300 °C α-TCP decomposes into tetracalcium phosphate. Figure 1.18 shows the effect that sintering temperature has on the growth of grains in hydroxyapatite sintered over a range of temperatures.

There is some evidence that biphasic ceramics comprising HA and TCP have certain advantages over pure HA or pure TCP ceramics. HA is mechanically stronger than TCP and has a much lower solubility in bodily fluids – up to 20 times lower than TCP. However the combination of the two means that the superior mechanical stability of the HA can be combined with the greater bioactivity of the TCP. The result is a biologically active material with good mechanical properties. Bioceramics can be successfully produced in many different densities and geometries. Some examples are shown in Fig. 1.19a – dense blocks that can be machined to shape and smaller granules for impaction grafting [16]. Figure 1.19b shows the more porous structure that replicates the cancellous bone architecture [15].

Until recently, because of their relatively poor mechanical properties, porous bioceramics have been limited to use in non-load bearing applications. However, technological advances mean that stronger ceramics can now be produced so this will no longer be such a constraint in the future.
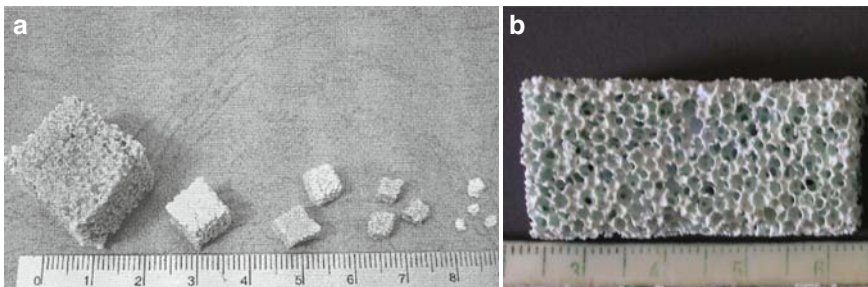
Fig. 1.19 (**a**) Different sized blocks and granules of dense HA ceramic (**b**) Open porous HA ceramic
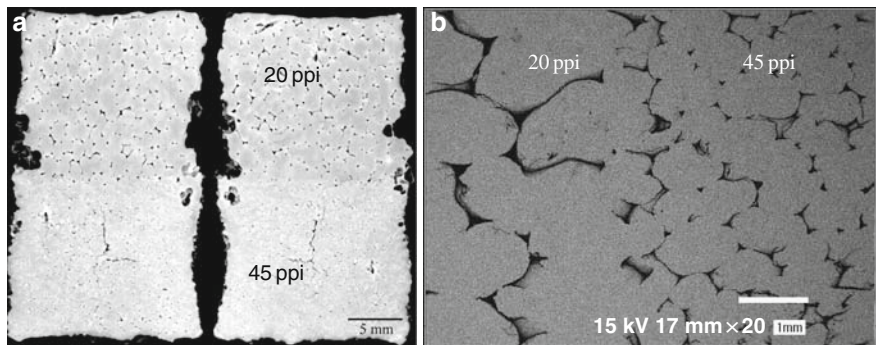
Fig. 1.20 (**a**) Sintered ceramic with two different porosity distributions (**b**) SEM of the interface of HA ceramic produced by joining together two different porosity foams to make a functional gradient material

1.16 Functional Gradient Materials

As has been seen in an earlier section, bone has a complex architecture with several levels of structural organization. In many ways it can be regarded as a functional gradient material in that on every level its structure varies to reflect the functional role of each constituent. Long bones such as, for example, the femur, have a functional gradient in terms of the relative distribution of compact and cancellous bone. The central region primarily comprises a cylinder of compact bone. This changes as the load bearing ends of the bone are approached where the bulk of the bone becomes cancellous, covered only in a thin layer of compact bone. There is therefore an interest in trying to replicate functional gradient materials to match more closely the structure of natural bone.

One approach has been to join together different porosity foams by sewing or press fitting. These have then been impregnated with slip and the foams burnt out (as described in the reticulated foam section earlier) to leave a porosity gradient within the porous ceramic structure as shown in Fig. 1.20a. Microscopical examination indicates that there is no discontinuity at the interface where the two foams have been joined as shown in Fig. 1.20b. Such combinations of material have potential for development as structures that more closely mimic the architecture of bone [16].

1.17 Bone Morphogenetic Proteins

Bone morphogenetic proteins (BMPs) occur naturally in bone where they have the ability to promote natural bone formation and healing. A family of bone morphogenetic proteins have been identified that can be extracted from demineralized bone matrix and processed for use as bone forming agents. A recent development has been the incorporation of BMPs in an appropriate carrier which is then implanted

in a defect site to stimulate repair. The carrier will obviously have an important role to play in future developments in this area. Ideally the carrier should be biocompatible and biodegradable and be absorbed without inducing any adverse tissue response. In addition, it must provide a scaffold for tissue growth and as such needs to be amenable to shaping to fill the void space accurately. Finally it must resorb at a controlled rate to allow sustained release of the BMP in vivo. Several types of carrier have been used experimentally as delivery systems, including HA, TCP and coral [29]. While challenges remain, BMP devices incorporating a range of bioceramic carriers can be seen as having significant potential as bone graft substitutes in the future.

1.18 Hydroxyapatite Coatings

Hydroxyapatite is used extensively as a coating material on prosthetic hip stems to promote early fixation in cementless devices. Some examples of commercial coatings are shown in Fig. 1.21. They are usually applied to the proximal section of the stem which in some cases has been fabricated to provide a mechanical interlock of the stem with the femur. Different manufacturers' designs include the incorporation
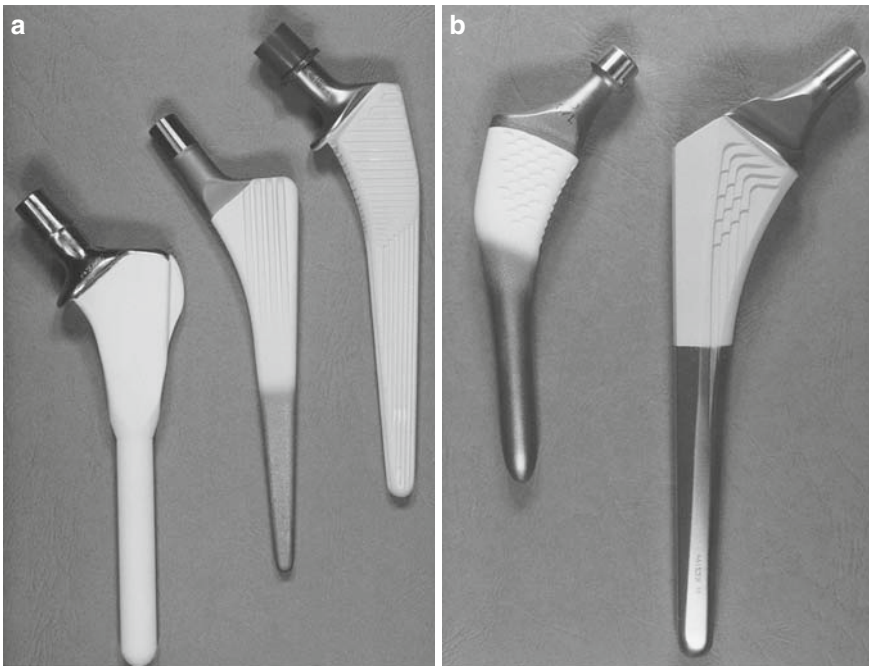
Fig. 1.21 (**a**) HA coated hip stems (**b**) HA coating of stems proximally

of grooves on the stem as in Fig. 1.21a; others have a fishscale geometry on the surface as seen in Fig. 1.21b.

Hydroxyapatite, in many different forms, has been reported to encourage direct apposition with bone. The early bone bonding properties of HA arise from its dissolution behavior. It is proposed that hydroxyapatite dissolves releasing calcium and phosphorus ions into the surrounding system. These can then be reprecipitated as new natural bone. Different hydroxyapatite phases have different solubilities in vivo –HA being up to 20 times less soluble than TCP. Amorphous HA has much greater solubility than crystalline HA. Therefore in terms of the behavior of coatings in vivo, the original morphology and composition will have a profound affect.

There is much debate about the ideal morphology of HA coatings. The original powder particle size and shape will influence the final structure of the coating. Maximum adhesion between impacting particles and the coating can only be achieved if the particles are completely molten and can flow into the substrate surface. However, molten particles freeze rapidly on impact with the surface and form metastable amorphous phases. Control of the degree of melting of the powder can thus control the relative proportions of amorphous and crystalline phase formed in the final coating. In practice, two different powder types are used commercially depending on the spray process. From the scanning electron micrograph of the particles used with the VPS process shown in Fig. 1.12a, it can be seen that the particles are discrete and angular in shape suggesting they have been cleaved. The dimensions vary, with an average particle size of 100–200 μm. By contrast, Fig. 1.12b shows a scanning electron micrograph of the powder used for the DGUN process. The particles, manufactured by a spray drying route, are spherical and relatively small with an average diameter of 50 μm.

A particular problem with applying a ceramic coating to a metal substrate can be the difference in the relative thermal conductivities of the two. This can be accommodated by the use of an intermediate bond coat. For example, before spraying an HA coating onto a titanium alloy substrate a bond coat of pure titanium may be applied to reduce the mismatch. The typical appearance of a thermal spray coating is shown in Fig. 1.22. The surface is very irregular with flat glassy regions that can be attributed to the amorphous phase present and irregular regions comprising the residual crystallites from the spray powder and splat debris. Small particles melt more rapidly, reach their maximum velocity faster, and freeze more rapidly on impact with the substrate in comparison to larger ones. The morphology of the starting powders used to produce the coating can therefore have as great an effect on the structure of the final coating as the spray process itself. Research has shown that for the VPS process the original crystallinity of 80% for the powder is reduced to 60% in the coating. Similarly, for the DGUN process the original crystallinity of 77% in the starting powder is reduced to 34% in the coating [19].

An observation with respect to the production of spray coatings is the need for careful control at all stages. The metallic substrates are normally grit blasted with alumina grit to roughen the surface to promote adhesion of the coating. Despite subsequent cleaning, it is possible for some of the blasting medium to remain embedded in the surface of the substrate. It is possible for the particles to be of sufficient size to

Fig. 1.22 An SEM
micrograph of an as-received
VPS HA coating
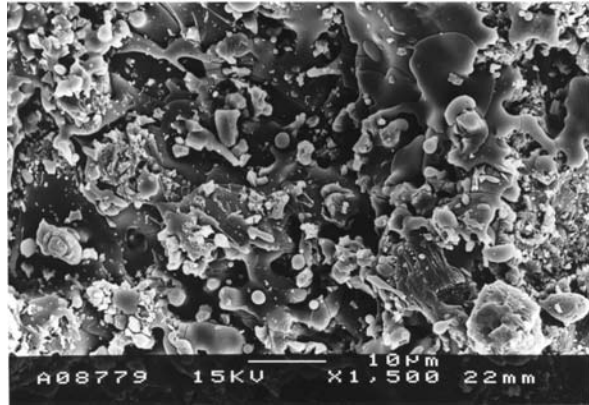
penetrate the thickness of the subsequent spray coating. A major advantage of apply-
ing thermal spray coatings to metallic stems is seen as their potential to encourage
interlocking of the metal stem with the surrounding bone. In this way the micro-
motion at the interface between the prosthetic stem and the bone will be minimized
and thus the potential problem of loosening reduced. Consequently, ceramic coated
implants for hip joint replacement are recommended for use with younger and more
active patients.

1.19 Bioactive Glasses

Several decades ago, in the 1960s, it was first discovered that certain compositions
of glass and glass-ceramics were capable of bonding directly with bone [30]. Since
then, research has shown that careful specification of the glass composition can
result in direct control over the rate of this bonding. Some very specific formula-
tions of glass have been developed that will form bonds directly with soft tissues, in
addition to bone [22].

A wide range of glass compositions have been studied and it has been found that
the bone bonding response can vary depending on the constituents. For example, it
is now known that having more than 3 wt% of alumina in the composition can result
in inhibition of the bone bonding. The glasses currently used for implant purposes
are based on SiO_2-CaO–Na_2O–P_2O_5 and Li_2O–ZnO–SiO_2 systems.

The success of the bone bonding mechanism depends on the formation of a bio-
logically active carbonated HA layer on the glass surface which promotes bonding at
the interface between the bioactive glass and the surrounding tissues. Three specific
criteria must be met in order to render the surface highly reactive when exposed to
an aqueous environment such as that found in the body. These are based on the com-
positional requirements of the glass and are that it must contain <60 mol% SiO_2, a
high Na_2O and CaO content, and lastly a high CaO:P_2O_5 ratio. Consequently, one

Table 1.4 Composition of bioactive glass Bioglass 45S5

45S5 Bioglass	Weight %
SiO_2	45
Na_2O	24.5
CaO	24.5
P_2O_5	6

of the most commonly used bioactive glasses, known as 45S5, has 45 wt% SiO_2, with a $CaO:P_2O_5$ ratio in the order of 5:1. The formulation is given in Table 1.4. It is known that glasses with a lower ratio than this will not bond with bone [24].

In an aqueous environment bioactive glass will undergo a rapid surface reaction which will result in a fast rate of bonding to tissue. The high level of biological activity is facilitated by the free exchange of ions between the bioactive glass and the surrounding bodily fluids. In some situations mineralizing bone is detectable at the interface after only 7 days. By 4 weeks there is complete bonding at the interface and no fibrous capsule has formed. The strength of bond at the interface can be considerable. For example, one study using pushout tests showed that the interfacial bond strength for Bioglass was 15–25 MPa in contrast to pushout strengths of 0.5–3 MPa for unreactive glass. Significantly higher interface strengths of up to 85 MPa have been reported.

One disadvantage of glass-ceramics is their inherent brittleness. However, they do have good resistance to scratching and abrasion. Their mechanical properties have been optimized for their composition and could not be further compromised without adversely affecting their potential for bone bonding. They have a 10-year history of use as middle ear replacements, dental implants, vertebral implants and in the area of periodontal surgery.

1.20 Conclusion

Historically, ceramics have been regarded as brittle materials whose range of applications were constrained by their relatively poor mechanical properties. This chapter has shown that this is no longer the case. Advances in processing technology and a deeper understanding of the structure and properties of ceramics have led to the expansion of their use into a whole host of different areas. Not least of these is the medical sphere where ceramics now have a proven history of clinical use over several decades with a broad range of applications. It is anticipated that future developments in the field of bioceramics will mean that mechanical properties can be further optimized to enable the range of useful applications to be expanded further. Exciting new prospects exist in the growing area of tissue engineering where undoubtedly ceramics will have a central role to play.

References

1. Ashby MF and Jones DRH. Engineering Materials 2: An Introduction to Microstructure, Processing and Design. Pergamon Press: Oxford, 1986.
2. Callister WD. Materials Science and Technology: An Introduction, 5th edn, John Wiley and Sons: New York, NY, 2000.
3. Edwards L and Endean M. Manufacturing with Materials, Butterworth Heinemann: Oxford, 1995.
4. Askeland DR. The Science and Engineering of Materials, 3rd edn, Chapman and Hall, London, 1996.
5. Swain MV, ed. Structure and Properties of Ceramics, Volume 2, Cahn RW, Haasen P, and Kramer EJ, eds. Materials Science and Technology, VCH: Weinheim Alemanha, 1992.
6. Engineered Materials Handbook Volume 4 Ceramics and Glasses, ASM International: S.J. Schneider, 1991.
7. Ravaglioli A and Krajewski A. Bioceramics: Materials, Properties and Applications, Chapman and Hall: London, 1992.
8. Williams DF, ed. Medical and Dental Materials Volume 14, Cahn RW, Haasen P, and Kramer EJ, eds. Materials Science and Technology, VCH: Weinheim Alemanha, 1992.
9. Kalpakian S. Manufacturing Engineering and Technology, Addison-Wesley: Boston, MA, 1989.
10. Watchman JB Jr, ed. Ceramic Innovations in the 20th Century, The American Ceramic Society: Westerville, OH, 1999.
11. Dean-Mo L. Fabrication of hydroxyapatite ceramic with controlled porosity. J Mater Sci Mater Med, 1996, 8: 227–232.
12. Binner JG. Processing of hydroxyapatite ceramic foams. J Mater Sci, 1996, 31: 5717–5723.
13. Fabbri M, Celotti GC, Ravaglioli A. Granulates based on calcium phosphate with controlled morphology for medical applications: physico-chemical parameters and production technique. Biomaterials, 1994, 15: 474–477.
14. Sepulveda P, Bressiani AH, Bressiani JC, Meseguer L, and Konog B. Synthesis and properties of ceramic foams for hard tissue repair. Bioceramics, 2002, 14: 413–416.
15. Gittings JP, Turner IG, and Miles AW. Calcium phosphate open porous scaffold bioceramics. Key Eng Mater, 2005, 284–286: 349–352.
16. Hsu YH, Turner IG, and Miles AW. Fabrication of porous calcium phosphate bioceramics as synthetic cortical bone graft. Key Eng Mater, 2005, 284–286: 305–308.
17. Grainger S, ed. Engineering Coatings – Design and Applications, Abingdon Publishing: Oxford, 1989.
18. Bernecki TF, ed. Thermal Spray Coating: Properties, Processes and Applications, ASM International: Materials Park, OH, 1992.
19. Gledhill HC, Turner IG, and Doyle C. Direct morphological comparison of vacuum plasma sprayed and detonation gun sprayed hydroxyapatite coatings for orthopaedic applications. Biomaterials, 1999, 20: 315–322.
20. Cowin SC, ed. Bone Biomechanics Handbook, 2nd Edn, CRC Press: Boca Raton, FL, 2001.
21. Ratner BD, Hoffman AS, Schoen FJ, and Lemons JE, eds. Biomaterials Science: An introduction to Materials in Medicine, 2nd Edition, Academic Press: New York, NY, 2004.
22. Silver FH. Biomaterials Medical Devices and Tissue Engineering, Chapman and Hall: London, 1994.
23. Park JB and Bronzino JD, eds. Biomaterials: Principles and Applications, CRC Press: Boca Raton, FL, 2003.
24. Nicholson JW. The Chemistry of Medical and Dental Materials, in Connor JA, ed. RSC Materials Monographs, The Royal Society of Chemistry: London, 2002.
25. Parks JB and Lakes RS. Biomaterials: An Introduction, 2nd Edn, Plenum Press: New York, NY, 1992.

26. Shackelford JF, ed. Bioceramics: Applications of Ceramic and Glass Materials in Medicine, Trans Tech Publications, Enfield, NH, 1999.
27. Hutchings IM, ed. Friction, Lubrication and Wear of Artificial Joints, Professional Engineering Publishing, London, 2003.
28. Lu JX, Flautre B, Anseelme K, Hardouin P, Gallur A, Descamps M, and Thiery B. Role of interconnections in porous bioceramics on bone recolonisation in vitro and in vivo. J Biomed Mater Res, 1983, 17: 769–784.
29. Laurencin CT, ed. Bone Graft Substitutes. ASTM International, West Conshohocken, PA, 2003.
30. Hench LL. Bioactive ceramics, in Ducheyne P and Lemons J, eds. Bioceramics: Materials characteristics versus in-vivo behaviour, Ann. NY Acad Sci, 1988, 523: 54.

Chapter 2
Metallic Biomaterials

Robert M. Pilliar

2.1 Introduction – Why Metals?

Metallic biomaterials continue to be used extensively for the fabrication of surgical implants primarily for the same reason that led to their initial selection for these devices many decades ago. The high strength and resistance to fracture that this class of material can provide, assuming proper processing, gives reliable long-term implant performance in major load-bearing situations. Coupled with a relative ease of fabrication of both simple and complex shapes using well-established and widely available fabrication techniques (e.g., casting, forging, machining), this has promoted metal use in the fields of orthopedics and dentistry primarily, the two areas in which highly loaded devices are most common although similar reasons have led to their use for forming cardiovascular devices (e.g., artificial heart valves, blood conduits and other components of heart assist devices, vascular stents), and neurovascular implants (aneurysm clips). In addition, the good electrical conductivity of metals favors their use for neuromuscular stimulation devices, the most common example being cardiac pacemakers. These favorable properties (good fracture resistance, electrical conductivity, formability) are related to the metallic interatomic bonding that characterizes this class of material. While the purpose of this chapter is to focus on the important issues pertaining to the processing and performance of metallic biomaterials and to review the metals that are currently used for implant fabrication, a brief review of fundamental issues related to the structure-property relations of metals in general follows.

Metal processing determines microstructure and that in turn determines properties (elastic constants being an exception since these are virtually structure-insensitive properties dependent on interatomic bond type and equilibrium atom packing as noted below) [1–3]. An understanding of material properties and processes used to achieve desired properties during fabrication of metallic components is critical for ensuring desired performance of implants in service. While mechanical failure is unacceptable for most engineered structures, it is particularly so for surgical implants where failure can result in patient pain and, in certain cases, death (heart valve component fracture, for example) or the need for complicated and life-threatening revision surgery.

2.2 Metallic Interatomic Bonding

Interatomic bonding in solids occurs by strong primary (ionic, covalent, and/or metallic) and weaker secondary interatomic bonding (van der Waals and hydrogen bonding). Metals are characterized by metallic interatomic bonding with valence shell electrons forming a 'cloud' of electrons around the atoms/ions. This is a consequence of the high coordination number, N, (i.e., number of nearest neighboring atoms) that occur with metals (i.e., N = 12 or 8 for many metals). As a result of the close positioning of neighboring atoms and the shared valence electrons, the interatomic bonds are non-directional and electron movement within metal crystal lattices is easier than in ionic or covalently bonded materials. This fundamental distinguishing characteristic of metals results in the relative ease of plastic deformation (i.e., permanent deformation on loading above their yield stress) as well as the high electrical and thermal conductivities of metals. Most metals used for implant fabrication have either close-packed atomic structures with N = 12 with face-centered cubic (fcc) or hexagonal close-packed (hcp) unit cells, or nearly close-packed structures with N = 8 forming body-centered cubic (bcc) structures. Less commonly, tetragonal and orthorhombic as well as other unit cells do occur with some metallic biomaterials. The equilibrium distance between atoms defining the unit cells of these crystals and the strength of their interatomic bonding are determined by intrinsic factors such as atom size and valency as well as extrinsic factors (temperature, pressure). In addition to ease of deformation to desired shapes, the ability to deform plastically at high loads results in another very important feature namely an ability to blunt sharp discontinuities (through plastic deformation) thereby reducing local stress concentrations resulting in relatively high fracture toughness. As noted below, these desirable characteristics are dependent on proper selection of processing conditions for material and part formation.

2.3 Crystal Structures – Atom Packing in Metals

The common metallic biomaterials (i.e., stainless steel, Co-based alloys and Ti and its alloys) form either face-centered cubic, hexagonal close-packed or body-centered cubic unit cells at body temperature with ideal crystal lattice structures as shown in Fig. 2.1. Real metal crystals, in contrast to these ideal atom arrangements, contain lattice defects throughout (vacancies, dislocations, grain boundaries – Fig. 2.2). The presence of these defects, (point, line and planar defects), have a strong effect on mechanical, physical and chemical properties.

Using a simple solid sphere model to represent atom packing, arrangement of spheres in the closest packed arrangement shown in Fig. 2.3 results in either a face-centered cubic structure (2-D planar layer stacking sequence as ABCABC... – Fig. 2.3a), or a hexagonal close-packed structure (ABABAB... stacking sequence – Fig. 2.3b). The selection of the preferred arrangement for a close-packed metal depends on the lowest free energy form under given extrinsic conditions (temperature and pressure). While many metals used for implant applications form close-packed structures over a certain temperature range (e.g., Ti and its alloys are hcp

Fig. 2.1 Unit cells for face centered cubic (fcc), hexagonal close-packed (HCP) and body-centered cubic (bcc) crystal structures (Courtesy of Scott Ramsay, Adjunct Professor, University of Toronto)
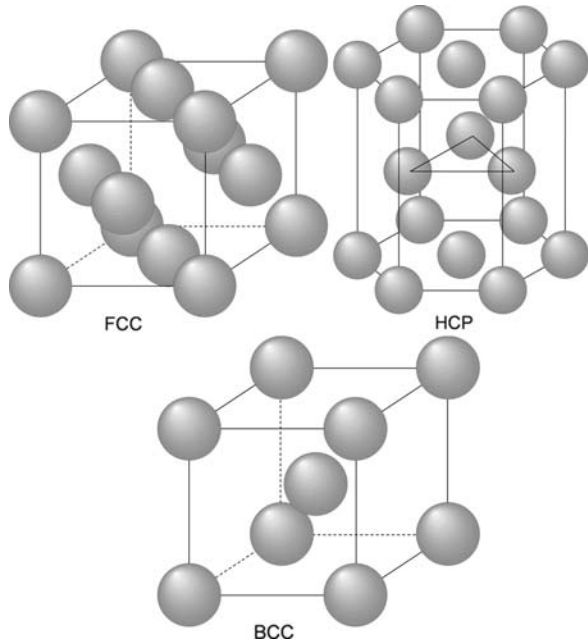
below about 900 °C, Co base alloys form fcc crystalline structures above approximately 850 °C, 316L stainless steel is fcc from its forging temperature ~1050 °C down to room temperature), others form less closely-packed structures (Ti and Ti-based alloys form bcc structures at elevated temperatures). Lowest free energy determines the most stable crystal structure under given conditions. Understanding the nature of the transformations that may occur during metal processing is important for control of properties. Phase transformations can be beneficial for achieving desired properties, but may also result in secondary phases with undesirable properties leading to unacceptable material performance. A good understanding of constitutional (equilibrium) phase diagrams is important for the design of processing methods for metallic implant formation. It should be appreciated, however, that these often over-simplified diagrams (i.e., limited to two- or three-element alloys rather than the multi-elemental compositions of most practical alloys) indicate, even for these simple compositions, equilibrium structures that may not be achieved during processing because of kinetic considerations as discussed below.

2.4 Phase Transformations – Diffusive and Displacive

The equilibrium arrangement of atoms in solids, as noted above, is determined by factors such as atomic size, valence, and chemical affinity between elements under specific extrinsic conditions (i.e., temperature, pressure). For given temperature and pressure conditions, the stable state is the structure of lowest free energy. Free
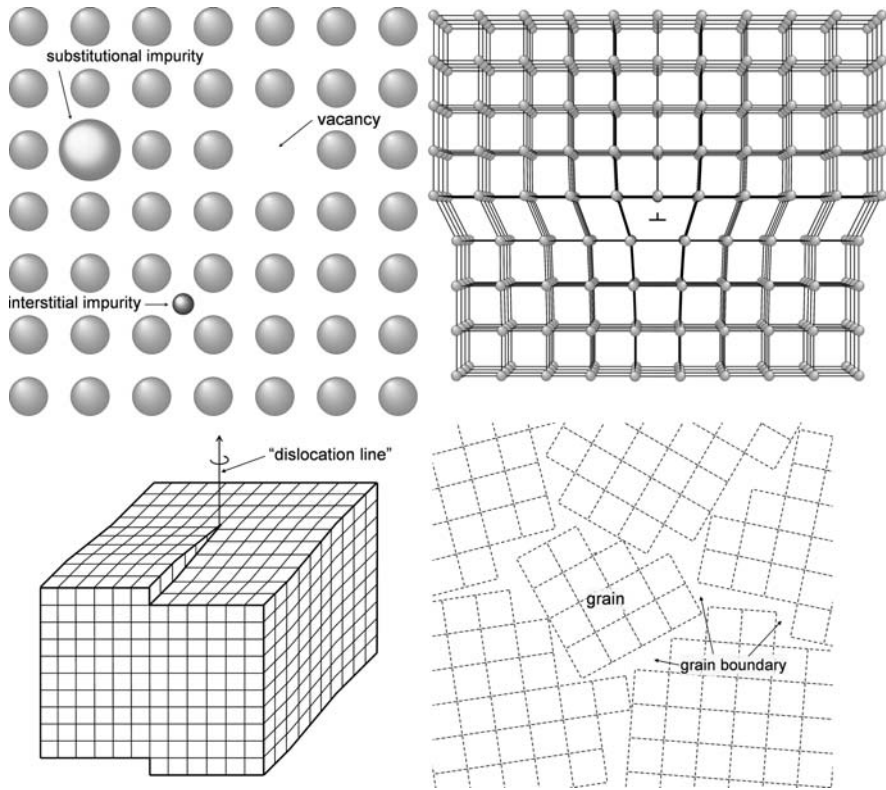
Fig. 2.2 Crystal structures showing point defects (substitutional or interstitial elements, vacancies), line defects (edge and screw dislocations), planar defects (grain boundaries) (Courtesy of Scott Ramsay, Adjunct Professor, University of Toronto)

energy, G, is related to enthalpy, H, (heat of formation), and entropy, S, through the relation $G = H - TS$, where T is the absolute temperature. Changes in state or phase transformations in solids are possible if a decrease in G occurs due to the transformation (i.e., $\Delta G = \Delta H - T\Delta S < 0$ at the transformation temperature).

While a reduction in G represents a necessary thermodynamic condition for the formation of a new phase, atomic rearrangement to form the new phase must occur in a reasonable time period (a kinetic consideration). For most solid-state phase transformations, atom rearrangement occurs through thermally controlled diffusion processes (diffusive transformations) that result in both crystallographic and regional compositional changes within the newly formed phase(s), the amount and microstructure of the new phase(s) being dependent on temperature and time at temperature. Phase transformations can also occur through diffusionless processes that are not dependent on time at temperature but rather on temperature only or mechanical straining of samples at an appropriate temperature. These are referred to as displacive or *martensitic* transformations [4]. They are *athermal* meaning that the
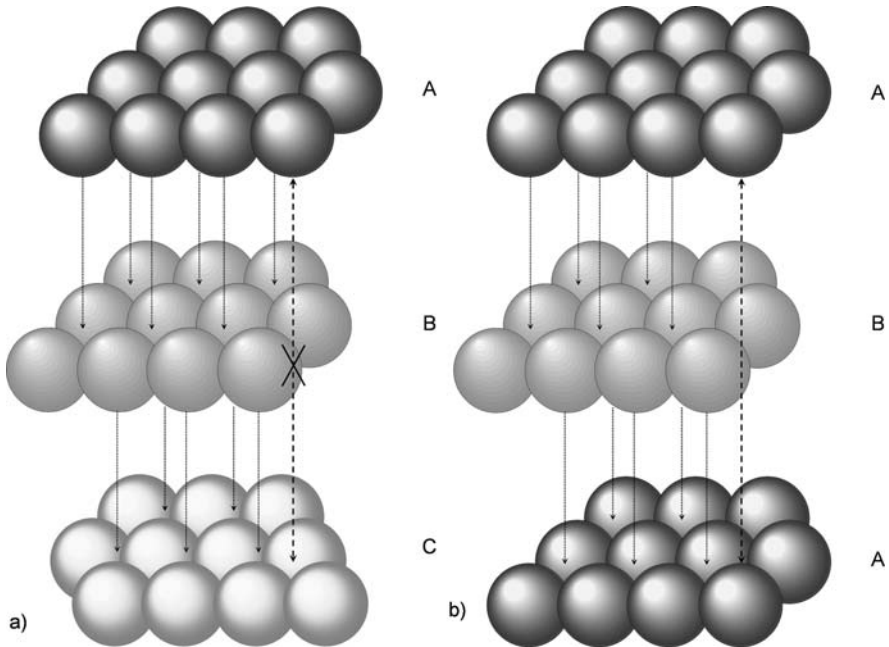
Fig. 2.3 Solid sphere models showing the two stacking sequences possible in close-packed crystal structures. The layering sequence of the close-packed planes is different but the nearest number of neighboring atoms (coordination number) is 12 for both arrangements (Courtesy of Scott Ramsay, Adjunct Professor, University of Toronto)

amount of new phase forming from a parent phase is dependent only on the final temperature (or degree of deformation at temperature). The transformation occurs very rapidly with the transforming wave-front sweeping across the sample at a rate corresponding to the speed of sound. Very slight atom displacements are needed to achieve the transformation. The temperature at which martensite first starts to form from the parent phase (austenite) is the M_s temperature (martensite start temperature) with 100% martensite forming at the M_f temperature (martensite finish temperature). Upon heating, the reverse transformation (martensite to austenite) starts at the austenite start (A_s) and finishes at the austenite finish (A_f) temperatures. There is hysteresis between the forward and reverse transformation temperatures reflecting the energy required for the transformation. Displacive transformations involve very slight repositioning of atoms within a crystalline solid in combination with a so-called *lattice invariant* strain that involves lattice *twinning* only in order to accommodate the overall shape change that would otherwise result from the cumulative small atom displacements. Some metallic biomaterials can transform through a displacive mechanism including the CoCrMo and Ti-based alloys that are used extensively for joint replacement implants. The Ni-Ti shape memory alloys (SMA) also transform by a displacive transformation resulting in two related phenomena that

are attracting considerable interest in biomaterials currently (i.e. the shape memory effect and superelasticity).

For Ni-Ti alloys, the net result of the austenite-to-martensite transformation is a highly twinned martensitic structure. The resulting twin boundaries are low energy boundaries that are quite mobile. Thus, components can be readily deformed at lower temperatures (where the martensite phase is stable) resulting in a change in part shape. On annealing above the A_f temperature, the martensite is transformed back to austenite *with an accompanying reversion to the initial undeformed shape.* This is referred to as the shape memory effect and can be used to advantage in biomaterial applications to impart a mechanical force on tissue following implantation as the implanted device warms up to body temperature. This is described further in the section on *Ni-Ti Alloys*.

A second effect due to the martensitic transformation in Ni-Ti alloys is that of *superelasticity* (also referred to as a *pseudoelasticity*). This effect results in a very low apparent elastic modulus for the Ni-Ti alloys. It is due to the formation of *stress-induced martensite (SIM)* during mechanical deformation of the austenite phase within a limited range of temperatures just above the A_s temperature and below a temperature defined as the M_d temperature, (the maximum temperature at which stress-induced martensite will form). Mechanical strain at this temperature provides the driving force for austenite-to-martensite transformation. However, the resulting stress-induced twinned structure forms with a preferred twin *variant* (a twin formed with a twin boundary defined by a specific crystallographic plane, e.g., a (111) plane) in contrast to the formation of a family of twins having common boundary planes but differently oriented, for example having {111} boundary planes as occurs for the thermally-driven shape memory transformation. The stress-induced transformation results in a significant overall elongation in contrast to the thermally-driven one. The elongation is fully recovered on release of stress. The resulting stress-strain curve (Fig. 2.4) shows complete recovery upon unloading to relatively large strains (as high as 10% for some shape memory alloys) thereby resulting in a low apparent elastic modulus.
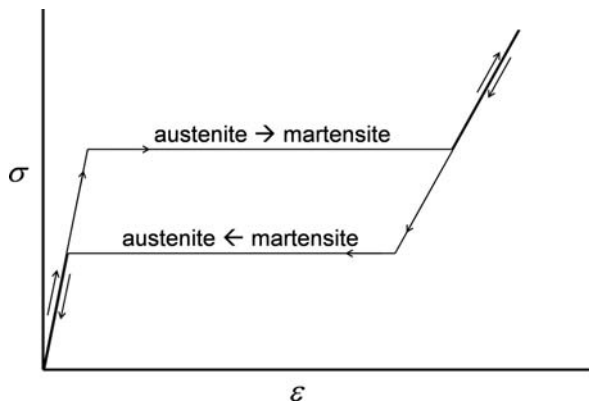
Fig. 2.4 Stress-strain curve for a metal (such as Ni-Ti alloy) displaying stress-induced martensitic transformation and superelastic behavior. Note the hysteresis on loading and unloading

2.5 Diffusion in Metals

Atomic diffusion occurs during metal processing and determines the structure and properties of metallic components. Hence, solidification during casting of a melt, grain growth during elevated temperature annealing of parts, the precipitation and growth of second phase particles, sintering of metal powders to form either dense or porous structures, bonding of films or coatings to substrates, recrystallization (in which relatively strain-free crystals nucleate and grow within mechanically-deformed materials), and the formation of protective oxides over metal substrates (passive film growth) all involve atom diffusion. Thus, an understanding of atomic diffusion rates and mechanisms is important for design of processes for formation of components with desired properties. The diffusion rate, D, is given by the equation $D = D_o \exp(-Q/RT)$ where D_o = diffusion coefficient (characteristic of an atom diffusing in a given material), Q = activation energy for atom movement, R = universal gas constant and T = absolute temperature. This relation shows that diffusion rates fall exponentially as temperature is reduced so that at some lower temperature, diffusion driven structural changes will not occur in practical time periods even though they are energetically favored by changes in free energy (i.e., $\Delta G < 0$). As a result, metastable structures can exist at ambient temperatures. The retention at room temperature of a fcc-structured phase in Co-based alloys used for orthopedic implant fabrication is an example of this. The fcc structure (γ-phase) is metastable at temperatures below about 900 °C where free energy considerations would predict the formation of the hcp ε-phase. Slow diffusion of Co, Cr, and Mo atoms at and below the transformation temperature prevents the ready formation of the hcp phase by a diffusive transformation. The reaction is "sluggish" and, only under special circumstances as discussed later, does the stable hcp phase form.

The formation of well-developed crystalline structures during solidification from the liquid state (or during other types of solid state formation reactions such as chemical or physical vapor deposition processes that are being used for formation of films and coatings) requires the movement of atoms to equilibrium lattice sites. Should very rapid cooling occur during solidification, such regular positioning of atoms is unable to occur and, instead, random atomic positioning and amorphous structures lacking long-range order will develop. Such amorphous structures occur more readily during solidification of complex ionically- or covalently-bonded non-metals. These have relatively lower diffusion rates due to the large molecules that define the structures of these materials and the higher activation energies for molecule movement. Some metals can also form amorphous structures on solidification but only if very high cooling rates ($>10^6$ °C/s) are used during solidification [5]. These display some interesting properties including high hardness and excellent corrosion resistance, characteristics that are desirable for certain biomaterial applications. The high hardness can be explained by the fact that amorphous metals do not deform readily by dislocation glide as occurs with crystalline metals (see below). The excellent corrosion resistance appears related to the fact that preferred sites for initiation of corrosion reactions such as grain boundaries, dislocations, and other lattice defects (vacancies) do not exist in these structures. These as well as

nano-crystalline metals (grain size <100 nm or so) that also display interesting properties such as excellent wear resistance, high strength and corrosion resistance are discussed briefly in the *New Directions* section. These represent exciting new directions for potentially forming novel metallic implants or implant surface coatings.

2.6 Interatomic Forces and Elastic Moduli (*Structure-Insensitive* Properties)

All crystalline materials, as noted above, are formed with characteristic 3-D equilibrium atomic structures with specific atom arrangement/packing and equilibrium interatomic spacings that characterize that material. The interatomic forces acting on atoms in these equilibrium positions will be equal to zero as a result of a balance of attractive and repulsive force fields acting on the atoms (i.e., F = dU/dr where F = interatomic force, U = bond energy, a function of 'r', and r = interatomic spacing – Fig. 2.5a,b). Small displacements from the equilibrium position in response to applied forces result in an increase in internal energy of the structure. The potential energy of an atom, U, can be described by an empirical equation of the form
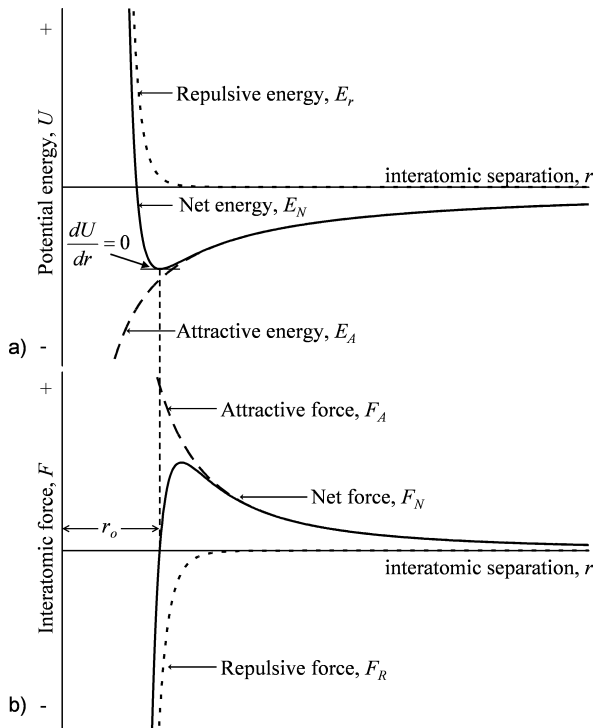
Fig. 2.5 Potential energy vs interatomic separation curve (**a**) showing how the net effect of attractive and repulsive forces acting between atoms, as shown in (**b**), result in an potential energy minimum that defines the equilibrium interatomic separation

$U = -A/r^m + B/r^n$ where m<n [1]. The first term of this equation is related to an attractive energy force (this force is lowered as atoms are brought closer together approaching the equilibrium spacing, r_o) while the second term is related to the repulsive force that develops as atoms are brought into too close proximity ($r < r_o$). This repulsive energy term predominates at $r < r_o$. The interatomic force-distance relation represents an intrinsic material characteristic relating small lattice deformations (atomic displacements) and the forces resisting these displacements. Stress and strain can be determined experimentally for a material with the initial slope of the stress vs strain curve describing an intrinsic elastic constant for a material (Young's modulus, E, for very small tensile or compressive forces or Shear modulus, G, for shear forces). For small deformations in metals, the relation between stress and strain is linear and, except at higher temperatures approaching the melting temperature where atomic diffusion is rapid and creep deformation occurs, is virtually strain-rate independent for metals. Assuming a regular 3-D cubic arrangement of atoms and an applied tensile force (one tending to increase the separation between atoms), a simple estimate of E can be made [2] giving Young's Modulus, $E = \sigma/\varepsilon = dF/r_o^2/dr/r_o = (1/r_o)(dF/dr)$ where σ = stress, ε = strain, F = interatomic force, r and r_o = interatomic and equilibrium interatomic spacing respectively. This suggests that E is directly related to the slope of the interatomic force-distance curve and inversely proportional to the equilibrium spacing between atoms. Metal processing has very little influence on these two parameters so that elastic constants (E and G) can be considered intrinsic material properties independent of method of processing and the resulting microstructural effects (i.e., *structure-insensitive* properties). A third elastic constant, Poisson's ratio, υ, that describes the lateral deformation occurring as a result of axial elastic deformation ($\upsilon = -\varepsilon_y/\varepsilon_x$ for a force applied in the 'x' direction) represents another *structure-insensitive* intrinsic material property. Table 2.1 summarizes some mechanical properties for a number of commonly used metallic biomaterials indicating the definite value for elastic constants in contrast

Table 2.1 Mechanical properties of metallic biomaterials. Small variations in E may be attributable to different measuring methods. The large range of strength and % elongation to failure properties are due to different material processing. Some polymer and ceramic, as well as cortical bone properties are shown for comparison

Material	E (GPa)	σ_{yield}(MPa)	σ_{ult}(MPa)	% elong
Fe-based	200–205	170–690	540–1000	12–40
Co-based	220–230	450–1500	655–1900	5–30
CP Ti	100–115	170–480	240–550	15–24
Ti-based	100–110	585–1050	690–1150	10–15
Ta	188	140–345	205–480	1–30
Ni-Ti (Ms)	28–41	70–140	895	~9
UHMWPE	0.5	–	~3	800
Al_2O3	350–380	–	400 (flexural)	–
$PS-ZrO_2$	200	–	800 (flexural)	–
Bone (cortical)	10–20	–	100–300	1–2

to the range of possible values for other mechanical properties. The range of these other properties is due to their dependence on material microstructure that, in turn, is a function of the method of processing used for material formation. Thus, in contrast to the elastic constants, these other properties are *structure-sensitive*. Also included in Table 2.1 are the elastic constants for some other biomaterials (polymers and ceramics) as well as bone (cortical and cancellous). The large difference in elastic constants for the metals commonly used for forming bone-interfacing implants for orthopedic applications (stainless steel, Co-based alloys, Ti and its alloys) and bone is noteworthy. This large disparity can lead to undesirable structural changes in bone situated next to a metallic implant in situations where the implants are (1) securely fastened to bone (usually a preferred situation for joint replacements and fracture stabilization) and (2) oriented with the length of the implant juxtaposing a significant length of bone (as for example a femoral hip implant stem component against the host femur). The resulting composite construct (metal implant and bone) forms, in effect, a reinforced composite so that major stresses are borne by the higher modulus metallic implant thereby 'stress shielding' the host bone. Because of this, disuse atrophy and significant bone loss can result over time. In addition, abnormally high stresses can develop in bone at regions where force is transferred from the implant to the host bone (e.g., at the distal tip region of the femoral stem – Fig. 2.6). Studies to develop lower elastic modulus metallic alloys to overcome this 'stress shielding' effect represent an active area of current biomaterials research.
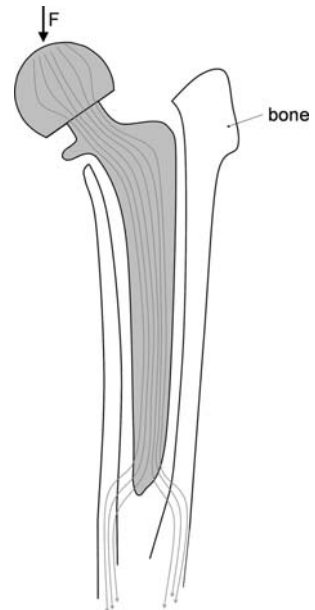
Fig. 2.6 Schematic illustration of a securely fixed metallic femoral hip implant component in the femur showing the direction and concentration of lines of force acting during loading. The stiffer implant component acts to shield the bone next to the proximal stem portion thereby promoting bone disuse atrophy (a stress shielding effect). Note also the concentration of forces resulting in stress concentration within bone at the distal stem region (Courtesy of Scott Ramsay, Adjunct Professor, University of Toronto)

2.7 Plastic Deformation and *Structure-Sensitive* Properties

Metals display high degrees of ductility if appropriately processed, a direct conse-
quence, as already noted, of the metallic interatomic bonding that allows relatively
easy movement of crystal line defects (dislocations) along certain crystallographic
planes (slip planes). Dislocations (edge and screw) are a type of crystal lattice defect
(line defects) that occur within crystalline solids (Fig. 2.2). Their movement through
the crystal lattice allows plastic deformation or yielding of metals at stress levels
far below those predicted for shearing of crystal planes in idealized dislocation-free
crystals. Plastic flow results in blunting of sharp geometric discontinuities that act as
stress concentrators at which crack initiation will occur. The force required to cause
dislocation movement or 'gliding' along a crystal plane is related to the strength
of interatomic bonding, since dislocation movement requires sequential breakage
and reformation of interatomic bonds, as well as dislocation interactions with other
lattice defects (other dislocations, grain boundaries, vacancies, impurity or alloying
elements). Significant dislocation movement and plastic deformation occur when
stresses are above a material's yield stress (Fig. 2.7). To increase yield strength, a
number of strategies can be used all aimed at increasing a metal's resistance to dis-
location glide. These include *strain hardening* (also referred to as work hardening or
cold working), alloying for substitutional or interstitial solid *solution strengthening*,
precipitation or second phase hardening, and strengthening by *grain refinement*.

Strain hardening is achieved by mechanical working of a metal above its yield
stress. It is a result of dislocations interacting with other dislocations during load-
ing resulting in dislocation entanglements and pile-ups that inhibit further dislo-
cation glide thereby resulting in higher yield stress. *Solid solution strengthening*
occurs through addition of other elements to a pure metal (or through inadvertent
impurity element contamination). This results in increased resistance to dislocation
movement along slip planes. *Precipitation hardening or strengthening* is a result
of the formation of second phase precipitates that interfere with easy dislocation
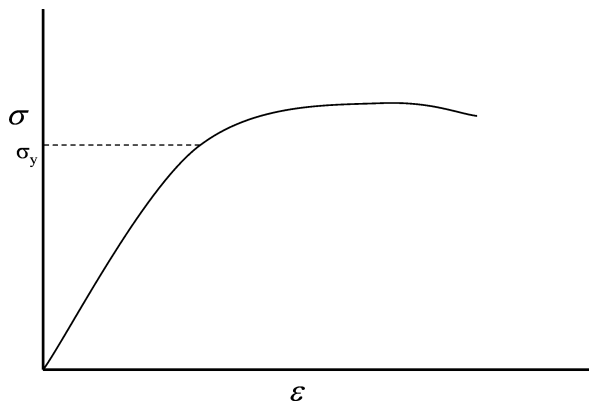
Fig. 2.7 Stress-strain curve for a metal showing the yield stress, σ_y. The yield stress is normally defined as the stress corresponding to either 0.1% or 0.2% plastic strain ($\sigma_{0.1\%}$ or $\sigma_{0.2\%}$)

glide. (*Dispersion strengthening* is similar but usually refers to non-metallic dispersoids within the metal matrix.) *Grain boundaries* and *interphase boundaries* also inhibit dislocation movement since glide directions along preferred slip planes must change at such junctions. Finer grain size (and consequently more grain boundary area) results in higher yield strength as described by the Hall-Petch relation ($\sigma_y = \sigma_o + k/d^{1/2}$ where d = mean grain size and σ_o and k are empirically determined constants [3]. This relation breaks down for metals with very small nano-sized or very large grains. All these possible strengthening mechanisms involve dislocation interactions with other dislocations or microstructural features. These result from processing routes used for forming implants for highly loaded applications such as those found in orthopedics and dentistry.

In addition to achieving sufficiently high yield strength to prevent unacceptable shape changes during functional loading, outright implant fracture must be avoided. Fracture can occur either as a result of a single overload event or through repeated (cyclic) loading at stresses well below the ultimate stress and even below the yield stress in the case of fatigue failure. Fatigue failure is a process resulting from the initiation and slow propagation of cracks leading to eventual catastrophic failure. This can occur in a relatively low number of cycles ($\leq 10^4$ cycles) (high repeated strains leading to low cycle fatigue, LCF) or over millions of cycles (low cyclic strain resulting in high cycle fatigue, HCF). For the former, relatively few loading cycles are required to initiate cracks so that crack propagation rates primarily determine fatigue strength. For HCF, the number of cycles for crack initiation can represent more than 95% of the total lifetime (measured in cycles to failure). Thus, prevention of fatigue crack initiation is of crucial importance for avoiding HCF failure. Fatigue crack initiation is a result of dislocation interactions leading to formation of microvoids and dislocation run-out creating surface irregularities (slip bands) that act as stress concentrators promoting local crack initiation. Thus, processing to increase resistance to dislocation movement is a strategy for achieving higher yield strength and, hence, higher fatigue strength although yield strength increases are accompanied by reduced ductility so that an optimal degree of strengthening will normally result in the highest fatigue strength. While fatigue fractures of contemporary orthopedic and dental implants are uncommon, they do occasionally occur and often after years of apparent satisfactory implant performance and millions of cycles of loading suggesting failure due to HCF conditions. This can be confirmed by examination of fracture surfaces using high resolution scanning electron microscopy (typically at magnifications >1000 times). These studies often reveal the tell-tale signs of HCF, namely very finely-spaced fatigue striations emanating from a fatigue crack initiation site (Fig. 2.8). In general, increased fatigue strengths are related to yield strength so that the processes used to achieve increased yield strengths normally benefit fatigue strength. However, as noted above, strain hardening also results in decreased ductility which can lead to easier fatigue crack initiation so that careful selection of material processing procedures is necessary to achieve optimal static and dynamic (fatigue) properties of metallic components.

Fracture toughness (K_c) or fracture energy (G_c) represents another important material property for consideration in device design. It is a measure of a material's

Fig. 2.8 Scanning electron micrograph of a CP Ti
dental implant that had failed as a result of fatigue.
The fatigue failure is clearly identified from the
striations observed on the fracture surface. Note the
high magnification needed to view the striations that
have an approximate 0.5 μm spacing
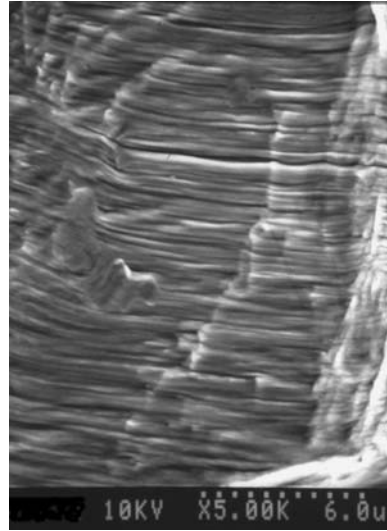
damage tolerance characteristic. The ability of metals to deform plastically has a
strong effect on fracture resistance since plastic deformation can result in blunt-
ing of sharp flaws thereby reducing local stress concentrations. This is the rea-
son for the much higher fracture toughness values of metallic biomaterials com-
pared to other high strength bioceramic materials such as Al_2O_3 or phase-stabilized
ZrO_2 that are characterized by high compressive and much lower tensile strengths,
(see Table 2.2).

Yield strength, ultimate strength, ductility, fatigue strength, fracture toughness,
hardness and wear resistance are examples of *structure-sensitive* mechanical prop-
erties. Lattice defects or imperfections such as point defects (vacancies, intersti-
tials) and planar defects (grain or crystal boundaries, twin boundaries, interphase
boundaries) and the interaction of these with dislocations determine these *structure-
sensitive* properties. How materials are processed determines the density of and

Table 2.2 Metals used for orthopedic implant applications

Metal	Major/(minor) application	Processing route
Stainless steel	Osteosynthesis/(joint arthroplasty)	Hot/warm forming, machining
CoCrMo alloys	Joint arthroplasty/(osteosynthesis)	Casting, hot/warm forming, p/m
Co-Ni alloys	Osteosynthesis/(joint arthroplasty)	Hot/warm forming
CP Ti	Osteosynthesis	Hot/warm forming, machining
$(\alpha + \beta)$ Ti alloys	Joint arthroplasty and osteosynthesis	Hot/warm forming, machining
β/near-β) Ti alloys	Osteosynthesis	Hot/warm forming, machining
Ni-Ti	Osteosynthesis	Hot/warm forming, machining
Ta	Bone augmentation	Chemical vapor infiltration

interactions between these microstructural features and consequently the acceptable limits of use of components made from the materials.

2.8 Corrosion Resistance

Concerns over possible high rates of corrosion of metallic biomaterials and the detrimental effect that this can have on biocompatibility represent the most important consideration in the selection of metals for implant use [6]. In fact, only relatively few metals are considered acceptable for surgical and dental implant fabrication (Table 2.1). This is based on the relatively low rates of in vivo corrosion of these metals if properly processed. Acceptable corrosion resistance for most of these relies on their ability to form well-adhered, dense, protective oxide surface layers (passive oxide films typically 5–10 nm thick) that are retained during in vivo use. Thus, Ti and its alloys, Ta, CoCrMo and Co-Ni alloys, Ti-Ni alloys and certain austenitic stainless steels rely on such passive oxide layers for corrosion protection. In addition, some metals made from more noble elements that do not rely on passive film formation and that display acceptable in vivo corrosion properties are used for some surgical implant and dental device applications. Pt and Pt-Ir components are examples used for fabricating neuromuscular stimulation electrodes, cardiac pacemakers being the most common example. Au alloys and Pd alloys are used for dental bridge construction and as components of dental crowns (porcelain-fused-to-metal crowns). These noble metals are able to resist corrosion to an acceptable degree even in the aggressive body or oral environment due to their inherent chemical stability. Those metals that rely on passive oxide film protection, in contrast, are extremely reactive in oxygen-containing environments. This characteristic is used to advantage, for forming the well-adhering, dense oxide layers that develop either spontaneously during use or through chemical (immersion in nitric acid solution, for example), electrochemical (anodic film formation), or thermal (air oxidation) treatment as the final step in manufacture of implants from these metals. The important characteristic of these passive oxide layers is their relative stability in vivo, providing effective barriers to electron and ion transport. And while film growth and ion release can continue in vivo, the rate is sufficiently low to allow the safe use of these alloys.

2.9 Metals and Processes for Implant Fabrication

As already noted, only a relatively small number of metals are used currently for making surgical implants, primarily because of concerns over metal corrosion and biocompatibility. The metals used for making orthopedic implants (joint replacements and implants used in fracture repair or spinal fusion procedures) are listed in Table 2.2. These represent major load-bearing applications requiring the use of materials with sufficient corrosion resistance, strength and fracture resistance

for long-term (joint replacements) or shorter term (fracture fixation) applications. Implant loading conditions can be complex and occur in an aggressive body environment so that good corrosion-fatigue resistance is a major requirement. (Corrosion-fatigue describes the additive effect of a corrosive environment on fatigue crack initiation and propagation.) Joint replacement implants must also display good wear resistance since wear debris generated at articulating joint surfaces can either of itself or following its degradation through corrosion result in unacceptable host reactions leading to implant loosening. Ideally, joint replacement prostheses should function reliably for decades. Unfortunately, even with today's biomaterials and implant designs, this is not always the case. For implants used in fracture repair procedures the requirements are less stringent since these usually can be removed following fracture healing (i.e., after months or so) although this is not always done in practice.

The commonly used metals and some newer alloys that are being investigated currently are reviewed below. Further information can be found in other review articles on this subject [7, 8]. The focus of the present chapter is on the processing of the metallic biomaterials and how processing determines properties (i.e., fatigue strength, wear resistance, corrosion properties). The basic mechanisms outlined above are used to rationalize the process-property relations and selection of methods for implant manufacture.

2.10 Austenitic Stainless Steel (ASTM F 138/139, F 1314, F 1586, F 2229) [1]

The stainless steel compositions (all fcc austenitic stainless steels) recommended for implant use are listed in Tables 2.3 and 2.4.

316L austenitic stainless steel (ASTM F 138/139), despite its greater susceptibility to crevice corrosion compared to other common metallic biomaterials, has over decades of use proved acceptable and the metal of choice for fracture repair devices. It has also been used for making some joint replacement components (although its use for forming these has become less common). Stainless steel's corrosion resistance is dependent on the formation of a thin Cr, Mo-containing passive surface oxide layer, the Mo imparting stability in a Cl^--containing environment. It forms a single phase (fcc austenite phase) from its forging temperature ($\sim$1050 °C) to room temperature and achieves its reasonable strength and fatigue resistance through strain hardening and solid solution strengthening mechanisms and a fine grain size. Typically, implants are forged at temperatures starting at 1050 °C and continuing through a series of forging and re-annealing steps until a near-final shape is achieved. Forging at the elevated temperature facilitates shaping since dislocation entanglements and pile-ups that would result in strain hardening at lower temperatures are eliminated. This is a result of microstructural *recovery* and

---

[1]American Society for Testing and Materials – recommended standards

Table 2.3 Compositions (in wt%) of austenitic stainless steels used for orthopedic implant fabrication [7]

ASTM#	C	Mn	P	S	Si	Cr	Ni	Mo	N	Cu	Other
F 138	0.03	2.0	0.025	0.01	0.75	17.0–19.0	13.0–15.0	2.25–3.00	0.10	0.50	
F 1314	0.03	4.0–6.0	0.025	0.01	0.75	20.5–23.5	11.5–13.5	2.00–3.00	0.20–0.40	0.50	0.1–0.3Nb 0.1–0.3 V
F 1586	0.08	2–4.25	0.025	0.01	0.75	19.5–22.0	9.0–11.0	2.00–3.00	0.25–0.50	0.25	0.25–0.8Nb
F 2229	0.08	21–24	0.03	0.01	0.75	19.0–23.0	0.10	0.50–1.50	0.90 (min)	0.25	

Table 2.4 Mechanical properties of austenitic stainless steels used for orthopedic implant fabrication; higher and lower values correspond to cold-worked and fully annealed samples respectively [7]

ASTM #	σ_{yield}(MPa)	σ_{ult}(MPa)	% elong	$\sigma_{fatigue}$ (10^7)
F 138	190–690	490–1350	<12–40	190–700
F 1314	380–860	690–1035	12–35	–
F 1586	430–1000	740–1100	10–35	–
F 2229	590–1550	930–1730	12–50	–

recrystallization, processes that occur during the elevated temperature mechanical working. (*Recovery* involves dislocation rearrangement to form lower energy networks while *recrystallization* involves the nucleation of new strain-free crystals with low dislocation density within a mechanically-worked microstructure.) Thus, easy plastic deformation and shaping (because of the low dislocation density) without the danger of component cracking are possible using elevated temperature forging. As parts cool during forging, they eventually reach the characteristic recrystallization temperature (the temperature above which nucleation of strain-free crystals occurs, ~1000 °C for stainless steel) so that continued forging on cooling below this temperature results in strain hardening of the alloy. This lower temperature 'finish forming' is useful for achieving a desired yield strength for the final component. The final properties (yield, ultimate strength and ductility and most importantly, fatigue strength) are dependent on the degree of 'cold' working that is imparted during this lower temperature mechanical deformation. Achieving a fine, uniform grain size is also important (for 316L an ASTM No. 5 grain size is recommended; i.e., ~60–65 μm cross-sectional dimension). Higher nitrogen content in some stainless steels (ASTM F 1314, ASTM F 1586, ASTM F 2229) results in higher strength (due to greater solid solution strengthening) as well as improved crevice and pitting corrosion resistance. The latter composition (ASTM F2229) is a Ni-free, nitrogen-strengthened austenitic stainless steel with a higher Mn content (a fcc stabilizing element like Ni) resulting in the retention of fcc austenite despite the lack of austenite-stabilizing Ni. As with the original 316L stainless steel, the corrosion resistance of all these alloys depends on the formation of a passive Cr(+Mo) oxide layer. While all these austenitic stainless steels are nominally single-phase fcc alloys, carbides can form within the structure due to the small amount of C in these alloys if they are exposed to temperatures in the 400–800 °C range for significant periods. In this range, $M_{23}C_6$ grain boundary phases can form (M being Cr primarily) with associated depletion of Cr from adjacent zones. The resulting denudation of Cr from regions adjacent to the carbides results in a structure that is more susceptible to intergranular corrosion because of the resulting less stable passive oxide film (due to reduced Cr content) next to the grain boundary region. This is referred to as alloy 'sensitization' and to minimize the possibility of its occurrence, low carbon levels are recommended for load-bearing austenitic stainless steels (hence the 'L' in 316L with C = 0.03 wt% maximum). An additional requirement for forming these steels

is the use of vacuum melting during solidification. This ensures low non-metallic inclusion levels. Such inclusions would act as stress concentrators thereby reducing mechanical properties (particularly fatigue strength) of components as well as making component fabrication to final shape more difficult. Following final shaping, implants are ground and polished to a desired surface finish and then given a 'passivation treatment' using a recommended procedure (e.g., exposure to a 40% HNO_3 solution or thermal oxidation treatment – ASTM F 86).

Other grades of stainless steel are used to a lesser extent in other surgical implant applications [7]. These include austenitic stainless steels such as types 304 and 316 (C levels up to 0.08 for the latter as opposed to 0.03 maximum for 316L) in wire form for surgical sutures (316L is also used for this application) and for microvascular clips (for treatment of aneurysms), precipitation hardenable stainless steels (17-7 PH) and martensitic stainless steels (type 420 and 431) also for neurosurgical and microvascular clips. Austenitic stainless steels are also used for fabricating vascular stents as well as electrodes, conducting lead wires and pulse generator housings of cardiac pacing systems (304, 316, 316L alloys). The fabrication of vascular stents presents a particular challenge since these devices experience extensive change in cross-sectional dimensions during expansion to their 'working' diameter in vivo. To attain this characteristic, intricate designs that allow such extensive deformation while keeping local strains in the metal within safe limits avoiding the danger of extensive yielding and fracture have been developed. The stents can be formed to these intricate designs by laser machining fine patterns into thin-walled cylindrical tubes with highly polished surfaces [9].

2.11 Co-based Alloys

Co-based alloy implants can be formed by casting or forging, the latter using bar or rod stock made by conventional forming of cast billets (rolling, extrusion), or by hot isostatic pressing of Co alloy powders. Additionally, novel methods for near net-shape formation of parts from metal powders (metal injection molding or MIM [10]) are currently being explored. CoCrMo implant alloys all contain Cr (~26–30 wt%), Mo (5–7 wt%), some Ni (1 wt% maximum in order to minimize concerns related to possible Ni sensitivity), other residual trace elements (Mn, Fe, Si, N), and C (either low-C ~0.05 wt% or high-C ~0.25 wt%) (Table 2.5).

2.12 Cast CoCrMo (ASTM F 75)

CoCrMo alloys cast to a final form (high-C) are made by investment casting procedures (the lost wax process). This involves the simultaneous casting of a number of components (e.g., femoral stem or ball components for hip implants) onto a so-called 'casting tree' and then cutting components from the 'tree' and grinding, welding (if necessary), honing or otherwise surface finishing to achieve the final implant

Table 2.5 Compositions (in wt%) of CoCrMo and Co-Ni alloys used for orthopedic implant fabrication [7]

ASTM #	Cr	Mo	Ni	Fe	C	Si	Mn	W	P	S	Other
F75	27–30	5–7	1.0	0.75	0.35 max	1.0	1.0	0.2	0.02	0.01	0.25 N; 0.3Al; 0.01B
F799 (low-C)	26–30	5–7	1.0	0.75	0.05	1.0	1.0	–	–	–	0.25 N
F799 (high-C)	26–30	5–7	1.0	0.75	0.25	1.0	1.0	–	–	–	0.25 N
F563	18–22	3–4	15–25	4–6	0.05	0.5	1.0	3–4	–	0.01	0.50–3.50Ti
F562 (MP35N)	19–21	9–10.5	33–37	1.0	0.025 max	0.15	0.15	–	0.015	0.01	1.0Ti
F90	19–21	–	9–11	3.0	0.05–0.15	0.40	1.0–2.0	14–16	0.04	0.03	–
F1058 (Elgiloy)	19–21	6–8	14–16	Bal	0.15	1.2	1.0–2.0	–	0.015	0.015	0.10Be; 39.0–41.0Co
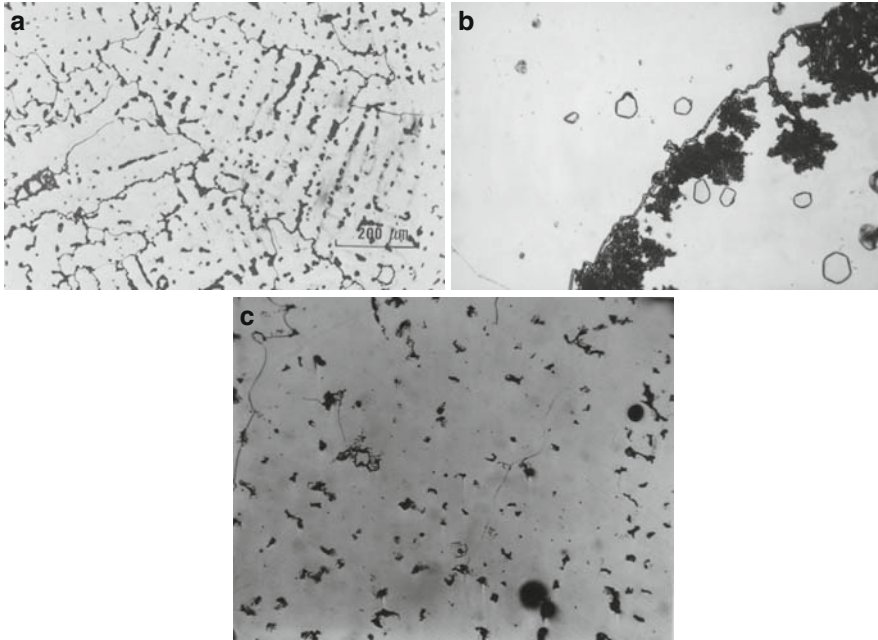
Fig. 2.9 Scanning electron micrographs showing microstructures of (**a**) a polished sample of as-cast high carbon CoCrMo alloy; note the relatively large grains, the heterogeneous composition evident from the different response to chemical etching used to show the microstructure; the darkest regions correspond to carbides dispersed throughout; (**b**) discontinuous $M_{23}C_6$ forming from γ-phase grain boundary region; (**c**) solution annealed CoCrMo alloy (**c**) showing more homogeneous structure but with some retained carbides throughout

form. The high-C alloy represents the most wear resistant clinically used metallic biomaterial, a result of the $M_{23}C_6$ (mainly), M_7C_3, and M_6C carbides (where M is primarily Cr) that form throughout their structure during solidification (Fig. 2.9a). The alloy, after solution treatment (see below), displays a high rate of work hardening [11]. This also may contribute to the alloy's excellent wear resistance as a result of localized hardening of the surrounding Co-based matrix phase during functional loading. The alloy melts between 1350 and 1450 °C depending on its exact composition and forms a typical as-cast, cored structure on solidification with the major phase being face-centered cubic (fcc) austenite designated as either γ- or α-phase in the literature and the interdendritic zones being enriched in Cr, Mo, and C. The γ-phase forms at temperatures above 890 °C while a hexagonal close-packed (hcp) structure is stable below this temperature. However, because of the sluggish nature of the fcc to hcp transformation, a metastable fcc phase is normally retained to room temperature. Subsequent aging at below 890 °C or extensive mechanical deformation temperatures below 890 °C results in the formation of either *faulted* fcc zones (*stacking faults* or regions where the ABCABC stacking sequence to form an fcc structure is altered to ABABAB stacking forming hcp zones within the fcc phase),

or hcp bands (formed by martensitic phase transformation). Aging at between 650 and 850 °C favors $M_{23}C_6$ precipitation in the hcp zones but while this does increase yield strength, it also reduces the material's ductility to unacceptable levels, making such an aging treatment impractical [12]. As already noted, during solidification, a highly cored structure develops with interdendritic Cr-, Mo- and C-rich regions forming extensive carbide networks and other brittle intermetallic phases (σ-phase) that result in low ductility. (Figure 2.10 represents a pseudobinary phase diagram that predicts the formation of the heterogeneous, cored structure during normal cooling from the melt.) To homogenize partially the structure, the as-cast alloy is treated with a short (~1 h) solution anneal at between 1200 and 1225 °C (below 1235 °C since this corresponds to a eutectic melting temperature for the Cr-, Mo-, C-enriched interdendritic regions and heating above this temperature would result in local melting of the Cr-, Mo-, C-rich zones [13] Fig. 2.10). Following the solution anneal, the alloy is cooled rapidly through the 1100–800 °C temperature range to avoid precipitation of embrittling $M_{23}C_6$ carbide networks. These can form as discontinuous carbide precipitates nucleated at and growing out from grain boundaries if sufficient time is allowed (Fig. 2.9b). The 1 h solution anneal treatment results in only partial homogenization of the cored structure but sufficient to give acceptable ductility (11–17% elongation). Some carbide dissolution does occur but enough of the hard carbide regions are retained to provide desired wear resistance (Fig. 2.9c). The major disadvantage of the cast and solution annealed CoCrMo alloy is its
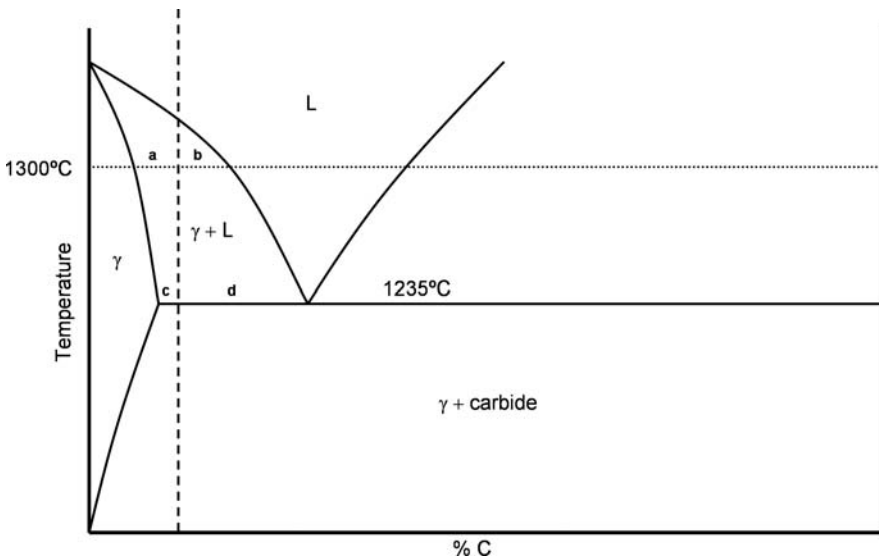
Fig. 2.10 Pseudobinary phase diagram showing a eutectic point at 1235 °C. The *vertical dashed line* corresponds to an alloy with a C-content that on sintering at 1300 °C and rapid cooling would result in a:b relative ratio of carbide to γ-phase. Slow cooling from 1300 °C to below the eutectic temperature would result in this ratio approaching c:d (the ratio at the eutectic temperature assuming equilibrium)

Table 2.6 Mechanical properties of CoCr and Co-Ni alloys after different treatments

Process description	σ_{yield} (MPa)	σ_{ult} (MPa)	% elong	$\sigma_{fatigue}$ (10^7)
CoCrMo alloys				
F 75 – cast + solution annealed	450–530	655–890	11–17	207–310
F 75 – cast + porous-coated	∼490	∼735	∼11	150–207
F 799 – forged (low C)	875–995	1320–1450	19–26	670–800
F 799 – forged(low C) + P-C	∼410	∼815	∼33	–
F 799 – forged (high C)	∼1175	∼1510	∼10	–
F 799 – forged (high C) + P-C	600–840	1030–1280	∼18	∼240
P/M D-S (as-forged)	840	1280	–	690–895
P/M D-S + P-C	–	–	–	345
Co-Ni Alloy (MP35N) – F 562				
Annealed (1050°C)	300	800	40	340
Cold-worked (50% red in area)	650	1000	20	435
Cold-worked + aged	1900	2050	10	405
Other Co Alloys				
F 1058 c-w + aged (wire)	1240–1450	1860–2275	–	–
F 563 c-w + aged	827–1172	1000–1310	12–18	–

relatively low mechanical properties (other than wear resistance) (Table 2.6). This is due to the coarse grain structure (>hundreds of microns and up to the millimeter size range), inherent casting defects and shrinkage porosity resulting upon solidification and relatively slow cooling of the castings. In addition, the carbides that are retained after the solution anneal, while beneficial for wear resistance, create sites for easy crack initiation and, if distributed to form networks along grain boundaries, easy crack propagation pathways. Internal porosity (shrinkage pores) can be eliminated by hot isostatic pressing (HIP), thereby improving properties. HIP treatment has been reported to successfully improve poor quality castings but, not surprisingly, it does not significantly benefit already sound castings. However, this process does not heal surface connected defects so these can still act as preferred sites for early crack initiation. Corrosion resistance, as with the stainless steels, is dependent on the formation of a passive Cr and Mo-containing oxide. This passive oxide layer is normally formed by nitric acid solution treatment as a final step in implant manufacture (ASTM F 86).

2.13 Wrought CoCrMo (Low- and High-Carbon) (ASTM F 799, F 1537)

Warm or hot forging of cast CoCrMo billets can result in significantly higher mechanical properties. Such thermomechanical treatment was applied initially only to low C-containing alloys (C ∼0.05 wt%). The lower carbon content results in fewer and smaller carbides throughout the structure thereby improving the alloy's

formability but at the cost of reducing wear resistance. For hot forging, billets are heated to temperatures between 1000 and 1150 °C. Re-annealing at stages during the forging process is used in order to prevent edge cracking during deformation. Final implant shapes can be achieved using closed-die forging and a series of forging and re-annealing steps. A lower temperature finish forging operation, as with the stainless steel alloys, is used to achieve a degree of strain hardening of the alloy and a final form with desirable mechanical properties. Yield, ultimate and fatigue strengths are significantly higher than for the high-C, cast alloy despite the lower carbide content. This is a result of the much finer grain size, possible stacking fault or hcp band formation, and strain hardening due to the lower temperature working operation all of which contribute to increased resistance to dislocation gliding and, hence, higher yield and fatigue strengths (Table 2.5). As noted in Table 2.6, the fatigue endurance limit for hot-forged, low-C CoCrMo alloy is greater than 600 MPa. The wear resistance of the low-C alloy is poor compared with the higher-C cast alloy thereby negating its use for femoral ball or other surface bearing components.

Closed-die forging of high-carbon CoCrMo alloys is also possible although the procedures for doing this are difficult and very close control on forging and re-annealing stages is necessary. The result is a fine-grained and strain-hardened alloy with the additional benefit of the break-up of larger carbides formed during solidification. This gives a high strength alloy with good wear resistance due to the finely distributed carbides throughout the structure. Whether cast or high-C wrought Co-based alloys provide superior wear resistance remains controversial. The fine carbide distribution does allow improved surface finishing of components and this may contribute to better wear characteristics although supporting evidence based on long term clinical comparisons has not been reported.

Fine-grained, high-C bar stock that is suitable for subsequent closed-die forging to final shape can also be formed using a *powder metallurgy* processing route. By this method, high carbon CoCrMo alloy powders are formed by atomization (either by inert gas atomization or by a rotating electrode process [8]). The powders so formed represent very rapidly solidified 'micro-castings' so that while retaining carbides and a cored structure (like the cast alloy), these are small and finely distributed throughout the atomized powder (Fig. 2.11). After atomization, the powders are sized by screening and then consolidated to full density by placement in a suitable containment vessel that is evacuated, sealed and hot isostatically pressed (at around 1100 °C for 1 h at 105 MPa pressure) or hot forged, to form full density CoCrMo alloy. During the period at elevated temperature, some grain coarsening occurs but it is limited in extent because of the grain growth inhibition effect of the finely distributed carbides. A dispersion strengthened powder-made alloy formed by adding La and Al to the melt stock prior to atomization resulting in additional finely-dispersed La and Al oxide particles throughout the HIP'ed alloy has been reported [14]. The rationale for introducing these finely-dispersed oxides in this case is to inhibit grain growth during a subsequent high temperature sinter anneal treatment that is used for forming a sintered porous surface coating on the implants; this process is described further in the *Surface Modification of CoCrMo Implants*

Fig. 2.11 Scanning electron
micrograph of ground and
polished atomized CoCrMo
powder showing the fine
cored microstructure within
the powder a result of the
relatively rapid cooling rate
during atomization
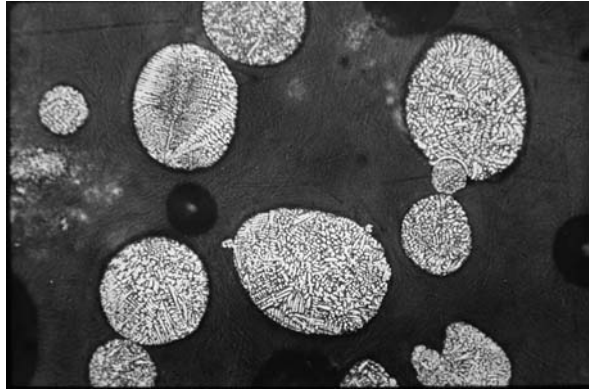
section. The containment vessel outer layer is subsequently removed from the dense
compact yielding a fine-grained CoCrMo alloy containing finely distributed car-
bides. The fine microstructure of the resulting bar allows easier hot or warm forging
to a near final form, (compared with coarser-structured alloys). Implants of rela-
tively high strength, with good honing and polishing characteristics and good wear
resistance can be formed using this metal powder-formed material.

Other novel methods such as metal injection molding (MIM) are being investi-
gated for forming near final shapes from CoCrMo alloy powders [10]. The process
involves mixing fine atomized powders with an organic binder, and extruding the
resulting slurry to form pellets. These are then treated to remove the binder using
a solvent and thermal decomposition during heating to the final sintering tempera-
ture that is just below the alloy's melting temperature (1340–1380 °C). (It is likely
that some liquid phase contributes to sintering during this treatment as a result of
localized melting of Cr-, Mo-, C-rich regions above the 1235 °C eutectic tempera-
ture). The sintered material is then hot isostatically pressed to remove any remaining
porosity and annealed at just above 1200 °C to minimize carbide networks by partial
carbide dissolution. This results in increased ductility.

The strengthening mechanisms for the CoCrMo alloys include solid solution
strengthening, dispersion strengthening (because of the fine $M_{23}C_6$ carbides and
La and Al oxides, if present), carbide phase reinforcement [11], strain hardening
(for the wrought alloys), and dislocation-grain boundary interactions. The coarse-
grained cast structures exhibit ultimate and yield strengths ∼860 and 550 MPa
respectively with fatigue strengths ∼250 MPa for samples in the solution-treated
condition; fatigue strengths of 450 MPa have been reported using casting processes
that result in ultrafine grain size [15]. The finer structure and strain hardening
effects of the wrought alloys result in much higher ultimate tensile strengths (1330–
1450 MPa), yield strengths (960–1000 MPa), and fatigue strengths (10^7 endurance
limit ∼690–830 MPa) [15]. These values are similar to those reported for the
powder-made bar stock materials formed by hot isostatic pressing (HIP) [16].

2.14 Surface Modification of CoCrMo Implants – Porous Coatings for Bone Ingrowth

To achieve secure implant-to-bone fixation without the use of acrylic bone cement, cast CoCrMo implant surfaces can be modified by either sintering CoCrMo powders to the bone-interfacing surfaces or plasma spray coating the surfaces [17, 18]. For the sintered coatings, typically 250–350 μm size powders (–45/+60 mesh) are used to form porous coatings of approximately 50–70% density (i.e., vol.% porosity ∼30–50%). Secure particle-to-particle and particle-to-substrate core bonding is achieved by sintering CoCrMo alloy powders at around 1300 °C for a period of 1 h or so. This high temperature sintering anneal significantly alters the microstructure and mechanical properties of both cast and wrought CoCrMo alloys. As noted previously for cast alloys, localized melting of the Cr-, Mo-, C-rich interdentritic zones occurs at approximately 1235 °C, the eutectic temperature for the solute-enriched interdentritic region [13]. Upon cooling, eutectic phase structures, (γ-phase + $M_{23}C_6$ carbides + possibly some σ-phase), form inter- and intragranularly (Fig. 2.12a,b). These eutectic structures form as networks at sinter neck regions and along the solid substrate grain boundaries where they form pathways for easy crack propagation thereby resulting in reduced ductility and fracture resistance of porous-coated implants (Fig. 2.12). This undesirable effect can be minimized by slow cooling from the normal sintering temperature (∼1300 °C) to below the incipient eutectic melting temperature (1235 °C) [8, 18] thereby minimizing the amount of γ-phase + carbide eutectic formed at the eutectic temperature as predicted by the lever rule (see Fig. 2.10). While early use of sintered porous coatings was limited to cast CoCrMo compositions because of concerns of recrystallization and grain growth during sintering of wrought CoCrMo alloys leading to low strength, the development of high-C and dispersion strengthened wrought Co-based alloys with fine non-metallic dispersoids throughout that inhibit grain growth (i.e., carbides or Al and La oxides) has made it
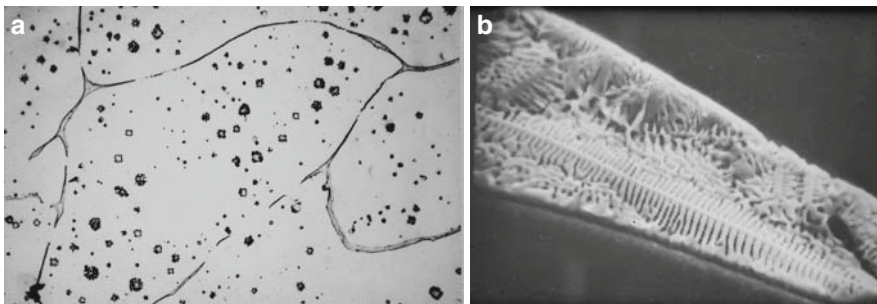
Fig. 2.12 Scanning electron micrographs of polished CoCrMo samples following a 1300 °C 1 h sinter annealing treatment followed by a normal furnace cool to room temperature showing (**a**) grain boundaries with eutectic structure ($M_{23}C_6$ + γ-phase primarily) and (**b**) higher magnification view of the grain boundary microstructure; the lamellar eutectic structure of $M_{23}C_6$ + γ-phase lamellae is clearly visible

practical to porous coat these alloys and retain relatively small grains (60–150 μm size range). Further, the more homogenous compositions of the wrought alloy substrates reduces significantly the amount of $M_{23}C_6$-γ-phase eutectic formation. As a result, higher mechanical properties of wrought high-C Co alloy porous-coated samples can be achieved (fatigue strengths equal to 241 MPa compared to 207 for porous-coated cast samples [14]; La and Al oxide dispersion strengthened porous-coated alloys are reported to have higher fatigue strengths equal to 345 MPa [16]).

2.15 Other Co-containing Implant Alloys (ASTM F 562, F 90, F 563, F 1058)

While the CoCrMo alloys described above represent the most common Co-based alloys used in orthopedics, other Co-containing implant alloys have been and continue to be used [7]. These are described briefly. All these contain alloying elements that significantly affect the nature of the fcc to hcp transformation of these alloys. This in turn has a strong effect on their properties. Cr, Mo, W, and Si are elements that stabilize the hcp phase while C, Ni, Mn, and Fe are fcc phase stabilizers. MP35N (F562 – Tables 2.5 and 2.6) contains 33–37 wt% Ni, an equal amount of Co and somewhat less Cr (19–21 wt%) compared with the F75 alloy. The high Ni content stabilizes the hcp phase so that at temperatures between 425 and 650 °C, a two-phase fcc (γ) + hcp (ε) equilibrium structure can form. However, mechanical working and the associated strain energy is required to nucleate the ε-phase at this lower temperature. The exact thermomechanical treatment used determines the size and distribution of the resulting hcp bands that form and, hence, the degree of strengthening that results due to the interaction of dislocations with the γ–ε interphase boundaries throughout the structure. To maintain acceptable corrosion resistance, MP35N has a higher Mo content than the CoCrMo alloys in order to compensate for its reduced Cr content. The higher Mo also can contribute to further strengthening of the alloy since on annealing in the 425–650 °C range, an intermetallic compound, Co_3Mo precipitates within the hcp zones. As shown in Table 2.5, the MP35N can develop very high fatigue strengths, a feature that made it particularly attractive for forming hip implant components. However, concern over its high Ni content and the reported occurrence of Ni-sensitivity in a significant percentage of the general population has limited its use for joint replacement implants although it remains popular for fabrication of fracture fixation (temporary) implants as well as conducting leads of cardiac pacing systems.

Other Co-based alloys that are used in implant applications include Elgiloy (ASTM F-1058) and the W-containing alloys (ASTM F-563) (Tables 2.5 and 2.6). Elgiloy is renowned for its high springback qualities when highly cold-worked, a property that makes it attractive for fabrication of neurosurgical and vascular implants (neural aneurysm and microvascular clamps) as well as conducting leads for pacemakers. The W-containing CoCrNi alloy (ASTM F-563) has been used for making fracture fixation implants. These alloys are not as corrosion resistant as

some of the Mo-containing Co-based alloys and they are strengthened primarily through work hardening. The common characteristic that all these Co- and CoNi-based alloys share in terms of response to thermomechanical treatment is the sluggish fcc to hcp transformation at relatively low transformation temperatures.

2.16 Titanium-Based Alloys

Titanium and its alloys have been used increasingly for fabrication of orthopedic implants (for fracture fixation and joint replacement) since the late 1960 s. They are also used virtually exclusively for forming endosseous dental implants, another mechanically-loaded application requiring implants with good fatigue resistance characteristics. The increasing use of Ti-based metals, in addition to their good fatigue resistance, is attributed to their excellent in vivo corrosion resistance, a feature related to the stable passive oxide layer (TiO_2) that rapidly forms, their lower elastic moduli compared to other metallic biomaterials (100–110 GPa compared to 200–220 GPa), and their strong osseointegration tendency (i.e., development of close bone-to-implant apposition after short implantation periods). This latter characteristic represents an important advantage for permanent bone-interfacing implants.

Pure Ti is body-centered cubic (β-phase) at temperatures above 883 °C (the β-transus temperature) and hexagonal close-packed (α-phase) at lower temperatures (Fig. 2.13). Addition of most other elements stabilizes either one phase or
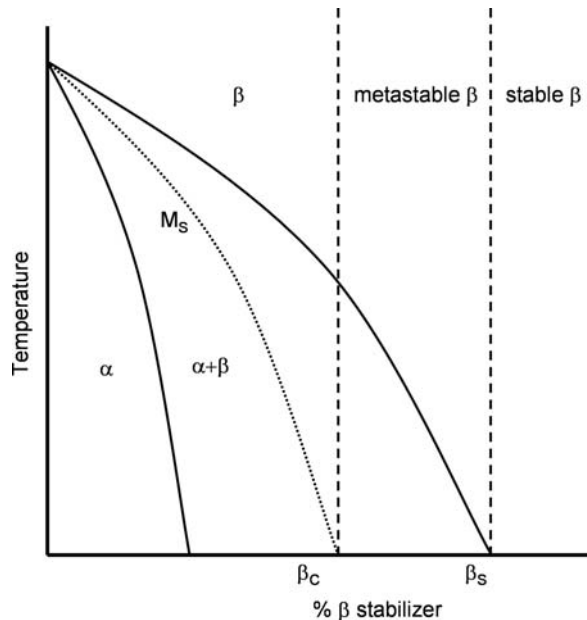
Fig. 2.13 Ti alloy equilibrium pseudobinary phase diagram. With increased β-stabilizer element additions, an (α + β) structure develops, above β$_c$ a metastable β-phase is possible and above β$_s$ only stable β-phase results. The *M$_s$ line* represents temperatures below which martensite can form depending on the cooling rate used

the other. α-Stabilizers include Al, O, N, and C while β-stabilizers are of two types, β-isomorphous (Mo, V, Nb and Ta) and β-eutectoid (Fe, W, Cr, Si, Ni, Co, Mn and H) [19, 20]. β-Isomorphous Ti alloys have attracted interest for implant applications more recently because of the low elastic moduli that are possible with these alloys if appropriately processed. The elements Zr and Sn that are found in some Ti alloys are considered to be 'neutral' alloying elements with no significant effect on either α- or β-phase stabilization.

The major disadvantage of Ti and its alloys is their very poor wear resistance. This makes them unsuitable for load-bearing articulating surfaces without some type of surface modification to give greater wear resistance. This can be achieved either through ion implantation (with N^+) or TiN film application using one of a number of physical vapor deposition (PVD) procedures. With the introduction of modular hip implant designs this has become less of an issue since, with such designs, it is possible to combine a more wear resistant bearing component made of either a Co alloy or a ceramic (Al_2O_3 or Phase Stabilized ZrO_2) with a Ti alloy stem, for example, to form a hip joint replacement with good wear and fatigue resistance. The issue, nevertheless, remains a concern since modular designs of necessity will have intercomponent junctions where relative movement and fretting can occur.

2.17 Commercial Purity Ti

Unalloyed Ti (commercial purity or CP Ti) can contain small amounts of interstitial elements including O, N and H. While the quantities are small (Table 2.7), they affect mechanical properties through interstitial solid solution strengthening [7]. CP Ti is available in four grades, Grade I having the lowest O content and yield strength but highest ductility and Grade IV the highest O content and strength but lowest ductility. The Grade III and IV types are used for fabricating implants for use in osteosynthesis (fracture repair and spinal fusion) but the mechanical strength (low fatigue strength in particular) precludes their use for joint replacement prostheses. CP Ti is used, however, for endosseous dental implants where its characteristic of promoting rapid osseointegration makes it particularly attractive. This is believed to be due to OH^- ion incorporation within the passive TiO_2 layer and reaction of the resulting hydroxylated surface zone with bone mineral phase constituents (Ca^{2+} and $(PO_4)^{3-}$). Strain hardening during mechanical forming of parts, its fine grain

Table 2.7 Interstitial element limits and mechanical properties for CP Ti (Grades 1–4)

Grade	O (max)	N (max)	H (max)	σ_{yield} (MPa)	σult (MPa)	% elong
1	0.18	0.03	0.015	170	240	24
2	0.25	0.03	0.015	275	345	20
3	0.35	0.05	0.015	380	450	18
4	0.40	0.05	0.015	483	550	15

Fig. 2.14 Scanning electron micrographs of CP Ti. The structure is single phase. Grain and twin boundaries can be seen
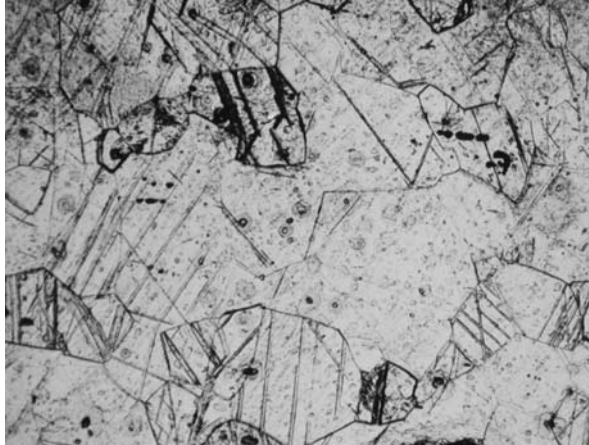
size, and interstitial solid solution strengtheners such as oxygen and nitrogen are responsible for strengthening CP Ti (Fig. 2.14).

2.18 (α + β) Ti Alloys

Alloying of Ti is used to form a two-phase (α + β) alloy of higher strength (yield, ultimate and fatigue) than CP Ti while maintaining excellent corrosion resistance and osseointegration tendency, again because of the TiO_2/OH surface film that rapidly forms.

The (α + β) Ti alloy with the longest history of use for major load-bearing applications is Ti6Al4V alloy with Ti6Al7Nb and Ti5Al2.5Fe being more recent alternatives that are similarly processed giving similar properties. All three alloys behave equally well in clinical use. Bar stock of these alloys is formed by thermo-mechanical processing (mill-annealing) giving high fatigue strength materials (Table 2.8). Combined with their excellent corrosion resistance, implants made from these alloys display superior corrosion-fatigue properties compared to other metallic biomaterials. The mechanical properties of the (α + β) Ti alloys, fatigue strength in particular, are strongly dependent on size and distribution of the α and β phase regions [8]. The so-called mill-annealed treatment (see below) results in the formation of small, equiaxed α grains surrounded by β-phase particles (Fig. 2.15a). This microstructure (developed initially for aircraft/aerospace applications) results in superior fatigue crack initiation resistance and excellent high cycle fatigue strength. The mill annealing process involves mechanically working the alloy to desired form at a temperature just below the β-transus temperature, rapid cooling (water quenching) to room temperature and then annealing to recrystallize the worked structure at around 750 °C (i.e., well within the α + β two-phase field). The rapid quenching treatment results in the formation of α, and a metastable

Table 2.8 Mechanical properties of Ti alloys

Alloy	E(GPa)	σ_{yield}(MPa)	σ_{ult}(MPa)	% elong	$\sigma_{fatigue}$ (10^7)
α–β Alloys					
Ti-6Al-4 V	110	860	930	10–15	610–625
Ti-6Al-7Nb	105	795	860	10	500–600
Ti-5Al-2.5Fe	110	820	900	6	580
β/Near-β alloys					
Ti-12M0-6Zr-2Fe (TMZF)	74–85	1000–1060	1060–1100	18–22	525
Ti-15M0-2.8Nb-0.2Si-0.26 (21SRx)	83	945–987	980–1000	16–18	490
Ti-35.5Nb-7.3Zr-5.7Ta (TNZT)	55–66	793	827	20	265
Ti-13Nb-13Zr	79–84	863–908	973–1037	10–16	500
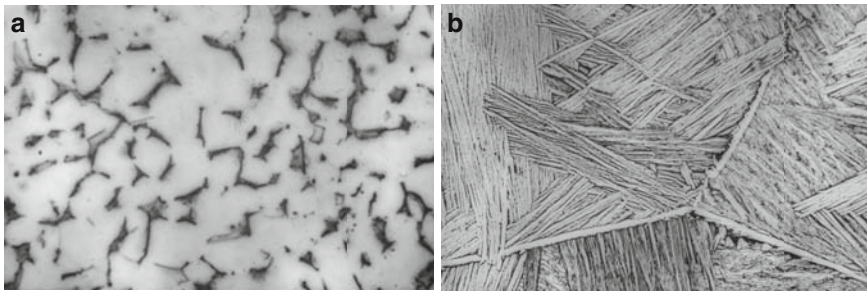
Fig. 2.15 Microstructure of Ti6Al4V alloy; mill-annealed condition (**a**) – *light regions* are α-phase and *darker zones* are β-phase regions; β-annealed condition (**b**) showing the lighter appearing a-lamellae separated by β-phase lamellae. Note also the domain structures (zones within a single grain with α to β formed on different habit planes, i.e., same α to β orientation relation but involving planes in a different crystallographic orientation)

α′-phase as well as the retention of some β-phase; α′ forms by a martensitic transformation mechanism. The 750 °C anneal, in addition to recrystallizing the worked structure, also results in the transformation of α′ to the stable α-phase. The amounts of α and β, as well as the size and distribution of these phases is dependent on the extent of mechanical deformation during component forming at near the transus temperature, the exact forming temperature and the recrystallization treatment. Fatigue strengths greater than 650 MPa (10^7 endurance limit) can be achieved.

The excellent corrosion resistance of Ti6Al4V makes it an attractive choice for forming high surface area, porous-coated or other surface-textured orthopedic implants either by sintering Ti or Ti alloy powders or fibers to the surface of machined substrates or by plasma spray deposition of Ti. Porous coatings formed by sintering the Ti-based materials involve sintering in a high vacuum or other non-oxidizing atmosphere at temperatures above 1250 °C for 1 h or so followed by furnace cooling to room temperature. This so-called β-anneal treatment results in a microstructure consisting of a lamellar (α + β) structure with 'colonies' of lamellae forming within prior β grains (Fig. 2.15b). Because of the higher annealing

temperature within the β-phase field (more than 250 °C above the α–β transus temperature), large β grains form. The 'colonies' that form throughout the structure represent zones in which the α and β lamellae are oriented in different directions but all with a common crystallographic α-to-β orientation relation (i.e., lamellae form on common 'habit' planes in the different colonies). The presence of these colonies and colony boundaries contributes to the strengthening of the alloy. However, the fatigue strength of the β-annealed samples is lower than for mill-annealed samples (i.e., 10^7 fatigue endurance limit of smooth-surfaced β-annealed samples ∼550 MPa compared to 620 MPa for mill-annealed samples [8, 21]. Of far greater consequence to fatigue strength is the effect of the porous coating per se [21]. It results in a reduction of fatigue strength to about a third (i.e., from 600 MPa to approximately 200 MPa). This is due to the notch fatigue sensitivity of hcp metals in general and specifically to Ti and other α-phase-containing Ti alloys. The sinter neck-substrate junctions of the sintered porous layer represent stress concentrators at which fatigue can readily initiate. Design and use of porous-coated Ti6Al4V components must take into consideration this reduced fatigue strength. Some femoral hip stem components are made with porous coatings on all but the lateral stem surface since the highest tensile stress (stresses primarily responsible for fatigue crack initiation) is expected to occur at this surface during normal functional loading of implants. Plasma spray deposition of Ti to form a textured surface suitable for cementless implant fixation is possible without significantly altering the mill-annealed implant core microstructure since the temperature within the alloy bulk remains low. Nevertheless a similar reduction in fatigue strength does occur since the resulting surface is irregular with many sites of surface stress concentration.

2.19 β-Ti and Near β-Ti Alloys

These alloys have higher levels of β-stabilizing elements, Mo displaying one of the strongest effects. A number of β-Ti alloys that have been developed for use in orthopedic implants are listed in Table 2.8. These alloys are characterized by a Mo equivalent >10. This is calculated by using different weighting factors for the elements added in forming the Ti alloy (i.e., Mo equiv = 1.0 (wt% Mo) + 0.67 (wt% V) + 0.44 (wt% W) + 0.28 (wt% Nb) + 0.22 (wt% Ta) + 2.9 (wt% Fe) + 1.6 (wt% Cr) + . . . [22]). While the Young's moduli of Ti and the (α + β) Ti alloys are significantly lower than those of CoCrMo or stainless steel alloys (∼110 compared to 220 and 200 GPa respectively), they nevertheless are 5–10 times greater than the modulus reported for cortical bone (10–20 GPa). The issue of stress shielding of bone next to well-fixed CP Ti or (α + β) Ti alloy implants is a concern in the case of long-term load-bearing implant use (although less so than with the higher modulus Co-based or stainless steel alloys). The β and so-called near-β Ti alloys if appropriately processed exhibit significantly lower elastic moduli (values as low as 44–51 GPa for water-quenched and cold-worked Ti-13Nb-13Zr, a near-β alloy [23]). These alloys if appropriately processed display good formability, high hardenability,

excellent corrosion resistance, and better notch sensitivity than the (α + β) Ti alloys. Annealing results in higher strength as a result of precipitation strengthening but this also results in an increase in modulus [20].

β-Ti alloys are characterized by retention of 100% β-phase on cooling from above the β-transus [22]. Considering the pseudobinary phase diagram of Ti + β-stabilizer alloying additions (Fig. 2.13), a metastable β-phase is retained on cooling from above the β-transus if the β-stabilizer content is greater than $β_c$ and less than $β_s$. For levels greater than $β_s$, β-phase represents the equilibrium structure. The metastable β structures are characterized by low elastic modulus (55–75 GPa) but relatively low fatigue strength (e.g., 10^7 fatigue endurance limit ∼265 MPa for the Ti-35Nb-5Ta-7Zr alloy (TNZT) (Table 2.8). As noted above, annealing the metastable β in the two-phase (α + β) field region results in some α-phase precipitation and an increase in strength but also a rise in elastic modulus to values approaching or just above 100 GPa (dependent on the amount of α-phase that forms).

Ti-13Nb-13Zr is an example of a so-called near β-Ti alloy [23]. Rapid cooling (water quenching) this alloy results in suppression of the β-transus temperature to a low enough temperature (<575 °C) to cause formation of acicular α′ martensite, an hcp phase formed by a displacive transformation. This structure is characterized by a Young's modulus ∼64–77 GPa (Table 2.8). Subsequent annealing in the (α + β) two-phase field (500 °C for 6 h – an aging heat treatment) causes the acicular α′ to coarsen and fine β-phase precipitates to form throughout, resulting in higher strength (because of precipitation strengthening) with an associated increase in elastic modulus to about 81 GPa. Heavy cold-working of α′ martensite is possible and results in the formation of a lower modulus material (∼45 GPa) with strength even greater than water-quenched and precipitation hardened material while maintaining good ductility. Surface hardening for improved wear resistance can be achieved by aging (at 500 °C, for example) in an oxygen-containing environment [23]. This is due to the formation of a hard surface oxide layer as well as interstitial solid solution strengthening of the subsurface region (diffusion hardening).

2.20 Zr-Nb Alloy

Zirconium, like titanium, is a highly reactive metal and will form a dense cohesive surface oxide layer (ZrO_2) spontaneously on exposure to an oxygen-containing environment. In addition to the resulting corrosion protection, ZrO_2 is very hard and can be used to form a good wear resistant surface assuming sufficient layer thickness. A Zr-2.5Nb alloy, developed initially for nuclear industry applications, when annealed in an oxygen-containing atmosphere at 500 °C develops a relatively thick (5 μm) monoclinic ZrO_2 layer over the alloy substrate [24]. The underlying alloy consists of a two-phase (α + β) structure with the dominant phase being hcp α with a small percentage of bcc β phase distributed throughout. Following the 500 °C surface oxidation treatment, the sub-oxide alloy zone is also hardened through oxygen diffusion into the lattice (interstitial solid solution strengthening). Surface oxidized

Zr-Nb alloy is used for making orthopedic components that are primarily intended for compressive loading and resisting wear (such as femoral hip implant and knee implant components). The structure offers the advantage over whole ceramic components such as Al_2O_3 or phase-stabilized ZrO_2 of reducing risks of catastrophic fracture as a result of rapid crack initiation and propagation within the ceramic phase since the ZrO_2 layer is reinforced by a Zr-Nb alloy body. Clinical outcomes of implants using this novel system are being followed with great interest.

2.21 Ni-Ti Alloys (Nitinol)

The equiatomic Ni-Ti alloy (Nitinol) is used currently in orthopedic, dental and cardiovascular applications [25]. The shape-memory effect that the alloy displays as well as its reported good corrosion resistance (the result of a TiO_2 passive surface layer) and its pseudoelastic property (E ~28–41 GPa for the martensitic phase) has attracted considerable interest in the biomaterials field in recent years. In vivo studies have indicated that the material is biocompatible despite its high Ni content [26, 27]. However, the concern over long-term consequences of even minor amounts of Ni ion released are likely to limit its use. Published reports indicate that with proper processing, corrosion properties comparable to 316L stainless steel are achievable [28–30] so that use in limited term applications such as fracture fixation is practical.

The shape memory effect is due to the thermoelastic martensitic transformation that occurs with the Ni-Ti shape memory alloys. Cooling from a high temperature (i.e., above body temperature for some Ni-Ti-based alloys), to a lower temperature results in the transformation of the austenite phase (an ordered intermetallic compound with a distorted bcc unit cell structure – a result of the ordering of Ni and Ti atoms giving a so-called B2 unit cell structure) to martensite. The transformation is isothermal with most of the phase change occurring over a fairly narrow temperature range between M_s and M_f. The temperature at which the austenite-to-martensite (or reverse) transformation occurs is dependent on the exact composition of the alloy. The Ni-Ti intermetallic compound, unlike most other intermetallic compounds, has a moderate range of solubility for excess Ni or Ti as well as other elements and this allows 'engineering' of the critical transformation temperatures (M_s, M_f, A_s and A_f). Thus, it is possible to develop shape memory alloys (SMAs) with transformation temperatures corresponding to body temperature. The austenite-martensite transformation can also be induced by mechanical deformation. While the overall change in shape due to cooling of austenite is minor due to the formation of a number of counteracting twin variants that develop thereby minimizing the overall shape change, as discussed previously, stress-induced martensite results in the development of a preferred twin variant. This involves a so-called *detwinning* reaction in which a twin variant preferentially oriented in the principal strain direction grows at the expense of other twins resulting in significant volume and net shape changes. If this shape change is constrained, appreciable stress can develop in the material

and this stress can be transferred to adjacent tissues, for example, (assuming secure anchorage of tissue to the Ni-Ti component). By way of example, a fracture fixation staple made of Nitinol, if deformed in the martensitic state (i.e., at room temperature) and then used to unite fractured bone fragments, will exert a strong force pulling the fragments together as it warms up in the body to above its A_f temperature and reverts to its prior austenitic shape. Other applications have been proposed in orthopedics including other fracture fixation devices, spinal rods for treatment of scoliosis, cages for use during spinal fusion and even self-locking joint replacement components [29].

Wires formed of Nitinol are used currently in orthodontics although, in this application it is the alloy's pseudoelasticity that makes it so attractive. SMAs are also used for cardiovascular stents where the pseudoelastic feature facilitates stent expansion/deployment following insertion into an artery. The ability to tailor the transformation temperature of Ni-Ti alloys through slight alloying modifications as noted above is important since this allows the formation of materials and devices displaying this superelastic behavior at a desired temperature corresponding to expected in-service conditions. Orthodontic wires formed from a Ni-Ti-based alloy are useful because of the large 'working range' that these wires provide during force application for tooth re-positioning.

The stress-strain curves for NiTi alloys are dependent on the M_s temperatures with yield strengths and fatigue strengths increasing with decreasing M_s. Accurate determination of alloy transformation temperatures is critical for the design of components and prediction of performance.

2.22 Tantalum

Porous tantalum foam structures are used as bone augmentation templates. Ta has been considered for many years to be biocompatible as a result of the stable tantalum oxide (Ta_2O_5) that forms on its surface. The metal has a long history of use in other implant applications such as implantable cranial plates. Ta foam structures for bone augmentation use [31] are formed by chemical vapor deposition (CVD) and infiltration (CVI) of tantalum onto vitreous carbon lattice structures, thereby building up Ta struts of desired dimensions. The vitreous carbon lattices are formed by pyrolysis of thermosetting polymer foams. The Ta is deposited onto the carbon filamentous structure to a desired thickness giving an open-pored Ta structure (approximately 99 wt% Ta + 1 wt% vitreous carbon) (Fig. 2.16). The dimensions of the Ta struts are controlled by the amount of Ta applied and this determines the mechanical properties of the foam. Structures with compressive and tensile strengths ~60 and 63 MPa respectively and compressive fatigue strengths ~23 MPa (5×10^6 endurance limit) have been reported [31]. Continuous interconnecting pores, or cells, about 550 μm in size, are formed and elastic and strength properties similar to those of cancellous bone result. Animal studies [32] have demonstrated the effectiveness of these structures for bone augmentation purposes. While the as-made mechanical properties of the foams are low, they are considered sufficient to allow bone ingrowth and filling

Fig. 2.16 Scanning electron micrograph of open-pored Ta foam (courtesy Dr JD Bobyn, McGill University, Montreal, Canada)
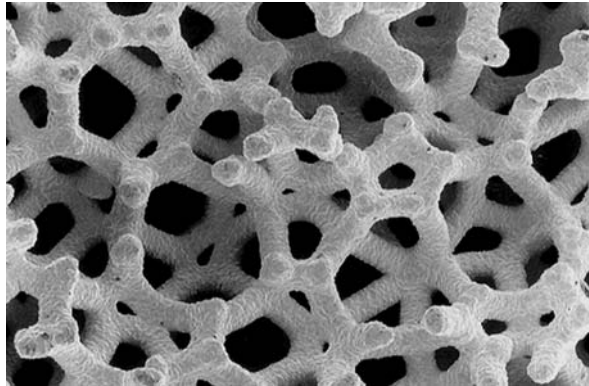
of the structures after which bone is expected to eventually satisfy the load-bearing requirements.

2.23 Platinum, Platinum-Iridium

Electrodes used for neuromuscular stimulation require the use of materials that are corrosion resistant under extreme voltage potential and charge transfer conditions. The noble metals (Pt and Pt-Ir alloys) satisfy these conditions and are the principal materials used for making electrodes for electrical stimulation and sensing purposes. The Pt-Ir alloys (10–30 wt% Ir) offer the advantage of higher mechanical properties (scratch resistance) compared with unalloyed Pt because of solid solution strengthening. Other possible noble metals that have been used for fabricating electrodes for electrical stimulation include rhodium, gold and palladium but these are not used extensively [7]. Some non-noble metals have also been used (stainless steel, Ti, Co-based alloys) but these are more prone to corrosion (passive film breakdown occurs at potentials above the breakdown potential). Hence, these alloys are not as suitable as the noble metals.

Conducting leads used with the electrodes are made of metals that, in addition to being corrosion resistant and electrically conductive, must display good flexural fatigue strength. Helically coiled lead wire designs are used currently to achieve good lead flexibility and to minimize local strains during flexing thereby improving fatigue characteristics. Stainless steel, wrought CoCrMo alloy (Elgiloy) and Co-Ni alloys (MP35N) are popular choices for conducting leads [7].

2.24 Dental Alloys

Metals are used in dentistry for direct fillings in teeth (dental amalgams), fabricating crowns and bridges (noble metal and base metal alloys), partial denture frameworks (base metal alloys), orthodontic wires and brackets (stainless steel, Ti alloys

and Ni-Ti alloys) and dental implants (CP Ti and Ti6Al4V). As in the orthopedic applications, the major advantage of metal for these dental applications is the high intrinsic strength and fracture resistance of this class of materials. Biocompatibility is again an important requirement since these materials also contact body tissues (tooth structure, soft supportive tissues) often for the remainder of a patient's lifetime although the ease of accessibility after placement relaxes the biocompatibility requirement considerably. A number of the metals already described (as indicated above) find application in dentistry. Dental implant materials with requirements very similar to materials used for orthopedic joint replacement implants are made almost exclusively from Ti and Ti6Al4V. Orthodontic wires and brackets are made of stainless steel (types 302, 303, 304 and 305), CoCrNiMo alloys (Elgiloy), β-Ti, and Ni-Ti alloys (because of their low elastic moduli, high strengths and consequently large working range, a desirable characteristic for this application).

2.25 Dental Amalgams

Dental amalgams are formed by adding Hg to dental amalgam alloys (alloys containing Ag, Cu and Sn plus some other minor elemental additions) (amalgamation process) [33]. Dental amalgam alloys are either low (Cu <6 wt%) or high Cu-containing (Cu >6 wt%), the latter being favored because it avoids the formation of an undesirable Sn-Hg phase (γ_2) that is susceptible to corrosion and results in lower strength properties of amalgams. Dental amalgam alloys are formed as powders either by lathe cutting Ag-Cu-Sn alloy billets (resulting in irregular particles – i.e., machining chips) or by atomization (to give spherical powders). Subsequent mixing of these alloy powders with liquid mercury results in their partial dissolution, complete consumption of the liquid Hg and the subsequent formation of a number of intermetallic compounds (Ag_3Sn, Ag_2Hg_3, $Sn_{7-8}Hg$, Cu_3Sn, Cu_6Sn_5) due to the Hg-dental amalgam alloy reactions and the condensation of the initial 'plastic' mass to form a load-bearing filling. During the reaction, the partially reacted powder can be manipulated to fill a tooth cavity. The final amalgam restoration exhibits reasonable mechanical properties. While susceptible to corrosion in the oral environment, buildup of the corrosion product serves to limit the rate of further corrosion and to form an acceptable marginal seal at the amalgam-tooth interface. The major advantage of this material is its easy in situ formability to a desired shape. Concerns related to Hg toxicity have been raised but, to date, these have not been proven to be valid although the issue remains controversial.

2.26 Dental Casting Alloys – (Au-based, Co- and Ni-based, Ti-based)

Dental casting alloys are used for making dental bridges, crowns (with porcelain fused to a metal substrate), inlays, onlays, and endodontic posts. Both noble and non-noble (base) metal alloys are used to form these often complex shapes.

Table 2.9 Dental casting alloys [7]

Alloy type	Ag	Au	Cu	Pd	Pt	Zn	Other
High noble							
Au-Ag-Pt	11.5	78.1	–	–	9.9	–	Ir (trace)
Au-Cu-Ag-Pd-*I*	10.0	75.0	10.5	2.4	0.1	1.0	Ru (trace)
Au-Cu-Ag-Pd-*II*	25.0	56.0	11.8	5.0	0.4	1.7	Ir (trace)
Noble							
Au-Cu-Ag-Pd-*III*	47.0	40.0	7.5	4.0	–	1.5	Ir (trace)
Au-Ag-Pd-In	38.7	20.0	–	21.0	–	3.8	In 16.5
Pd-Cu-Ga	–	2.0	10.0	77.0	–	–	Ga 7.0
Ag-Pd	70.0	–	–	25.0	–	2.0	In 3.0
Base metal	**Ni**	**Cr**	**Co**	**Ti**	**Mo**	**Al**	**V**	**Other**
Ni-Cr	69–77	11–20	–	–	4–14	0–4	–	Fe, Be, Ga, Mn, B
Co-Cr	–	15–25	55–58	–	0–4	0–2	–	Fe, Ga, Nb, W, B, Ru
Ti	–	–	–	90–100			0–4	

Requirements include sufficient strength, toughness, wear resistance, corrosion resistance and biocompatibility. The noble metal alloy compositions are primarily Au- or Pd-based with alloying additions of Ag, Cu, Pt, Zn and some other trace elements. These can be divided into 'high noble' alloys (noble metal content $\geq$60 wt%) and 'noble' alloys (noble metal content $\geq$25 wt%) [7]. Some of the Au-based alloys are heat treatable with precipitation hardening or order-disorder reactions (similar effect to precipitation hardening for some alloys) contributing to increased strengths. The base (non-noble) metal alloys contain $\leq$25 wt% noble metal elements and are either CoCr or NiCr alloys (Table 2.9). Cast Ti components for crowns, partial and complete dentures, while limited because of difficulties associated with casting because of the high melting point of Ti (compared to Au-based dental alloys), its high reactivity, difficulty in surface finishing and other problems, can be made (using special vacuum melting and casting equipment) but it is not common. Such cast Ti components are useful, however, for individuals with allergies to Ni- or Co-based alloys. Cast Ti-based alloys are also possible including Ti6Al4V, TiCuNi, TiV, TiCu, TiPd and TiCo alloys (some of these are still in the experimental stage of development). To satisfy esthetic requirements for dental crowns, porcelain-fused-to-metal (PFM) restorations are made with the silicate-based porcelains being bonded to a cast metal substrate. Attainment of good bonding of porcelain to the dental alloy substrate is achieved through micromechanical interlock and interfacial chemical reactions, the latter usually involving a substrate surface oxide reacting with the porcelain coating. Compatible thermal expansion coefficients of metal coping and porcelain cladding materials are important to ensure slight residual compressive stresses in the porcelain for inhibition of cracking. Corrosion resistance of the underlying dental alloy is necessary to avoid unacceptable staining and discoloration of a PFM crown.

2.27 Wrought Dental Alloys

As noted previously, wrought stainless steel, CoCrNi (Elgiloy), β-Ti and Nitinol alloys are used for making orthodontic wires where high yield strength and preferably low elastic modulus provide high 'working range' characteristics. The requirement of low elastic modulus favors the selection of β-Ti and Ni-Ti alloys for orthodontic wires although all four alloys are used at present. Wire drawing procedures result in strain hardening and high yield strengths but can be performed such that the wires are sufficiently ductile to allow their manipulation and shaping during placement.

CP Ti and (α + β) Ti alloys (Ti6Al4V) are used for making endosseous dental implants. These components are prepared by machining shapes from bar stock followed by appropriate surface finish operations. In order to achieve more rapid osseointegration, implant surfaces are modified using plasma spray coating, acid etching, grit blasting, laser ablation or addition of metal powder sintered surface layers [34]. These additional treatments can result in final products with either mill-annealed microstructures or, as in the case of Ti6Al4V implants with sintered surface layers, β-annealed structures. In addition to geometric or topographic surface modifications, surface chemical modifications (addition of calcium phosphate surface layers using one of a number of possible methods) are used to promote increased osteoconductivity resulting in faster rates of osseointegration.

2.28 New Directions

The metals and alloys used for forming biomedical implants to date have consisted of conventional polycrystalline structures. Methods of formation, as noted previously, have involved casting, forging, milling, machining and honing, metal powder compaction and sintering including the new approach of metal injection molding (MIM) to form near net-shape parts, hot isostatic pressing as well as other conventional metal processing methods (wire drawing, rolling, extrusion and so on). All these result in metallic microstructures with grains in the micron to millimeter size range. In order to achieve preferred mechanical properties grain refinement, alloy additions, strain hardening, development of multi-phase structures and precipitation hardening are used to achieve high strength with acceptable ductility, and fatigue strength in particular, for highly loaded components. Wear resistance of these conventional metals is dependent on the formation of hard non-metallic phases ($M_{23}C_6$ for example) or surface modification through ion implantation or physical vapor deposition (PVD) of hard non-metallic surface layers. Corrosion resistance for all but the noble metals relies on formation of stable passive oxide layers. Over the past 20–30 years there have been increasing studies on the formation of nanocrystalline (grain size <100 nm) and amorphous (glassy) metals for various applications. Much higher yield strength, wear resistance and corrosion resistance has been shown for some pure metals formed with such fine microstructures. The higher strengths are explained by the greater resistance to mechanical deformation due to inability for

dislocation movement within the nanoscale crystals. Studies on optimizing structures to achieve high strength and hardness with a required degree of ductility have involved the use of bimodal grain size distributions, with nanocrystalline and some microcrystalline regions [35]. The much greater hardness of the nanocrystalline metals suggest their potential as surface coatings for achieving better wear resistance than conventional alloys without the potential disadvantage of hard nonmetallic particles that might become dislodged forming hard 3rd body wear particles. Improved corrosion resistance has also been observed with nanocrystalline metals. While the reason for this improved corrosion resistance is not fully understood, it is suggested that this may be due to reduced electrochemical potentials between anodic and cathodic zones within the nanocrystalline structures [36]. In effect the anodic intergranular and the cathodic intragranular regions are less distinct in the refined structure so the driving potential for corrosion is reduced. More rapid re-passivation due to the high grain boundary area has also been proposed as a reason for the enhanced corrosion resistance.

Methods that have been reported for forming nanocrystalline and amorphous metals and coatings involve rapid melt cooling rates ($\sim 10^6$ °C/s) through atomization or melt spinning methods, extreme mechanical deformation to give heavy mechanical deformation followed by recrystallization anneals to yield high nucleation rates with limited grain growth, mechanical attrition or milling of gas atomized powders followed by plasma spraying (for coatings), pulsed electrodeposition and other PVD and CVD methods (for coatings and possibly bulk components). In addition to developing very fine grain structures, some of these approaches also offer the possibility of unique non-equilibrium phases forming at room temperature. The determination of the properties of the novel compositions and structures possible by these methods (including their biocompatibility characteristics) presents an exciting area for future biomaterials research.

References

1. Ashby MF and Jones DR. Engineering Materials 1. An Introduction to their Properties and Applications, and 2. An Introduction to Microstructures, Processing and Design, Pergamon Press: London, 1980.
2. Newey C and Weaver G. Materials Principles and Practice, Butterworth Scientific Ltd.: Guildford 1990.
3. Callister WD Jr. Materials Science and Engineering: An Introduction, 6th ed, Chapter 7, John Wiley & Sons, Inc, 2003.
4. Wayman CM and Duerig TW. An introduction to martensite and shape memory, in Deurig TW, Melton KN, Stockel, D, and Wayman, CM, eds. Engineering Aspects of Shape Memory Alloys, 1990, pp. 3–20.
5. Pilliar RM. Modern metal processing for improved load-bearing surgical implants. Biomaterials, 1991, 12: 95–100.
6. Williams DF. Corrosion of orthopedic implants, in Williams DF, ed. Biocompatibility of Orthopedic Implants, Vol 1, chap. 6, CRC Press: Boca Raton, FL, 1982.
7. Davis JR, ed. Metallic Materials, Chapter 3, in *Handbook of Materials for Medical Devices*, ASM International, Materials park, Ohio, USA, 2003.
8. Pilliar RM and Weatherly GC. Developments in implant alloys, CRC Critical Reviews in Biocompatibility, Vol 1, 1986, pp. 371–403.

9. Sigwart U. Endoluminal Stenting, WB Saunders Co Ltd: London, UK, 1996.
10. Tandon R. Ne-shaping of Co-Cr-Mo (F75) via Metal Injection Molding, in Disegi JA, Kennedy RL, and Pilliar R, eds. Cobalt-Base Alloys for Biomedical Applications, ASTM STP 1365, American Society for Testing and Materials: West Conshohocken, PA, 1999, pp. 3–10.
11. Kilner T, Laanemae WM, Pilliar RM, Weatherly GC, and MacEwen SR. Static mechanical properties of cast and sinter-annealed cobalt-chromium surgical implants, J Mater Sci, 1986, 21: 1349–1356.
12. Taylor RNJ and Waterhouse RB. A study of aging behaviour of a cobaltbased implant alloy, J Mater Sci, 1983, 18:3265–3280.
13. Kilner T, Pilliar RM, Weatherly GC, and Allibert C. Phase identification and incipient melting in a cast co-Cr surgical implant alloy. J Biomed mater Res, 1982, 16:63–79.
14. Wang KK. A dispersion strengthened Co-Cr-Mo alloy for medical implants, in Disegi JA, Kennedy RL, Pilliar R, eds. Cobalt-Base Alloys for Biomedical Applications, ASTM STP 1365, American Society for Testing and Materials West: Conshohocken, PA, 1999, pp. 89–97.
15. Berlin RM, Gustavson LJ, and Wang KK. Influence of post processing on the mechanical properties of investment cast and wrought Co-Cr-Mo alloys, in Disegi JA, Kennedy RL, Pilliar R, eds. Cobalt-Base Alloys for Biomedical Applications, ASTM STP 1365, American Society for Testing and Materials: West Conshohocken, PA, 1999, pp. 62–70.
16. Mishra AK, Hamby MA, and Kaiser WB. Metallurgy, microstructure, chemistry and mechanical properties of a new grade of cobalt-chromium alloy before and after porous-coating, in Disegi JA, Kennedy RL, Pilliar R, eds. Cobalt-Base Alloys for Biomedical Applications, ASTM STP 1365, American Society for Testing and Materials: West Conshohocken, PA, 1999, pp. 71–88.
17. Pilliar RM. Powder metal-made orthopaedic implants with porous surfaces for fixation by tissue ingrowth, Clin Orthop Rel Res, 1983, 176:42–51.
18. Pilliar RM. Porous-surfaced metallic implants for orthpedic applications, J Biomed Mater res, 1987, A1:1–33.
19. Ankem S and Seagle SR. Heat treatment of metastable beta titanium alloys, in Rosenberg H and Boyer RR, eds. Beta Titantium Alloys in the 1980's, AIME: New York, 1984, pp. 107–126.
20. Long M and Rack HJ. Titanium alloys in total joint replacement – a materials science perspective, Biomaterials, 1998, 19:1621–1639.
21. Yue S, Pilliar RM, and Weatherly GC. The fatigue strength of porous-coated Ti-6Al-4 V implant alloy. J Biomed Mat Res, 1984 18: 1043.
22. Bania PJ. Beta titanium alloys and their role in the titanium industry, in Eylon D, Boyer RR and Koss DA, eds. Beta Titanium Alloys in the 1990's, The Minerals, Metals & Materials Soc.: Warrendale, PA, 1993, pp. 3–14.
23. Mishra AK, Davidson JA, Poggie RA, Kovacs P, and Fitzgerald TJ. Mechanical and tribological properties and biocompatibility of diffusion hardened Ti- 13Nb-13Zr – A new titanium alloy for surgical implants, in Medical Applications of Titanium and its Alloys: The Material and Biological Issues ASTM STP1272, 1996, pp. 96–113.
24. Benezra V, Mangin S, Treska M, Spector M, Hunter G, and Hobbs LW. Microstructural investigation of the oxide scale of Zr-2.5Nb and its interface with the alloy substrate, in Biomedical Materials-Drug Delivery, Implants and Tissue Engineering, Mater Res Soc Symp Proc, 1999, 550: 337–342.
25. Haasters J, v Salis-Salio G, and Bensmann G. The use of Ni-Ti as an implant in orthopaedics, in Deurig TW, Melton KN, Stockel D, and Wayman CM, eds. Engineering Aspects of Shape Memory Alloys, 1990, pp. 426–444.
26. Wever DJ, Veldhuizen AG, Sanders MM, Schakenrad JM, and van Horn JR. Cytotoxic, allergic and genotoxic activity of a nickel-titanium alloy. Biomaterials, 1997, 18:1115–1120.

27. Kapanen A, Ryhanen J, Danilov A, and Tuukkanen J. Effect of nickel-titanium shape memory alloy on bone formation. Biomaterials, 2001, 22: 2475–2480.
28. Firstov GS, Vitchev RG, Kumar H, Blanpain B, and Van Humbeeck J. Surface oxidation of NiTi shape memory alloy. Biomaterials, 2002, 23:4863–4871.
29. Cisse O, Savadogo O, Wu M, and Yahia LH. Effect of surface treatment of NiTi alloy on its corrosion behavior in Hanks' solution. J Biomed Mater Res, 2002, 61: 339–345.
30. Assad M, Chernyshov AV, Jarzem P, Leroux MA, Coillard C, Charette S, and Rivard C-H. Porous titanium-nickel for intervertebral fusion in a sheep model: Part 2. Surface analysis and nickel release assessment. J Biomed Mater Res Pt B: Appl Biomater, 2003, 64B:121–129.
31. Zardiackas LD, Parsell DE, Dillon LD, Mitchell DW, Nunnery LA, and Poggie R. Structure, metallurgy and mechanical properties of a porous tantalum foam. J Biomed Mater Res (Appl Biomater), 2001, 58:180–187.
32. Bobyn JD, Stackpool GJ, Hacking SA, Tanzer M, and Krygier JJ. Characteristics of bone ingrowth and interface mechanics of a new porous tantalum biomaterial. J Bone Jt Surg, 1999, 81-B: 907–914.
33. Phillip's Science of Dental Materials, in Anusavice KJ, ed. Dental Amalgam: Structures and Properties, chap 17, 10th Edition, 1996.
34. Pilliar RM. An Overview of Surface Variability of Metallic Endosseous Dental Implants: Textured and Porous surface-Structured Designs. Implant Dent, 1998, 7: 305–314.
35. Baker SP. Plastic deformation and strength of materials in small dimensions. Mat Sci Eng, 2001, 319–321: 16–23.
36. Cheng D, Tellkamp VL, Lavernia CJ, and Lavernia EJ. Corrosion properties of nanocrystalline Co-Cr coatings. Ann Biomed Eng, 2001, 29: 803–809.

Chapter 3
Polymeric Biomaterials

Teerapol Srichana and Abraham J. Domb

3.1 Introduction

Recognition of the chemical nature of polymers began in the early 1920, when Staudinger concluded from his research on starch, natural rubber and cellulose that these compounds consisted of giant chains of carbon atoms (plus oxygen atoms in the case of the polysaccharides) held together by covalent bonds. "They were truly macromolecules." This notion was met with disbelief and even ridicule by many of his colleagues. Not until the 1930s, when linear polymers were synthesized, especially by W.H. Carothers in the Du Pont laboratories, did Staudinger's concept gain universal acceptance. The Nobel Prize in chemistry was finally given to him in 1953.

Medical devices have been employing polymers in every aspect for many years. Typical examples are polyethylene and polyolefin bottles, polystyrene vials, rubber closures, plastic tubing for injection sets, flexible bag of plasticized polyvinyl chloride to hold blood and intravenous solutions. Barrels and plungers of hypodermic syringes are made of polypropylene. In addition to polymers used as excipients, some drugs themselves are polymers including insulin, heparin and its antagonist, protamine sulfate, the plasma extenders dextran, human serum albumin and bulk laxative methylcellulose. Thus, it is seen that polymers are essential to all medical fields.

3.2 Nomenclature

Generally, polymers are named by either common names or by the IUPAC naming system. The common name is given on the basis of starting monomer used for polymerization. According to this system, the polymer is named as a prefix 'poly' before the name of monomer used written in brackets. For example, the name poly(ethylene glycol) is derived for the polymer prepared from ethylene glycol. Similarly, the name poly(ethylene oxide) is derived from the ring opening polymerization of ethylene oxide, although it has the same chemical structure of poly(ethylene glycol). According to the IUPAC system the polymers are named by writing the repeating units within bracket and 'poly' as a prefix [1]. The common and IUPAC names of some polymers with their monomers used are listed below:

Table 3.1 Common and IUPAC Nomenclature of polymers

Monomer Used	Common Name	IUPAC Name
Ethylene	Poly(ethylene)	Poly(methylene)
Vinyl acetate	Poly(vinyl acetate)	Poly(1-acetoxyethylene)
Styrene	Poly(styrene)	Poly(1-phenylethylene)
Vinyl chloride	Poly(vinyl chloride)	Poly(1-chloroethylene)
Formaldehyde	Poly(formaldehyde)	Poly(oxymethylene)
Ethylene glycol	Poly(ethylene glycol)	Poly(oxyethylene) etc.

3.3 Biopolymer in Medical Applications

Biopolymers based on medical applications can be categorized into four groups, namely inert, natural, bioactive and biodegradable polymers. This classification does not clearly separate one group of polymers from the others. The first group that was introduced into the medical field is inert polymers, which are widely used. They suit very well with some applications that are aimed to last for a lifetime, such as breast implant polymer. When the concern over the load of non-biodegradable polymeric materials increases dramatically, the situation drives to the use of biodegradable instead. In particular, if the functions of the polymers are complete, they should be removed from the body in biomedical applications. Biomedical polymers in medicine and surgery are currently widely used and include intracorporeal, para-corporeal and extracorporeal applications (inside, interfacing or outside the body, respectively). Applications are given below.

Intracorporeal (implanted) materials
Temporal devices
>Surgical dressing
>Sutures
>Adhesives
>Polymeric intermedullary nails
>Polymer-fiber composite bone plates

Semipermanent devices
>Tendons
>Reinforcing meshes
>Heart valves
>Joint reconstruction and bone cement
>Tubular devices
>Soft tissue replacement
>Interocular and contact lenses
>Drug delivery implants

Complex devices
>Artificial kidney/blood dialysis

Artificial lungs/blood oxygenator

Artificial pancreas/insulin delivery system

Artificial heart

Paracorporeal or extracorporeal materials

Catheters

Blood bags

Pharmaceutical containers

Tubing

Syringes

Surgical instruments

The science and technology of polymers for biomedical applications has been dramatic for several years. Medical polymers must adhere to very rigid standard and must be non-toxic, non-carcinogenic, biocompatible and in no way injurious to the biological environment.

Artificial organs and prostheses constitute some of the most valuable remedies in present medical treatment. They have contributed immeasurably to the welfare and health of the human race. Although the production of medical polymers is small compared with other industrial plastics, the number of individuals who benefits from these uses is large (Table 3.2).

Table 3.2 Medical applications of biomaterials in the United States (1983) (2002)* [2, 3]

Applications	Number per year
Permanent implantations	
Heart valve	30,000
Pacemakers	130,000*
Vascular grafts	250,000*
Hips prostheses	110,000
Knees prostheses	65,000
Shoulders and finger joints	50,000
Retinal surgery	35,000
Intraocular lenses	2,700,000*
Breast prostheses	100,000
Nose and chin prostheses	10,000
Urinary incontinence	2,500
Hydrocephalus shunts	21,500
Intermediate applications	
Contact lenses	30,000,000*
Dental prostheses	15,000
Renal dialyzers	16,000,000*
Ventricular assists	>100,000*
Transient applications	
Cardiopulmonary bypass	150,000
Blood oxygenators	200,000
Over the needle catheters	200,000,000*
Therapeutic catheters	150,000
Infusion catheters	300,000

Table 3.3 Medical uses of polymers

Polymer	Applications
Polyvinyl chloride	Extracorporeal devices; hemodialysis or hemoperfusion, blood tubing, cardiac catheters, blood bag and IV infusion set, endotracheal tubes surgical tapes, sheet oxygenator, artificial heart, blood pump, artificial limb
Ultrahigh MW polyethylene	Acetabulum in total hip prostheses, artificial knee prostheses
Polypropylene	Membrane oxygenator, finger joint prostheses, IV cannulae, unabsorbable sutures
Silicone rubber	Hydrocephalus shunts, catheters, membrane for oxygenator, artificial skin for burn dressing, plastic surgery implant, artificial heart, heart assisted pump, drug release system, atrioventricular shunts, ear prostheses, facial prostheses, artificial heart valve, tendon, finger joint repair, tracheal prostheses, bladder prostheses, bladder patch, intestine patch, dura-mater prostheses, retinal detachment, impressing materials, heart pacemaker leads
Polycarbonates	Membrane for oxygenator, hemodialyzer, plasmapheresis membrane
Polyester	Vascular graft prostheses, fixation device for tissue, hernia repair, patches for heart, bladder, arteries, suture
Polytetrafluoroethylene	Vascular graft prostheses, heart patch, retinal detachment, femoral stems
Polyurethane	Artificial heart pump material, balloon, heart valve prostheses, vascular graft prostheses, coating for blood compatibility
Polymethyl methacrylates	Bone cement, artificial teeth, denture material, bone prostheses, cranial bone replacement, intraocular lenses, membrane for dialysis

Polymers penetrate virtually every aspect of medicine, although the science of polymeric biomaterials is much more recent than that of other high molecular weight polymers. A few polymers have been designed for medical use, e.g., hydrogels for soft contact lenses, poly(glycolic acid) for absorbable sutures, special ion exchange resin, semipermeable membranes, and more are likely to be created (Table 3.3). Although production may be low in some cases, patient benefits are high enough to justify industry's support of the high costs of research and development.

Most engineering progress through the centuries was dependent on the discovery and availability of new materials. In fact many modern technologies require materials with unusual combinations of mechanical properties, lightness and ease of processing. For example, biomedical applications are increasingly searching for structural materials that have low density, are strong, stiff, abrasion and impact resistant and are not easily corroded. Table 3.4 gathers some characteristics of polymers used in medical applications.

Table 3.4 Characteristics of typical thermoplastic polymers [4]

Material	Tensile modulus (GPa)	Tensile strength (MPa)	Glass transition temperature (°C)	Melting point (°C)	Processing temperature (°C)
Polypropylene	1.1–1.6	30–40	–10	165	200–240
Polyamides	3–3.3	80–90	47–57	225	240–270
Polyethyleneterephthalate	2.7–4	50–70	70	265	280–310
Polycarbonate	2.3–3	60–70	150	Amorphous	280–330
Polyglycolic acid	6.5	57	35	225	200–205
Polylactic acid	2.1	59–79	60	180	180–200
Polycaprolactam	0.3	19	–60	210–255	<300
Polyhydroxybutyrate	2.5–3.5	36–40	1	171	95
Polyorthoester	0.8–1.2	20–27	55–95	Amorphous	95
Polyanhydride	0.04	4	–	46–49	40–50

3.4 Inert Polymers

Inert durable polymers (non biodegradable) do not undergo any chemical change in vivo. Both hydrophobic and hydrophilic polymers may fit this category. Table 3.5 gives a list of the commonly used inert polymers.

3.4.1 Silicones

Silicones were one of the first polymers studied due to their excellent biocompatible and ease of fabrication into various medical devices. A number of medical products employing silicones have been commercialized including implants for contraceptives, and ear replacement (Fig. 3.1). A class of polymers considerable importance is based on a linear, cyclic or crosslinked arrangement of alternating silicon and oxygen atoms, where the silicon is substituted with organic radicals or hydrogen. They are called polyorgano-polysiloxanes or simply "silicone" polymers.

The usual procedure for preparing silicone polymers is to hydrolyze single or combinations of R_3SiCl, R_2SiCl_2, $RSiCl_3$ and $SiCl_4$. The intermediates in the reaction are believed to be silanols which condense very rapidly with the elimination of water and formation of the Si-O-Si link (Scheme 1).

Table 3.5 Commonly used inert polymers

1. Silicones
2. Polyacrylates
3. Polyethylene and related polymers
4. Polyamides
5. Polyurethanes
6. Polyesters
7. Polyethers

Fig. 3.1 Silicone ear replacement
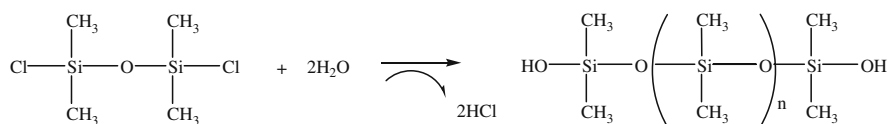
Scheme 1

The linear silicones $[(CH_3)_3Si(OSi(CH_3)_2]_n–OSi(CH_3)_3$, where n is in the range of 10–20, are known as silicone oils. The cyclic silicones, formed in hydrolysis reactions of the silane dihalides, especially $[(CH_3)_2SiO]_{3-4}$, are convertible to high molecular weight linear silicone elastomers. Various curing techniques are available for converting linear and cyclic materials to crosslinked elastomers and resins. Linear silicone polymers are conveniently prepared from a base-catalyzed ring opening polymerization of the cyclic tetramer or trimer.

Silicones have some unique chemical and physical properties. They are inert to cells and tissues and they are stable to heat and irradiation, which suggests that silicone based products are sterilizable. It was these properties that inspired scientists and physicians to consider the use of silicones for medical applications [5]. Silicones are now used in many surgical and non-surgical procedures and their use extends to a variety of medical instruments.

Silicones have a specific gravity lower than water and are commonly used fluids in medical practice. The viscosity of oil is directly related to molecular weight but also the ability to slip past one another. Silicones are produced as oils and gels, gums and elastomers. The strength and elasticity of silicone elastomers is a function of polymer chain length, organic side groups and the degree of cross-linking. The basic ingredients to make silicone elastomers clear and highly viscous are vinyl group substituents. Other radicals in the silicone rubber determine the physical properties of the resultant elastomer. The most common mechanism of crosslinking occurs by a radical attack on the pendant alkenyl groups from the silicon initiates by heat

(heat vulcanization) or benzoyl peroxide. The reaction intermediates containing carbon free radicals then combine to form the carbon-carbon covalent cross-linking bond. In some cases, crosslinking facilitated substitution of vinyl groups takes place because the double bond of this group is quite susceptible to free radical attack. In order to increase tensile strength, the prevulcanized silicone is usually mixed with fillers such as fumed silica, talc and TiO_2. By changing the amount of filler added to a silicone rubber elastomer, one can change its degree of hardness.

The prototype of a silicone rubber sac filled with silicone gel was seen as a mammary prosthesis and was later marketed by Dow Corning after undergoing the first human trial. During 1960–1970, Dow Corning introduced many improvements to enhance the mammary prosthesis, which became a multimillion dollars product. Lately, silicone breast implants have been associated with complications such as capsular contracture, enlargement of lymph nodes draining the implants area, occasionally rupture of the silicone sac and bacterial contaminations [6]. The drying effects of silicone in the surrounding soft tissues enhance the scar formation and this is another possible contributing factor to the formation of the capsule. Physical changes and alterations of the silicone implants as well as use of wetting agent such as poly(vinyl pyrrolidone), have reduced the incidence of capsular contracture.

3.4.2 Polyacrylates

Polyacrylates may be safely bulk polymerized under mild conditions (Scheme 2). In this way it is possible to prepare a casting by in situ polymerization at $40\,°C$. The polymerization of acrylate esters can be carried out readily. The acrylate esters form glassy materials similar to the methacrylates but superior in light transmission. For this reason, methacrylate esters have achieved considerably greater popularity as clear plastic materials. The largest single factor in determining whether a polymerization is successful or not is the purity of the monomer used. It is absolutely essential that the material be pure. This group of polymers is widely used for intraocular lenses, bone cement, dentures and middle ear prostheses.
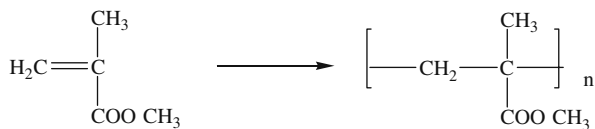
Scheme 2

Polyacrylates have been used in different applications as summarized in Table 3.6.

One of the most popular polymers in this group are cross-linked poly(2-hydroxyethyl methacrylate) (HEMA) hydrogels, invented in the 1950s by the Czech chemist Otto Wichterle. The permeability and hydrophilicity of these gels are dependent on the cross-linking agent. Poly(HEMA) has been used for ophthalmic

Table 3.6 Polyacrylates and their applications

Structure	Applications
Poly(methyl methacrylate)	Intraocular lenses, bone cement, dentures and middle ear prosthesis, orthopedic surgery
Poly(2-hydroxyethyl methacrylate)	Contact lenses, cartilage, matrix in drug delivery system, burn treatment
Poly(2-(dimethylamino)ethyl methacrylate)	Radical scavenging agent, drug delivery devices
Poly(methyl methacrylate)-*co*- (methacrylic acid)	Enteric coating in tablets, gel for encapsulation of biological materials, biofilm on medical implants

R = H or CH_3

products including contact lenses as well as in many drug delivery systems. Macroporous poly(HEMA) gels have been prepared by freeze/thaw or particulate leaching techniques for cartilage replacement. Many different types of molecules and cells have been encapsulated into poly(HEMA) gels and this approach has been reported to be successful for delivery of insulin and other proteins. Poly(HEMA) gels are not degradable in physiological conditions.

Copolymers of MMA with methacrylic acid at a 1:1 ratio are insoluble at acid pH, but not at neutral pH. These copolymers are used as enteric coating of tablets that pass safely through the stomach and dissolve when the tablet reaches neutral pH in the small intestine.

3.4.3 Polyethylene and Related Polymers

Low density branched polyethylene was first prepared in 1933 (Scheme 3) by radical polymerization at high pressure and temperature (>200 °C) without a catalyst. Polyethylene made at high pressure with free radicals will vary in properties from a linear, high density to high branched low density material depending on

polymerization variables. For the preparation of high density polyethylene with a low degree of branching, a Ziegler Natta catalyst should be used. The product is usually obtained as a white powder which can be molded to clear, tough films or extruded to tough fibers. In medical applications, it has been used for disposable product such as tubing, shunts, syringes or packaging materials.

Ultra high molecular weight polyethylene (UHMWPE MW >2,000,000 Da) has been used as acetabular cups (Figs. 3.2 and 3.3) due to its high durability and resistance to abrasion.

$$H_2C = CH_2 \longrightarrow \left[CH_2 - CH_2 \right]_n$$
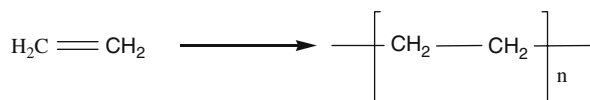
Scheme 3

Polypropylene may be polymerized over the catalyst (transition metallic halide). The polymers are generally high in molecular weight. In normal polymerization of a substituted olefin, a new asymmetric center is produced as each monomer unit is added, but there is no control of the configuration of each succeeding center (Scheme 4). The result is a completely random configuration of the chain. The stereoregularity permits the chains to crystallize; hence the properties of the polymers differ markedly from the random counterpart. It is widely used as sutures catheters, colleting bags, prosthetic valve structures, plasmaphoresis membrane and packaging materials.

$$2HC = CH\overset{\displaystyle CH_3}{|} \longrightarrow \left[CH_2 - CH\overset{\displaystyle CH_3}{|} \right]_n$$
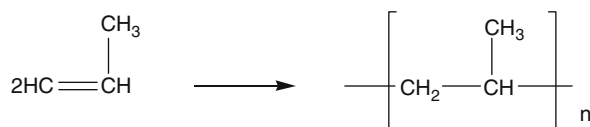
Scheme 4

Polymerization of vinyl chloride was known to be a useful plastic material before World War I (Scheme 5). The polymerization is exothermic and great quantities of heat have to be dissipated through the walls of the polymerization tank to the cooling medium in the jacket. In order to obtain better heat transfer once the polymerization begins, refrigerated brine is circulated in the cooling jacket at about –20 °C. Accurate control of the temperature is necessary since the molecular weight is extremely

$$H_2C = CHCl \longrightarrow \left[CH_2 - CH\underset{\displaystyle Cl}{|} \right]_n$$
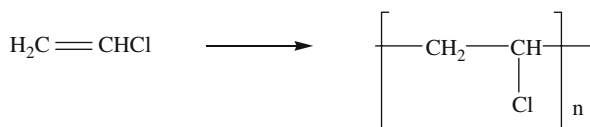
Scheme 5
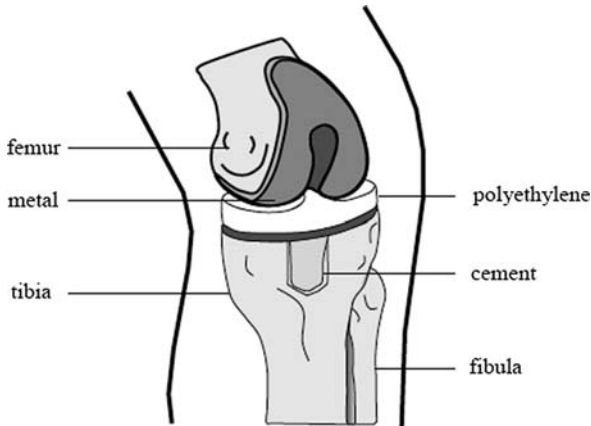
Fig. 3.2 Total knee replacement
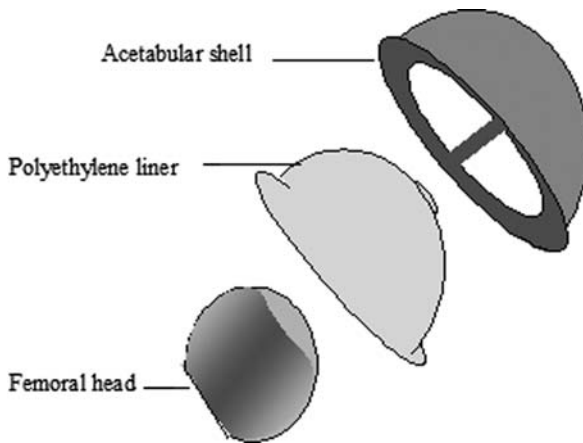
Fig. 3.3 Total hip replacement

sensitive to the variation in temperature. High molecular weight material is obtained with the internal temperature of 48–50 °C. The polymer is converted to dry powder by spraying onto a contoured rotating heated roller. The medical use of PVC is as tubing, plasmaphoresis equipment, and blood bags.

3.4.3.1 Hydrogel Polymers in this Group

Poly(vinyl alcohol) (PVA) is generally obtained from poly(vinyl acetate) by, hydrolysis. The hydrophilicity and solubility of PVA can be controlled by the extent of hydrolysis. PVA forms hydrogels by chemical cross-linking with glutaraldehyde or epichlorohydrin. To avoid the toxicity and leaching of chemical cross-linkings, a repeated freeze/thawing method or electron beam has been applied to form PVA

hydrogels. The gels formed by the repeated freeze/thawing method were reported to be stable at room temperature and highly elastic. These gels are not degradable in physiological situations and are therefore useful as long term or permanent scaffolds. PVA hydrogels have been used in tissue engineering for regeneration of artificial cartilage and bone-like apatite formation.

3.4.4 Polyamides

Polyamides are prepared from acids and amines. Regular high molecular weight polyamides such as polycaprolactam (Nylon 6) and Nylon 6.6 posses high mechanical strength, elasticity and biocompatibility, and resist enzymatic attacks. Adipic acid (6-carbon atom monomer) and hexamethylene diamine salts are prepared first and then the salts are heated at higher temperature to form Nylon 6.6 polyamide [7]. The polymerization reaction takes place at a higher temperature than the melting points of the reactants and the polymer (Scheme 6). In order to polymerize, this

a. *Preparation of Nylon 6.6*:

Adipic acid   Hexamethylenediamine

Hexamethylene diammonium adipate salt
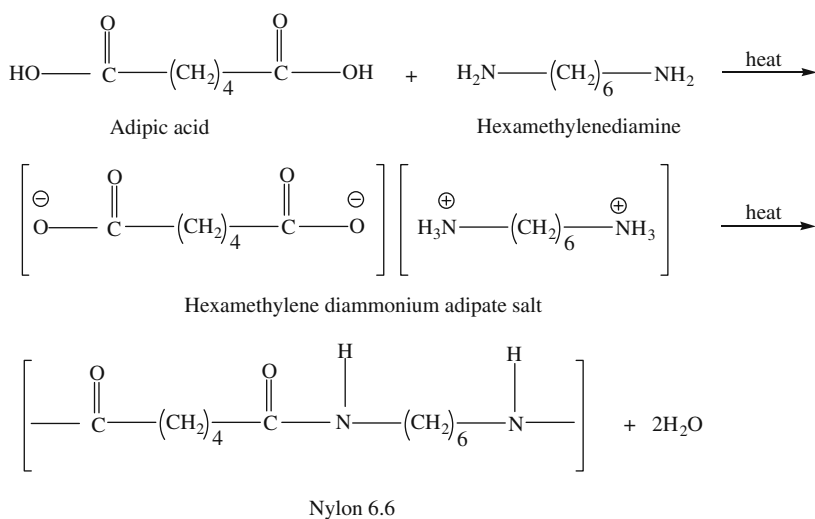
Nylon 6.6

b. *Preparation of Nylon 6*:
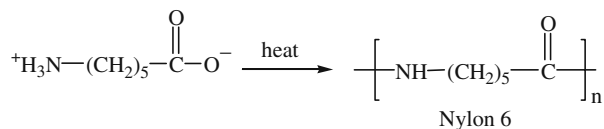
Nylon 6

Scheme 6

mixture is brought to a temperature of 255–265 °C under nitrogen. Nylon 6 is prepared from the 6-carbon monomer – 6-amino hexanoic acid. Nylon 6 is a tough plug with melting point of about 215 °C. It may be fabricated into a tough film by pressing or extruding in the form of filament. Both film and filament may be stretched over a hot plate to give highly oriented crystalline products.

These polymers have applications in surgical sutures and hemodialysis membranes.

3.4.5 Polyurethane and Polyurea

Polyurethanes are prepared by stepwise polymerization reaction, but without the elimination of byproducts such as water or hydrochloric acid [7]. The addition of diol (1,4-butanediol) to a diisocyanate (1,6-hexa diisocyanate) gives a polyurethane (Scheme 7).
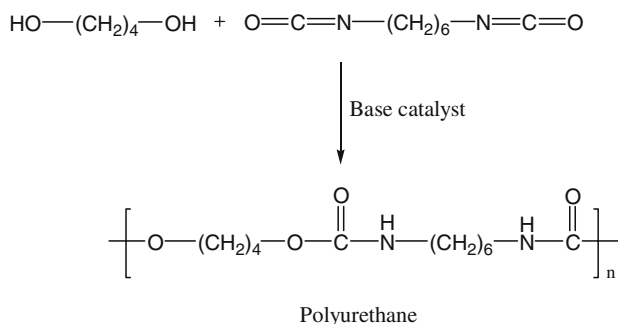
Scheme 7

Similarly, polyurea is prepared from the reaction of a diamine such as hexamethylenediamine with a diisocyanate (Scheme 8). Several applications are found for this group of polymer such as catheters, pacemaker leads, tubing, intra-aortic balloons, wound dressing, artificial heart components, cardiac prosthetic valves (Fig. 3.4), device coatings, etc.
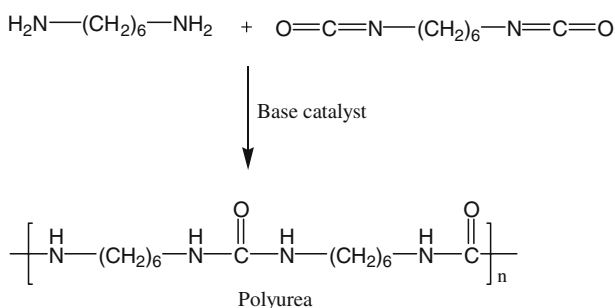
Scheme 8

Fig. 3.4 Artificial heart valve (adapted from
http://imti-itfi.nrc-cnrc.gc.ca)
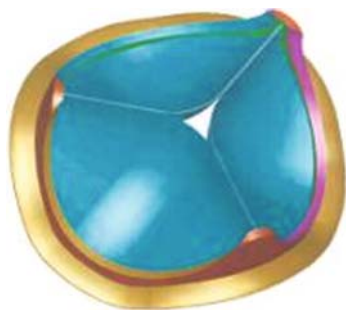
3.4.6 Polyesters

Poly(ethylene terephthalate) is a common polymer prepared by a melt polymeriza-
tion. The following polymerization process is applicable in general to any system in
which the monomers and polymers are thermally stable at temperatures above the
polymer melting point and the glycol is sufficient volatile to permit the excess to
be removed under vacuum. In this preparation there are two ester exchange reac-
tions. The first forms monomer from excess glycol and dimethyl terephthalate,
with elimination of methanol. The second eliminates glycol and forms the polymer
(Scheme 9).
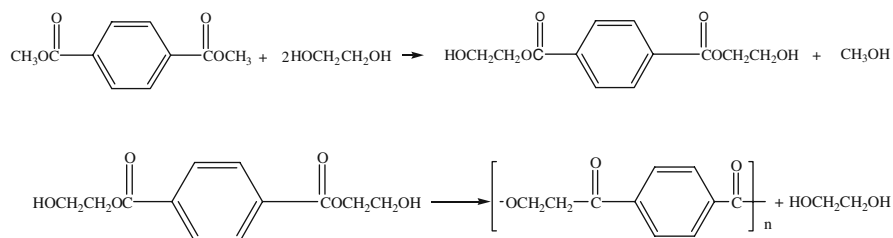
Scheme 9

Dimethyl terephthalate and ethylene glycol at a 1:1 mole ratio heated at 200 °C in
the presence of calcium acetate forms oligomers which at 300 °C gave poly(ethylene
terephthalate) and methanol as byproducts. This polymer is used in synthetic fibers,
films bottles, vascular grafts, tissue patches and shunts.

3.4.7 Polyethers

Poly(ethylene oxide) (PEO) and poly(ethylene glycol) (PEG) have the same poly-
mer structure made of different monomers; one is made from ring opening of ethy-
lene oxide and one from the condensation of ethylene glycol, respectively. PEG has
been approved by the FDA for several medical applications due to its biocompati-

bility and low toxicity. It has been extensively used as excipient in pharmaceutical formulation for oral and injectable administration to stabilize proteins by chemical conjugation of PEG, surface modification of biomaterials and induction of cell membrane fusion. PEG itself is very hydrophilic and can be synthesized by anionic or cationic polymerization of ethylene oxide. PEG gels can be prepared by UV polymerization of the precursor that consists of PEG with acrylate terminal at each end in the presence of α-hydroxy acids. Star-shaped PEG has been cross-linked by irradiation to form hydrogels and modified with galactose moiety to enhance the interaction with liver cells. Various PEO based polymers have been reported and utilized especially in drug delivery. One interesting copolymer is a triblock copolymer of PEO and poly(propylene oxide) (PPO) which is known under the trade name of Pluronics® or Poloxamers® and is available in various lengths and compositions. These polymers form thermally reversible gels without any permanent cross-links. Moreover, PEO-PPO-PEO triblock copolymers can be designed to form gels at body temperature. A few PEO-PPO-PEO copolymers are in clinical use as surfactants and solubilizers in injectable formulations.

3.5 Natural Biopolymer

Biomaterials are mainly from natural origin such as collagens, polysaccharides, fibrins, alginates, chitin and chitosan. etc. Such materials are often complex and difficult to characterize. Various polymers in this group have been studied and utilized to date as follow.

3.5.1 Collagen and Gelatins

Collagen is the most widely used tissue derived natural polymer and it is a main component of extracellular matrices of mammalian tissues including skin, bone, cartilage, tendon and ligament. Physically formed collagen gels are thermally reversible and offer a limited range of mechanical properties. Chemical cross-linking of collagen using glutaraldehyde or diphenylphosphoryl azide can improve the physical properties. However, these gels are still short of physical strength, potentially immunogenic and can be expensive. Furthermore, there can be variations between batches. However, collagen meets many of the biological design parameters, as it is composed of specific combinations of amino acid sequences that are recognized by cells and degraded by enzymes secreted from cells (i.e., collagenase). Collagen has been used as a tissue culture scaffold or artificial skin due to the ready attachment of many different cell types and its cell based degradation. The attachment of cells to collagen can be altered by chemical modification, including the incorporation of fibronectin, chondroitin sulfate or low levels of hyaluronic acid into the collagen matrix. Collagen gels have been utilized for reconstruction of skin, blood vessel and small intestine.

Gelatin is a derivative of collagen, formed by breaking the natural triple helix structure of collagen into single strand molecules. There are two types of gelatin, A and B. Gelatin A is prepared by acidic treatment before thermal denaturation, while gelatin B is processed by alkaline treatment that leads to a high carboxylic content. Gelatin forms gel by changing the temperature of its solution. It has been used in many tissue engineering applications due to its compatibility and ease of gelation. Gelatin gels have also been used for delivery of growth factors to promote vascularization of engineered new tissue. However, the weakness of the gels has been a problem and a number of chemical modifications have been investigated to improve the mechanical properties.

3.5.2 Fibrin

Fibrin has been used as a sealant and an adhesive in surgery as it plays an important role in natural wound healing. Fibrin gels can be produced from the patient's own blood and used as an autologous scaffold for tissue engineering. No toxic degradation or inflammatory reactions are expected from this natural component of the body. Fibrin forms gels from enzymatic polymerization of the fibrinogen at room temperature in the presence of thrombin. An interesting feature of fibrin is the degradation and remodeling by cell associated enzymatic activity during cell migration and wound healing.

Fibrin gels may promote cell migration, proliferation and matrix synthesis through the incorporation of platelet derived growth factors and transforming growth factor. Fibrin gels have also been used to engineer tissues with skeletal muscle cells, smooth muscle cells and chondrocytes. However, fibrin gels are limited in mechanical strength and this prevents their use in certain applications.

3.5.3 Polysaccharide Hydrogels

Celluloses offer attractive drug delivery matrixes (Fig. 3.5). Carboxymethyl-cellulose (CMC) is one example of hydrogel in clinical use. It is prepared by swelling cellulose in sodium hydroxide solution, followed by reaction with monochloroacetic acid. The acid can react with hydroxyl on C2, C3 and C6 on each glucose unit to give a maximum degree of substitution. Presence of ionizable side
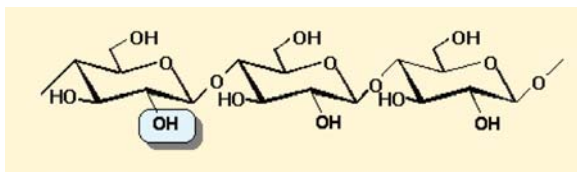
Fig. 3.5 Structure of cellulose

groups disrupts the cellulose crystal structure, hence making CMC water soluble. The polymer forms high viscosity solution. For example, CMC has been combined with hyaluronate to form hydrogel membranes used for prevention of adhesion after intra-abdominal surgery. CMC also forms insoluble ionic complexes with cationic polymers such as chitosan, polyethyleneimine and diethylaminoethyl dextran.

3.5.4 Glycosaminoglycans

Glycosaminoglycans (GAGs) are linear heteropolysaccharides consisting of disaccharide units with the general structure of uronic-amino sugar (Figs. 3.6, 3.7, and 3.8). In their native form, several GAG chains are covalently linked to a central protein core and the protein-polysaccharide conjugates are termed proteoglycans.

There are six different types of GAGs – chondroitin sulfate, dermatan sulfate, keratan sulfate, heparin sulfate, heparin and hyaluronic acid. Heparin and hyaluronate are interesting candidates for development of implantable biomaterials. Heparin is being examined for tissue regeneration and wound healing.

Many of current available hydrogels are based on synthetic non-biodegradable polymers. Hyaluronate has shown excellent potential for tissue engineering applications such as artificial skin, facial intradermal implant and soft tissues augmen-
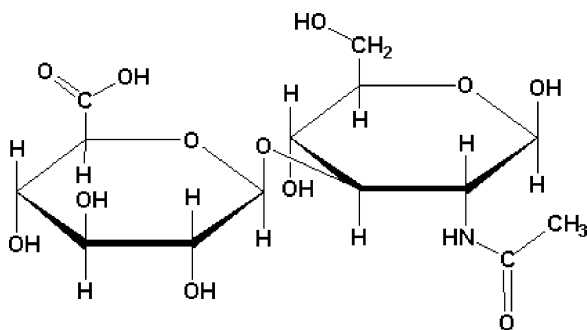
Fig. 3.6 Structure of hyaluronic acid
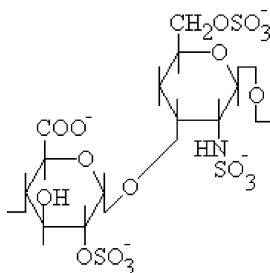
Fig. 3.7 Structure of heparin sulfate

Fig. 3.8 Structure of chondroitin sulfate
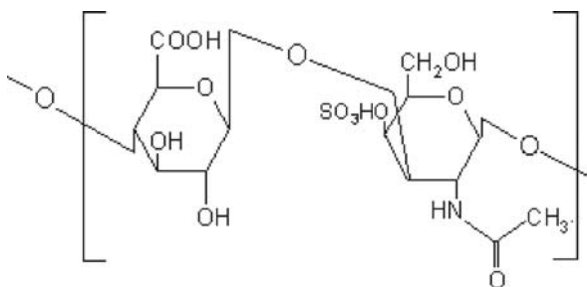
tation. However, hyaluronate requires thorough purification to remove impurities and endotoxins that may potentially transmit diseases. In addition, hyaluronate gels possess low mechanical properties.

3.5.5 Alginates

Alginic acids (Fig. 3.9) or alginates are isolated from several species of brown algae (e.g., *Macrocystis pyrifera*). Alginates are block copolymers of α-L-guluronic acid (G) and β-D-mannuronic acid (M) with average molecular weight of 200,000 to 500,000 (Fig. 3.10). G blocks are linked by α(1→4) and M blocks are linked by β(1→4) glycosidic bonds. Simple gelation will be obtained from adding divalent cation such as Ca^{+2} into a solution of alginates. Alginate have applications as an injectable cell delivery vehicle as well as wound dressing, dental impression and immobilization matrix. Alginate gel beads have also been prepared and used for transplantation of chondrocytes, hepatocytes, and islets of Langerhans. Despite its advantages features, alginate itself may not be an ideal material because it degrades via a process involving loss of divalent ions into the surrounding medium and subsequent dissolution. This process is generally uncontrolled and unpredictable. Therefore, covalent cross-linking with various types of molecules has been attempted to control precisely the mechanical and swelling properties of alginate gels. In addition, molecular weights of many alginates are above the renal clearance threshold of the kidney.
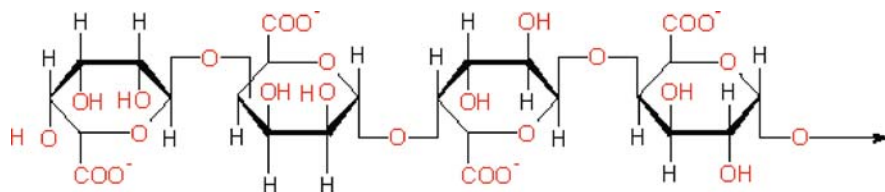
Fig. 3.9 Structure of alginic acid

$$\rightarrow 4ManUA\beta 1 \rightarrow 4ManUA\beta 1 \rightarrow 4GulUA\alpha 1 \rightarrow 4GulUA\alpha 1 \rightarrow$$
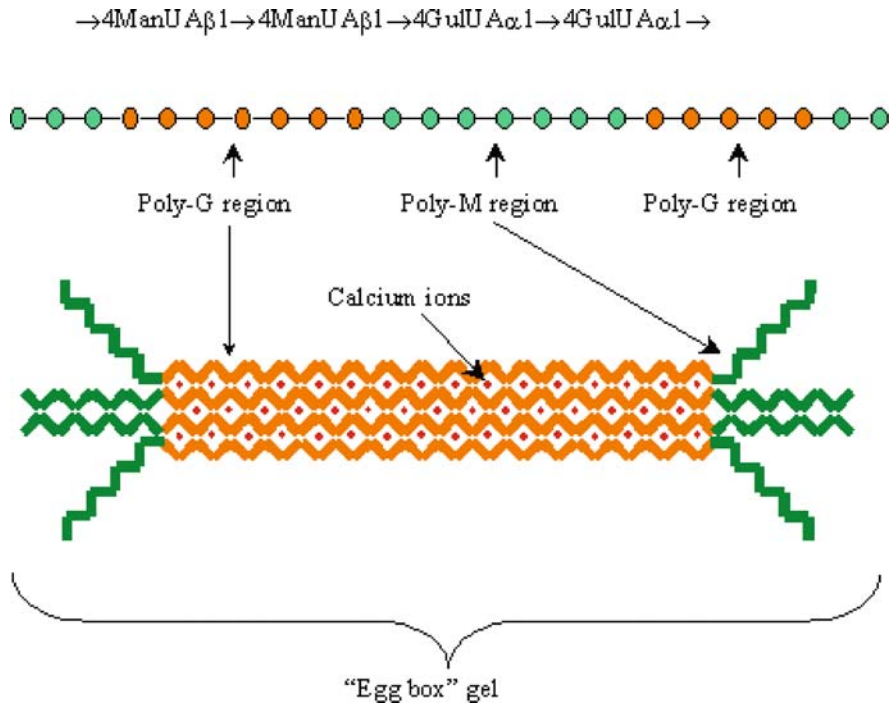
Fig. 3.10 Alginate gel (adapted from www.ifis.co.uk)

The limitation in tissue engineering of using alginate gels is the lack of cellular interaction. Alginate is known to discourage protein adsorption due to its hydrophilic character and it is unable to interact with mammalian cells. Therefore, alginate has been modified with lectin, a carbohydrate specific binding protein, to enhance ligand specific binding properties.

3.5.6 Chitin and Chitosan

Chitosans are partially or fully deacetylated derivatives of chitin (Fig. 3.11), the primary structural polymer in arthropod exoskeletons. Chitosans are the most
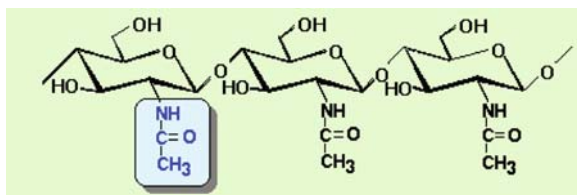
Fig. 3.11 Structure of chitin
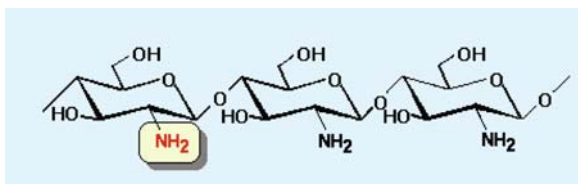
Fig. 3.12 Structure of chitosan

promising polysaccharide biomaterials for development of resorbable and biologically active implants. The primary source of chitin and chitosans is shells from crab, shrimp and lobster. Shells are ground, demineralized with HCl, deproteinized with a protease or dilute NaOH and then deacetylated with concentrated NaOH. Structurally, chitosans are very similar to cellulose, except for their amino or acetylamido substitute. The polymer is linear, consisting of $\beta(1\rightarrow4)$ linked D-glucosamine units with a variable number of randomly located *N*-acetyl-glucosamine groups. The molecular weight of chitosan is in the range of 50,000–1,000,000. Commercially, available preparations carry 10–30% acetylamido residues.

Chitosan is crystalline and its degree of crystallinity depends on the degree of deacetylation (Fig. 3.12). Crystallinity is maximum when the polymer is fully acetylated with varying crystallinity at intermediate deacetylation. Chitosan is normally insoluble in aqueous solution above pH 7 but is readily soluble in dilute acid (pH <5) where free amino groups are protonated.

The high charge density of chitosan means that the linear chains are of semirigid rod conformation. At pH higher than 5, amino groups are increasingly deprotonated and become available for hydrogen bonding, and at some critical pH, the chains in solutions develop enough H bonds to establish a gel network. Gelation pH depends on both degree of deacetylation and average molecular weight. pH dependent solubility of chitosan provides a convenient mechanism for its processing under mild conditions. Viscous solutions can be extruded into higher pH solutions or a nonsolvent such as methanol. The gelled fibers can be subsequently drawn and dried to form high strength fibers. The polymer has been extensively studied for industrial application based on film and fiber formation. Numerous derivatives have been developed to alter biological functions of chitosan, including enhancement of cellular interactions for tissue engineering approaches. Chitosan has been modified with sugar residues such as fructose or galactose for culture hepatocytes and with proteins such as collagen, gelatin and albumin for neural tissue engineering. In addition, methylpyrrolidone derivatized chitosan has been reported to promote bone formation.

3.5.7 Dextran

Dextran is a nontoxic biodegradable polymer and is widely used in many biomedical areas. It is primarily based on $1\rightarrow6$-α-D-glucopyranose and carries an average
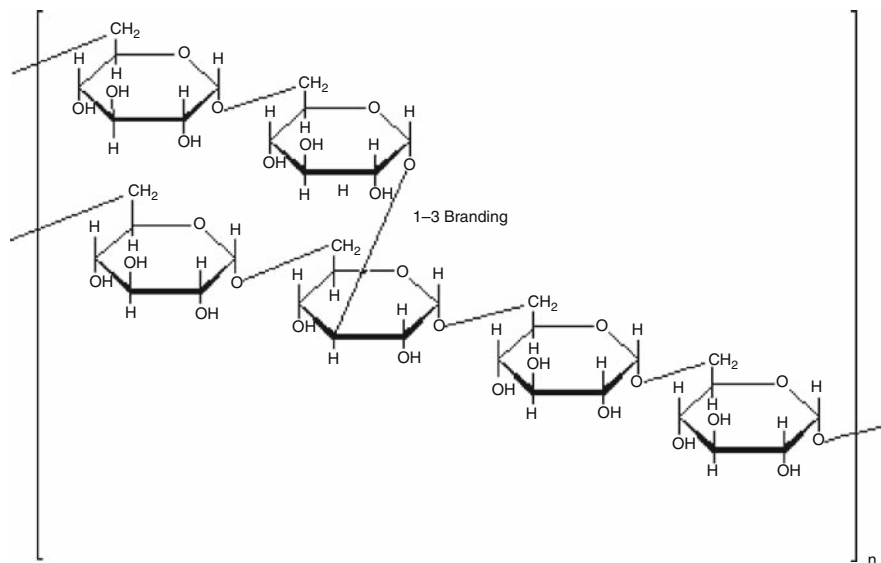
Fig. 3.13 Structure of dextran

of three hydroxy groups per anhydroglucose unit (Fig. 3.13). Solution of Dextran has been used as surgical aids for reducing tissue adhesion. Derivatization of dextran with maleic anhydride gives the dextran-maleic acid half ester, followed by UV crosslinking and formation of dextran-maleic acid hydrogel. Swelling of dextran-maleic hydrogel increased with increasing the degree of maleic substitution over a wide range of pH. Dextrans have been in used for drug conjugation, via reductive amination, to obtain large molecular weight drug derivatives for the purpose of altering drug distribution after intravenous injection and accumulation of the drug conjugates in cancer or in flamed tissues ,where large molecules leak out from broken blood vessels. Also, conjugation of water insoluble drugs to the water soluble dextran provides a water soluble derivative, which can be injected to a patient.

3.6 Bioactive Polymers

Interest continued to grow in polymers which have inherent biological activities or are covalently bound with drugs. This part provides an introduction to the polymeric drug, polymeric drug conjugates, polymeric prodrug and targeted polymeric drug.

3.6.1 Polymeric Drugs

Many synthetic polymers are biologically inert while some exhibit wide ranges of biological activities. They can be categorized into four kinds of polymeric drugs: polycations, polyanions, polynucleotides/polypeptides, and polysaccharides.

3.6.1.1 Polycationic Polymers

These macromolecules have positive charges attached to the polymer chain or as a pendant to the chain. They are active against a number of bacteria or fungi. Recent reports indicate that polycationic polymers can enhance cellular antigen uptake and exhibit antitumor activities. Of the many polycations initially explored as non-viral vectors, the polyethyleneimine (PEI) have been most widely studied. PEI promotes endosomal escape via the proton sponge mechanism; linear polymers of MW 22,000 are able to overcome the nuclear barrier and also yield the highest transfection efficiency. Recent promising results obtained using the PEI-polyplexes in vivo in AIDS and cancer are moving forward into clinical evaluation. PEI has the limitation of relatively high toxicity and this could prove problematic for repeated systemic rather than local administration. PEI derivatives have been found to have a strong antimicrobial activity as well as anti-prion and viral activities.

3.6.1.2 Polyanionic Polymers

Polymers with negative charges exhibit various biological activities. They can enter biological functions by distribution throughout the host and behave like proteins, glycoprotein and polynucleotides which modulate a number of biological responses related to the host defense mechanism. These are enhanced immune responses and activation of the reticuloendothelial system macrophages. Examples of polyanions are pentosane sulfate, dextran sulfate, poly(maleic acid) copolymers and poly(vinyl amino acids).

3.6.1.3 Polynucleotides/Polypeptides

Polynucleotides are potent interferon inducers. A mismatched double stranded synthetic polyribonucleotide ampligen and the double stranded acids, polyadenylic-polyuridylic acid and polyinosinic polycytidylic acids, have been studied for cancer therapy. Although these materials elicit excellent activity with rodents, therapeutic effects are dramatically decreased within primates.

3.6.1.4 Polysaccharides

Sulfation of dextrin – a polysaccharide routinely used for peritoneal dialysis in patients with end stage renal failure at the 2 (or 6) position produces a polymer that blocks infection of T cell lines by adapted strains of HIV-1. Dextrin-2-sulfate (MW 25,000 g/mol) given to patients intraperitoneally daily for 28 days was well

tolerated up to the maximal daily dose of 150 mg and in phase III clinical trials it reduced replication of HIV-1 in patients with AIDS. Coincidentally, dextrin 2 (and 6) sulfate inhibits morphological differentiation of endothelial cells into tubes; the reduced lesions is probably the result of anti-angiogenic effect. In gel form, dextrin sulfate is now approved for an intravaginal virucide.

3.6.2 Polymeric Drug Conjugates/Polymeric Protein Conjugates

The drug is covalently bonded to an appropriate polymer carrier. These large molecules diffuse more slowly and are adsorbed at pharmacological interfaces. Therefore, drug-polymer conjugates can prolong therapy. The major attributes of polymeric drug carriers are their depot effects, unique pharmacokinetics and pharmacological efficacy. A model for polymeric drug-conjugates is shown in Fig. 3.14 [8].

It can be seen that four different groups are attached to the polymer backbone. One group is the drug, the second is a spacing group, the third is a transport system and the last is the group to solubilize the entire polymer system. The drug can be attached permanently by a stable bond or temporarily and removed by hydrolysis or enzymatic processes. The transport system can be made specific for tissue cells. Solubilizing groups are added to increase the hydrophilicity and solubility of the whole macromolecular system in aqueous system while non-polar groups enhance the hydrophobic character in lipid regions. The specificity of polymer is related to molecular size which alters the transport rate across compartmental barrier.

By controlling the weight of the polymer carrier of a polymer drug carrier, it is possible to regulate whether the drug passes through the blood membrane barrier or accumulates in some organs. The macromolecular transport theory of biopolymers through tissues has been successfully applied to the design, fabrication, and prediction of in vivo performance of polymeric drug conjugate systems.
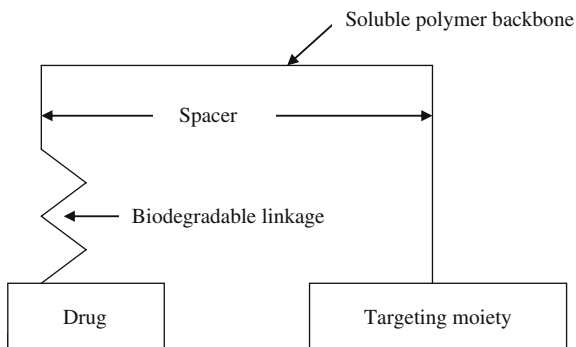
Fig. 3.14 Model for polymer drug carrier
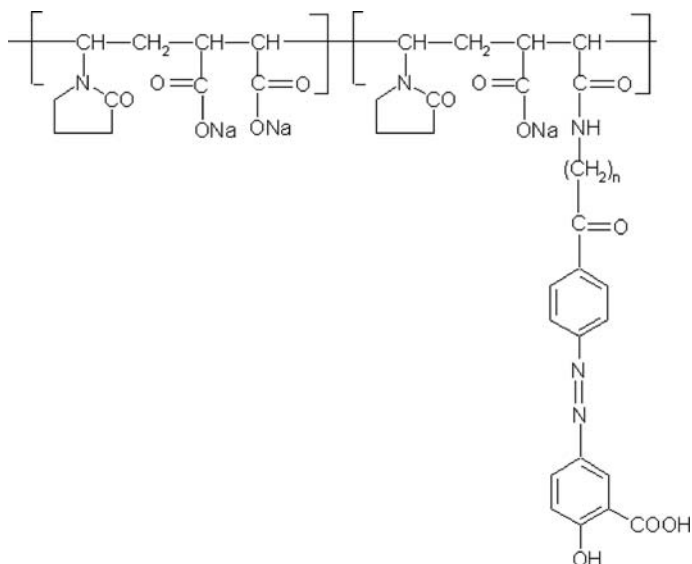
Fig. 3.15 Polymeric prodrug containing 5-ASA conjugate covalently linked to poly(methyl vinyl ether/*co*-maleic anhydride) and poly(1-vinyl-2-pyrrolidone *co*-maleic anhydride)

3.6.3 Polymeric Prodrugs

Polymeric prodrugs are designed to protect against rapid elimination or metabolism by adding a protective polymer to the therapeutic materials (Fig. 3.15). Therapeutic activity is usually lost with this attachment and reinstated with the removal of the protective group. The protective group is designed to be easily removed usually by hydrolysis.

A drug which is active only after being cleaved from the polymer chain is called a prodrug. The drug is usually attached away from the main polymer chain and other pendant groups by means of a spacer moiety that allows for more efficient hydrolysis.

3.6.4 Targeted Polymeric Drug

Polymeric drug targeting to a specific site is an enormous advantage in drug delivery because only those sites involved are affected by the drug. Ideally, a targetable drug carrier is captured by the target cell to achieve optimum drug delivery while minimizing the exposure to the host. Most administered macromolecules are eliminated by the host before any significant cellular uptake takes place. If the macromolecules contain a moiety that is compatible with a receptor on a specific cell surface, then the molecule is attracted to the cell surface and the uptake is enhanced. This maximizes the opportunity for specific cell capture. This type of cell specific targeting has been developed.

Table 3.7 Examples of biodegradable polymers

Polyesters
    Polylactic acid
    Polyglycolic acid
    Polycaprolactone
    Polyhydroxybutyrate
    Polyhydroxyvalerate
Polyorthoesters
Polycarbonates
Polyanhydrides
Polyphosphate esters
Polyphosphazenes

3.7 Biodegradable Polymers

The term "biodegradable polymers" denotes water insoluble polymers which, by means of a chemical reaction in the body, are converted slowly to water soluble materials. The polymer can have a side chain that undergoes hydrolysis in the body to produce hydroxyl, carboxyl or other hydrating groups. These groups make the polymer fragments and degradation products water soluble. Another approach is to crosslink a water soluble polymer with a hydrolyzable cross-linking agent. Once crosslinked, the polymer is insoluble. When placed in the body, the crosslinking group is hydrolyzed or degraded to give a water soluble polymer. Water insoluble polymers which contain hydrolyzable functional groups directly in the polymer chain are the most frequently used. As these groups in the chain are hydrolyzed, the polymer chain is slowly reduced to shorter and shorter chain segments which eventually become water soluble. The main advantage of the latter group of polymer is that polymer will have good mechanical properties. Table 3.7 lists examples of these biodegradable polymers.

3.7.1 Polyesters

Polyesters are synthesized by condensation polymerization of dicarboxylic acids and diols, or hydroxy alkyl carboxylic acids. Examples for both polymerizations are aliphatic polyester based on monomers other than α-hydroxy-alkanoic acids which have been developed and evaluated (Scheme 10). These include the polyhydroxybutyrate and the polyhydroxyvalerate developed by ICI from a fermentation process and the polycaprolactone. The homopolymers in these series are hydrophobic and crystalline in structure, and therefore they have long degradation times in vivo (1–2 years). However, the use of copolymer in the case of polycaprolactone have led to materials with a shorter degradation time as a result of changes in the crystallinity and hydrophobicity of these polymers.

$$HO-(CH_2)_x-COOH \xrightarrow{H^+/ \text{ heat}} \left[O-(CH_2)_x-\overset{\overset{\displaystyle O}{\|}}{C}\right]_n + H_2O$$
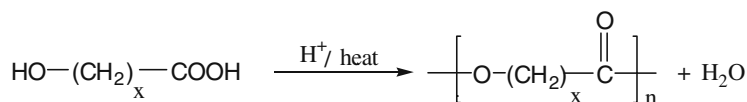
x=1-poly(glycolic acid); x=3-poly(butyric acid); x=5-poly(caprolactone).

$$HO-(CH_2)_x-OH \quad + \quad HOOC-R-COOH$$

$$\Big\downarrow H^+ \text{ heat}$$

$$\left[O-(CH_2)_x-O-\underset{\underset{\displaystyle O}{\|}}{C}-R-\underset{\underset{\displaystyle O}{\|}}{C}\right]_n \quad + \quad H_2O$$
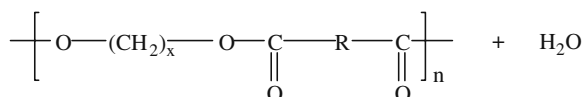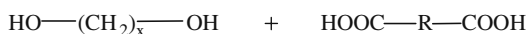
x = 2, R= (CH$_2$)$_2$-poly(ethyl succinate); x = 2, R=(CH=CH)-poly(ethylene)

Scheme 10

The ω-hydroxycaproic acid on condensation gave polycaprolactone. At every condensation reaction the polymer chain grows but remains an ω-hydroxy carboxylic acid; hence it further reacts to form a longer polymer chain. For example poly(lactic acid) upon overheating converts into lactide (Scheme 11).
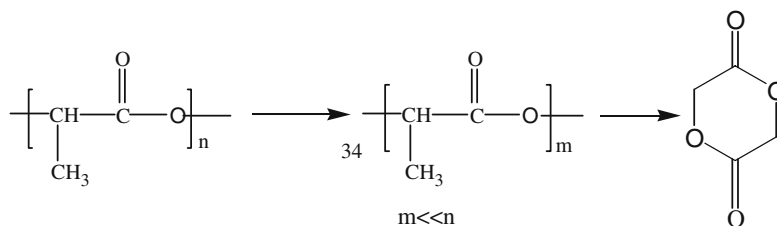
Scheme 11

Polyesters can also be polymerized by stepwise polymerization and ring opening polymerization as shown in Scheme 12. One of the most versatile and widely used synthesized polymers are aliphatic polyesters prepared from lactic and glycolic acids. These polymers were first utilized as sutures and orthopedic plates and nails, and their biocompatibility and biodegradability are well known. Moreover, the commercial availability of these polymers and the accurate biodegradation rates has made these biodegradable polymers the first choice of medical devices. The applications are also found in controlled release and tissue engineering.

1. Stepwise polymerization:

Poly(lactic acid)
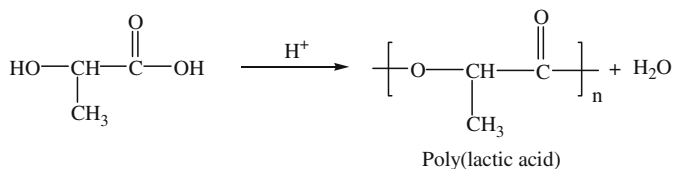
2. Ring-opening polymerization:

Poly(lactide)
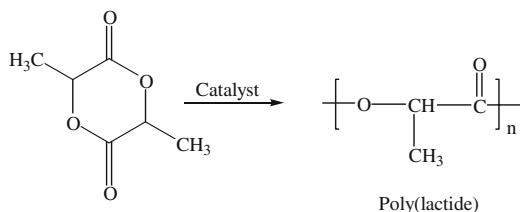
Scheme 12

3.7.2 Poly(ortho esters)

Poly(ortho ester) can be synthesized by the reaction of ketene acetal and an alcohol, to form the orthoester bond (-O-COO-) between the polymer units (Scheme 13). For example, 3,9-bis(methylene) 2,4,8,10-tetraoxaspiro[5,5] undecane on condensation with 1,6-hexanediol forms a poly(ortho ester) (Scheme 6). The reaction proceeds at room temperature using tetrahydrofuran as solvent and acid catalyst. The preparation of poly(ortho esters) from diol and diketene acetals is similar to the preparation of polyurethanes from diol and diisocyanate (see previously). Both reactions proceed without the formation of a byproduct. Polyorthoesters contain hydrophobic units linked together along the polymer chain by functional groups susceptible to hydrolysis. Exposure of the outer layer of polymer to water causes hydrolysis of the
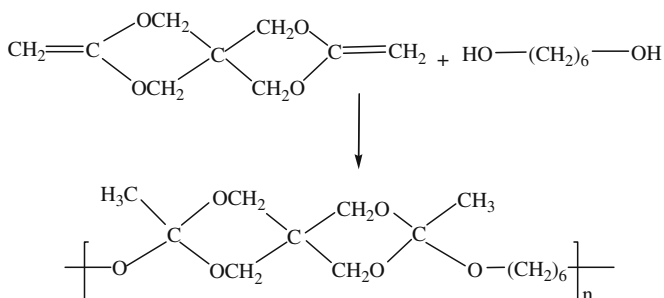
Scheme 13

water labile linkages and subsequent degradation and loss of the outer layer of the polymer. As the outer layer is lost, another layer of polymer is exposed to water and the process continues such that polymers are slowly eroded from the surface. Such polymers are termed bioerodible and the rate of hydrolytic degradation at the surface of the polymer is much faster than the rate of water penetration into the polymer matrix. Poly(ortho esters) have been tested as carriers of drugs like insulin and antibiotics [9].

3.7.3 Polycarbonates

Biodegradable polycarbonates are synthesized by the condensation of an aliphatic diol with the derivatives of carbonic acid such as phosgene or biphenyl carbonate. Phosgene, on bubbling into a solution of a diol in pyridine at 20–35 °C, forms a polymer [10] which is isolated by precipitation in water or methanol (Scheme 14).
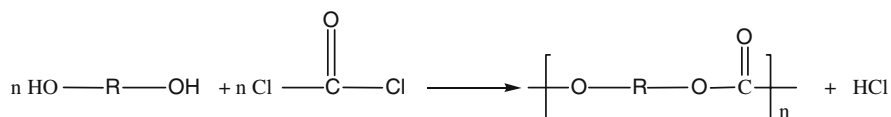
Scheme 14

3.7.4 Polyanhydrides

Polyanhydrides are synthesized by stepwise polymerization from diacid monomers using a dehydrative agent or from the reaction of a diacid chloride and a dicarboxylic acid (Scheme 15). This hydrophobic monomer leads to a polymer with low water permeation. However, the anhydride linkage is quite susceptible to hydrolysis and experiments have demonstrated that surface erosion of these polymers does occur. Polyanhydrides have been used for controlled delivery of anticancer agents to the brain for treating brain tumors, which is now FDA approved [11].
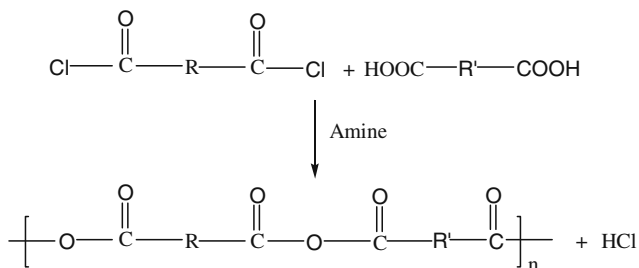
Scheme 15

3.7.5 Poly(phosphate ester)

Poly(phosphate ester) can be synthesized by the interfacial condensation of either ethyl or phenylphosphorodichloridates and various dialcohols under phase transfer conditions (Scheme 16). The polycondensation was found to be dependent on the catalyst concentration. Polyphosphate ester copolymers have been tested as drug carriers [12].
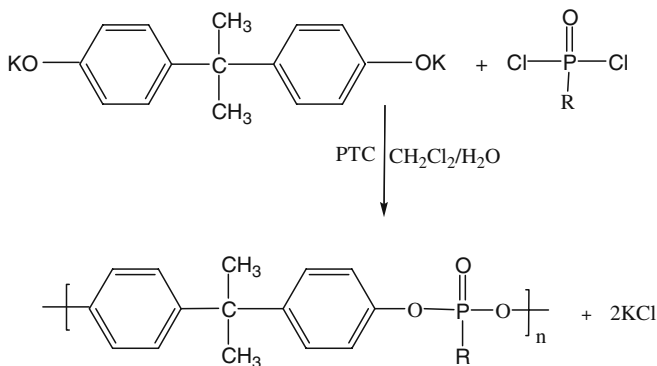
PTC = Phase Transfer Catalyst

Scheme 16

3.7.6 Poly(phosphazenes)

Poly(phosphazenes) are polymers consisting of an inorganic backbone of alternate nitrogen and phosphorus atoms linked by alternating single and double bonds. The different polymers can be prepared by nucleophilic displacement reactions.

Poly[(dichloro)phosphazene] is prepared by condensation reaction. Monomeric phosphoranimine on condensation forms N-dichlorophosphoryl-P-trichloromono phosphazene [$Cl_2P(O)N=PCl_3$] with the elimination of phosphoryl trichloride $POCl_3$ (Scheme 17). Poly(phosphazenes) were also prepared by ring-opening polymerization methods. Different polyphosphazenes has been used as inert biomaterials for cardiovascular and dental applications. Bioerodible and water soluble polyphosphazenes are also used for controlled drug delivery [12].

The degradation of polyphosphazene can be controlled by changes in side chain structure rather than the polymer backbone. A hydrophilic backbone as well as the structure versatility as a result of various substitutions offers possibilities in designing new classes of polyphosphazene gels. Various modifications of polyphosphazenes have been reported including poly(aryl/alkyl) phosphazenes, poly(amino acid-ester) phosphazenes and methoxy-poly(ethylene glycol) substituted polyphosphazenes with temperature responsive features. Two types of hydrogels
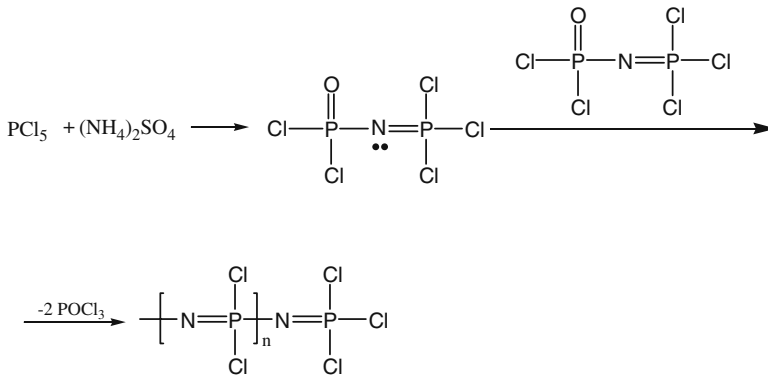
Scheme 17

exist – non-ionic and ionic. Non-ionic polyphosphazene gels are based on water soluble phosphazene containing glucosyl or glyceryl side groups. Ionic polyphosphazene hydrogels are formed with divalent ions. These polymers may be useful for skeletal tissue regeneration or encapsulation of hybridoma cells.

3.8 Characterization of Biomaterials

The understanding of the interactions which take place between a material surface and the compositions of the biological systems is an important requirement of biomaterial development (Table 3.8). The uppermost layers of a biomaterial which present certain chemical and physical properties have to be defined as it is a contact surface. Also, biological parameters such as adsorption of protein, cell growth and blood compatibility need to be established.

These properties are essential for biomaterials, although perhaps not required in industrial applications. This is necessary as the living cells are very vulnerable to and readily killed by physical and chemical stimuli, for instance, by toxic substance invasion. Toxic compounds associated with polymers are listed in Table 3.9.

Table 3.8 Minimum requirements for biomaterials [13]

Requirements	Examples
I. Non-toxicity	Non-pyrogenic, non-hemolytic, non-inflammatory (chronic), non-tumorigenic, non-allergic
II. Functionality	Organ and tissue replacement, tissue reconstruction, internal organ support, disposable medical devices, drug delivery
III. Sterilizability	Radiation, ethylene oxide gas, autoclave, dry heat
IV. Biocompatibility	Mechanical and interfacial

Table 3.9 Toxic compounds related to polymers

Components	Remarks
I. Monomers	Not polymerized, depolarized, oligomers
II. Initiators	Intact and decomposed
III. Catalysts	Crosslinking, curing and other reactions
IV. Additives	Antioxidants, plasticizers, UV absorbants, lubricants, antistatic agents, dyes, pigments
V. Others	Byproducts, degradation products

Table 3.10 The features of biocompatibility [14]

Materials
    Bulk chemistry
    Surface chemistry
    Surface roughness
    Surface energy
    Surface charges
    Chemical stability
    Chemistry of degradation products
    Physical characteristics of degradation products
Devices
    Size
    Shape
    Elastic modulus/rigidity
Host
    Species (animal study)
    Tissue type and location
    Age
    Sex
    General health status
    Pharmaceutical regimens
Systems
    Operative technique
    Implant tissue attachment
    Infection

In general, we would expect the features shown in Table 3.10 to be influenced in determining the specific characteristics of biocompatibility that will be seen under different circumstances. It is important to note that it is not only the characteristics of the materials and device that control events but also host variables as we get older; repair processes are less efficient and the general state of health will influence defense mechanisms.

3.8.1 Chemical Properties on the Surfaces

The compositions of biomaterial surface can be measured by different methods such as attenuated total reflectance (ATR-IR), X-ray photoelectron spectroscopy (XPS)

and secondary ion mass spectroscopy (SIMS) [15]. ATR-IR supplies the absorption spectra of functional groups with an informational depth of 0.1–10 μm. XPS is a more sensitive surface analytical method which gives information not only about the type and amount of elements present but also their oxidation state and chemical surroundings. In SIMS, primary ions interact with the polymer surface and the mass spectra of the secondary ions are obtained which also give the chemical compositions of the outermost atomic layer (1 nm thickness).

3.8.2 Physical Properties of the Surfaces

Several methods are available to determine the physical parameters of polymer surfaces. Biomaterials penetrate liquids like blood or water present in soft tissue. It is known that the surface free energy at the biomaterial/water interface is the driving force for the reorientation processes of the polar groups of the uppermost molecular layers of the polymer surface. The investigation of surface wettability by means of contact angle determination is of special interest in characterization of a surface.

Electrochemical properties are also physical surface parameters. The existing surface charge density, i.e., surface potential, has a strong influence on protein adsorption and blood compatibility. Table 3.11 shows surface analysis use in characterization of biomaterials.

Table 3.11 Summary of surface analysis techniques in biomaterials [16]

Chemical compositions
Electron spectroscopy for chemical analysis
Auger electron spectroscopy
Ion scattering spectroscopy
Backscattering spectroscopy
Secondary ion on mass spectrometry
Surface extended X-ray absorption fine structure
Thermodynamic analysis
Wettability
Sorption-desorption
SEM
TEM
Scanning tunneling microscopy
Atomic force microscopy
Infrared spectroscopy
Electron energy loss spectroscopy
Thin film X-ray diffraction
Electron diffraction
Low energy electron diffraction

3.8.3 Adsorbed and Immobilized Protein Determination

ELISA was employed to detect the adsorption of a protein at a biomaterial surface based on a specific reaction of antibodies with surface bound antigens [17]. The detection of binding capacity of antibodies is useful to establish the amount of adsorbed and immobilized fibronectin. When a large amount of adsorbed fibronectin exists at the polymer surface, enhanced growth of the cells, e.g., endothelial cells, can be expected. Low fibrinogen adsorption may indicate good blood compatibility.

3.8.4 In Vitro Cell Growth

Cell culture permits the determination of cell growth influencing factors and thereby the selection of cell toxic materials. In addition, it allows the investigation of tissue compatibility of biomaterials. In cell culture two cell types can be used; organ specific cells and cells belonging to permanent cell lines which have lost their specificity. The evaluation is done by measurement of cell density after a period of time.

3.8.5 Blood Compatibility

The contact of biomaterial surfaces with the blood provokes activation of the intrinsic coagulation pathway at blood/biomaterial interface. The following coagulation parameters obtained after the contact are evaluated as compared with citrate plasma values: activated partial thromboplastin time (APTT), partial thromboplastin time (PTT), platelet adhesion, and leucocyte number.

   The most important parameters to characterize blood compatibility are thrombocyte adhesion and thrombocyte number. Materials which show strong platelet adhe-

Table 3.12 Interactions between living systems and biomaterials
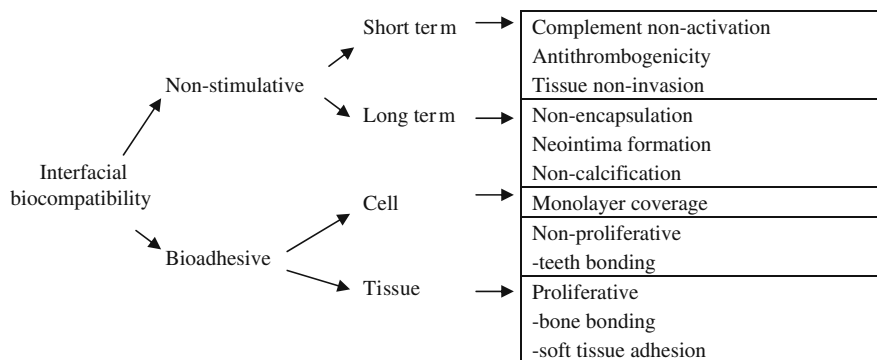
sion as a consequence of their contact with a foreign body or provoke a decrease in the number of blood platelets are considered as thrombogenic [17]. The decrease in the quantity of blood leukocytes after contact is a sign of cellular immunoresponse of the biological system. When a biomaterial comes in contact with a living body, interactions occur between the foreign material and the living body, as shown in Table 3.12.

3.9 Fabrication Technology

Thermoplastic resins are fabricated in the molten state by extrusion into films or fibers and by molding into three-dimensional objects. Two processes are employed, namely extrusion and injection molding.

3.9.1 Extrusion

A single screw extruder is a conveyer similar to a meat grinder. It consists of a screw, driven by a motor connected to its shaft through a gear reducer, rotating inside a cylindrical barrel. The rotating screw moves the resin pellets forward and generates by shear most of the heat required to melt pellets, as well as the hydrostatic pressure to force the molten plastic through the die (Fig. 3.16). The size of the extruder is described by the inside diameter of the barrel. The screw consists of three
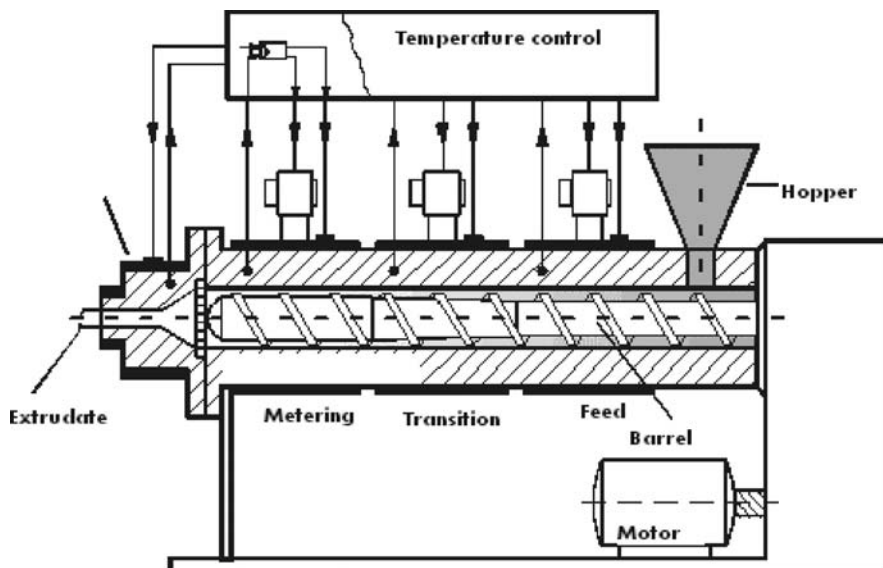
Fig. 3.16 Schematic drawing of a single screw extruder (adapted from www-rt.uni-paderborn.de/)

zones, namely a feed section, a transition section, and a metering section. The screw usually has the same pitch or helical angle in the three sections. The process variables, temperature and speed of revolution, do not afford a range of conditions wide enough for the effective extrusion of different polymers. For example, polyamides have lower melt viscosities than polyethylene resins, while polyvinyl chloride resins degrade readily at high temperature. Therefore, different screws are available for a single extruder, having different lengths of the three zones and different compression ratios. Resin pellets, granules or flakes are fed from the hopper into the feed section of the extruder, from where they were conveyed to the transition section. Here the pellets are compressed and melted. A large portion of the heat required to melt the resin is generated by viscous friction as the pellets are sheared between the rotating screw and the stationary wall of the barrel. Another portion of the heat is supplied externally through the barrel, usually by electric band heaters mounted on the barrel. As the resin advances through the transition zones, it is plasticized, melted, and mixed. By the time it reaches the metering section, it is a homogenous, very viscous liquid. The metering section of the screw has the shallowest channel depth. It pumps the melt through a screen-pack filter into the die cavity. The filter removes solid impurities and lumps of unmelted resin. If the extruded film is to have a uniform thickness, it is essential that the melt exits through the die slit at a constant flow rate, free of sudden surges. The metering section ensures a constant delivery rate. Flat film is extruded downward through a die with a long slit for an opening onto highly polished chilled rolls that are water cooled. From there, the sheet is rolled up on a windup roll. Tubular film is produced by extruding the melt upwards through an annular die around a mandrel (Fig. 3.17). As the tube is pulled upwards, it is blown up to a bubble by air injected through the mandrel, stretched and biaxially oriented. The hot tubular film is cooled by air issuing through the holes of a hollow ring surrounding the tube near the point where it leaves the die, below the zone where it is inflated to a bubble. As the inflated, solidifying film moves
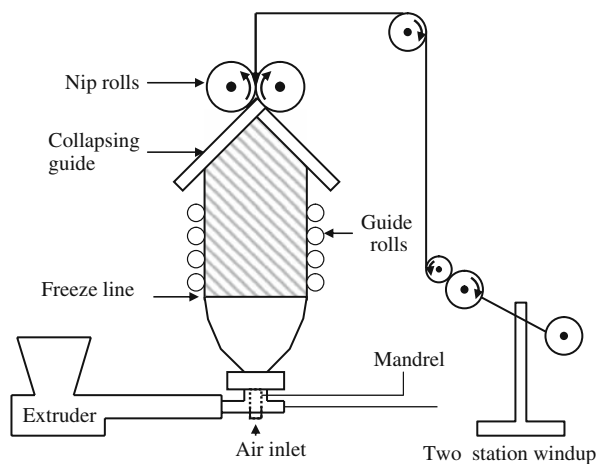
Fig. 3.17 Schematic drawing of tubular film extrusion

upwards, it is gradually deformed into the layflat form by the action of guide rolls. A pair of rubber-covered nip rolls collapses the film completely and thereby seals in the inflating air that expanded the molten film issuing from the die into a bubble. A windup roll then rolls the flattened film up, pulling it upwards continuously away from the die.

3.9.2 Injection Molding

Of the various processes, injection molding is the most widely used. The following three operations are carried out successively. The thermoplastic resin, in the form of pellets, is heated, melted, and pushed into the die cavity, which is filled with the melt. The molten plastic cools and solidifies in the mold while under pressure. Finally, the mold is opened and the part is ejected. Molds may have several cavities for the simultaneous molding of several parts. Two types of machines are used to melt and inject the resin into the mold, a plunger injection molding machine and a reciprocating screw injection molding machine. The plunger molding machine is fitted with a hydraulic arm that compresses the resin pellets at the same time that they are heated and melted. The molten resin is pushed through a nozzle into the mold cavities and cooled under pressure to below its melting or glass transition temperature. The molten resin shrinks on cooling, the mold is opened, and the solidified plastic parts are ejected.

The second molding machine, equipped with a reciprocating screw, mixes and homogenizes the melting resin. The screw resembles that of the single screw extruder depicted in Fig. 3.17, except that it is moved forward and backward in the barrel by a hydraulic mechanism. The resin is fed from a hopper into the barrel, plasticized and melted by the rotating screw. The screw, then acting as a plunger, forces the melt into the mold cavities, where the plastic cools and solidifies. The molded plastic is then ejected from the mold cavities. The molten resin shrinks on cooling and solidifying because its density decreases with rising temperature. Amorphous plastics shrink far less on solidifying than semi-crystalline plastics owing to crystallization of the latter. Molds are filled under high pressure. By compressing the melt, more material is made to flow into the mold cavity, reducing the shrinkage on cooling. The average linear shrinkage of polystyrene on cooling from $\sim$100 °C to room temperature ranges from 3.5% at atmospheric pressure to 1.4% at 10,000 PSI. The corresponding linear shrinkage of polyethylene is 7% and 5%, respectively. These values are higher because polyethylene is semi-crystalline whereas polystyrene is amorphous.

3.10 Future Trends in Biomedical Uses of Biopolymers

This chapter illustrates some of the polymers use in medical sciences. In the past, medical sciences used polymers in much smaller amounts than other industries such as textile, packaging and tire industries. Therefore, the manufactures of plastics and

elastomers found no economic incentive to tailor make polymers for the special needs of industry, nor to seek clearance from the FDA for the application of existing polymers.

An example of the latter situation is methyl cyanoacrylate, which has been used in an industrial glue for many years. It has been employed successfully to repair incisions or tears in a variety of tissues and organs.

Presently, medical scientists, commercial laboratories, and some pharmaceutical companies design, prepare, manufacture and process or fabricate their own polymers in accordance with their specialized requirement. In addition to sutures, implants of plastics and elastomers in human bodies are widely used for repair or replacement of tissues, organs or parts of organs. Some of the problems encountered with implants are degeneration of adjoining tissues and formation of blood clots and thrombi on the surfaces of implant synthetic polymers, enlargement of implants through sorption of lipids or moisture, which constitutes a problem with precision parts and corrosion, and weakening of implants due to the interaction with enzymes or other physiologic compounds and/or mechanical stresses. A skin substitute to cover burns has recently been introduced. It is a composite made of an outer layer of silicone elastomer to protect the wound from infection and dehydration, bonded to a porous bottom layer that is adhesive and biodegradable consisting of cross-linked collagen combined with glycosaminoglycan. After the skin has grown back underneath this composite film, the bottom layer is biodegraded and the elastomeric top layer sloughed off. Considerable research is in progress to achieve biocompatibility of synthetic polymers through surface modification to prevent blood clotting and to search for strong and more durable polymers.

References

1. Brandup J and Immergut EH. Polymer Handbook, A Wiley-Interscience: New York, 1989.
2. Halpern BD and Tong YC. Medical applications, in Kroschwitz JI, ed. Polymers: Biomaterials and Medical Applications, John Wiley and Sons: New York, 1989.
3. Cast DG and Ratner BD. Biomedical surface science: Foundations to frontier. Surf Sci, 2002, 500, 28–60.
4. Migliaresi C and Pegoretti A. Fundamentals of polymeric composite materials, in Barbucci R, ed. Integrated Biomaterials Science, Kluwer Academic: New York, 2002.
5. El-zaim HS and Heggers JP. Silicones for pharmaceutical and biomedical applications, in Dumitriu S, ed. Polymeric and Biomaterials, 2nd edn, Marcel Dekker, Inc.: New York, 2002.
6. Cook RR, Harrison MC, and Levier RR. Arthritis & Rheumatism, 1994, 37, 153.
7. Ebewele RO. Polymer Science and Technology, CRC: Boca Raton, FL, 2000.
8. Ringdorf H. Macromelecular reviews. J Polym Sci, 1975, 10(1): 1–230.
9. Heller J. Polyortho esters, in Domb AJ, Kost J, Wiseman DM, eds. Handbook of Biodegradable Polymers, Harwood. Academic: Amsterdam, 1997.
10. Legrend DG and Bendler T. Handbook of Polycarbonate Science and Technology, Marcel Dekker: New York, 2000.
11. Domb AJ, Elmalak O, Shastri VR, Ta-Shma Z, Masters DM, Ringer I, Teomin D, Langer R. Biodegradable phosphazenes for biomedical applications, in Domb AJ, Kost J, and Wiseman DM, eds. Handbook of Biodegradable Polymers, Harwood Academic Publishers: Amsterdam, 1997.

12. Vandorpe J, Schacht E, Dejardin S, and Lemmouchi Y. in Domb AJ, Kost J, and Wiseman DM, eds. Handbook of Biodegradable Polymers, Harwood: Amsterdam, 1997.
13. Ikada Y. Interfacial biocompatibility, in Shalaby SW, Ikada Y, Langer R, and Williams J, Polymers of biological and biomedical significances, American Chemical Society: Washington DC, 1994.
14. Williams DF. Biofunctionality and biocompatibility, in Cahn RW, Haasen P, Kramer EJ, eds. Materials Science and Technology, VCH publishings: New York, 1992.
15. Klee D and Hocker H. Biomedical applications polymer blends. Adv Polym Sci, 1999, 149: 1.
16. Lyman D. Characterization of biomaterials, in Barbucci R, ed. Integrated biomaterials Science, Kluwer Academic: New York, 2002.
17. Anderson JM and Kottke-Marchant, K. CRC Critical Reviews in Biocompatibility, 1985, 1: 111–120.

Chapter 4
Biomaterials: Processing, Characterization, and Applications

Damien Lacroix and Josep A. Planell

4.1 Introduction

Biomechanics is the study of the mechanics of a part or function of a living body and of the forces exerted by muscles and external loading on the skeletal structure. Biomechanics dates back to ancient times where the study of arthritis was known to be induced by joint disease. But it is only at the beginning of the twentieth century that biomechanical studies of joint materials such as articular cartilage, ligament, and bone began. Living tissues have some similarities with conventional engineering materials although they usually have complex structures that make them more difficult to study. In this chapter, a description of the composition and structure of the main tissues found in mammals is given. The relations between composition, structure and biomechanical properties are presented for bone, cartilage, skin, tendons and ligaments, muscles, and blood vessels and arteries. Finally, some aspects of joint biomechanics are described.

4.2 Bone Biomechanics

4.2.1 Bone Composition and Structure

4.2.1.1 Composition

Bone is a hard connective tissue that fulfils three main functions: (1) it gives support to the body structure, (2) it serves as a protection shield against external loadings, and (3) it provides a framework that allows skeletal motion. Bone is made up of a mineral or inorganic phase (60–70% of the tissue), of water (5–%), and of an organic matrix that makes up the remainder. Approximately 90% of the organic matrix is collagen and 10% noncollagenous proteins. Bone strength is given mainly by its mineral phase made of impure hydroxyapatite crystals $(Ca_{10}(PO_4)_6(OH)_2)$ with carbonate ions. The small crystals are in the shape of needles, plates and rods located within and between collagen fibers. The plate-like crystals have dimensions of 20–80 nm long and 2–5 nm thick.

Bone contains four types of cells: (1) osteoprogenitor cells, (2) osteoblasts, (3) osteocytes, and (4) osteoclasts, of which osteocytes are the most abundant. Bone contains a small number of mesenchymal cells called osteoprogenitor cells that have the ability to proliferate and differentiate into osteoblasts. The bone matrix is produced by the osteoblasts that differentiate into osteocytes when surrounded by bone matrix. Osteocytes are mature bone cells with extensive cell processes that project through the canaliculi. Through a network of cells, they establish contact and communication between adjacent osteocytes and the central canals of osteons via gap junctions. The transmission of information is possible between osteocytes themselves and between osteocytes and osteoblasts on the bone surface through rapid fluxes of bone calcium across the gap junctions. In addition to the gap junctions, the interstitial fluid that surrounds the osteocytes and their processes provides an additional route for the diffusion of nutrients and waste products. This information network is also linked to osteoclasts activity that are giant cells with 50 or more nuclei, and have the function of removing bone matrix.

Bone can be viewed as a composite material made of solid and liquid phases. The solid phase is represented by the mineral phase whereas the liquid phase is represented by the interstitial fluid. The solid phase is made up of collagen type I produced by osteoblast bone cells. Apart from collagen type I, osteoblast produces a variety of noncollagenous proteins, including osteocalcin, osteopontin, osteonectin, and proteoglycans; regulatory factors, such as cytokines, growth factors, and prostaglandins; and neutral proteases, alkaline phosphatase, and other enzymes that degrade the extracellular matrix and prepare it for calcification.

Bone collagen is constructed in the form of a triple helix of two identical $\alpha1(I)$ chains and one unique $\alpha2$ chain stabilized by hydrogen bonding between hydroxyproline and other charged residues. This configuration gives a fairly rigid linear molecule 300 nm long. Each molecule is aligned with the next in a parallel fashion in a quarter-staggered array to produce a collagen fibril. The collagen fibrils are then grouped in bundles to form the collagen fiber.

4.2.1.2 Bone Structure

There are two types of bone: trabecular (spongy or cancellous) bone and cortical (dense or compact) bone (Fig. 4.1). Trabecular bone is found in the epiphysis and metaphysis of long bones and inside flat or small bones. Trabecular bone has an extensive network of small and interconnected plates and rods of individual trabeculae oriented according to the external loading. Cortical bone consists of layers with vascular channels surrounded by lamellar bone. This arrangement is called the osteon or Haversian system. The central canal of an osteon contains cells, vessels, nerves and the canals connecting osteons are called Volkmann's canals.

At the microscopic level, trabecular and cortical bones consist of two forms: woven and lamellar. Woven bone is considered immature bone, or primitive bone, and is found in the embryo, the newly-born, and in fracture healing. Woven bone has a matrix of interwoven coarse collagen fibers with osteocytes distributed more or less at random. It is less organized than lamellar bone. Woven bone is replaced
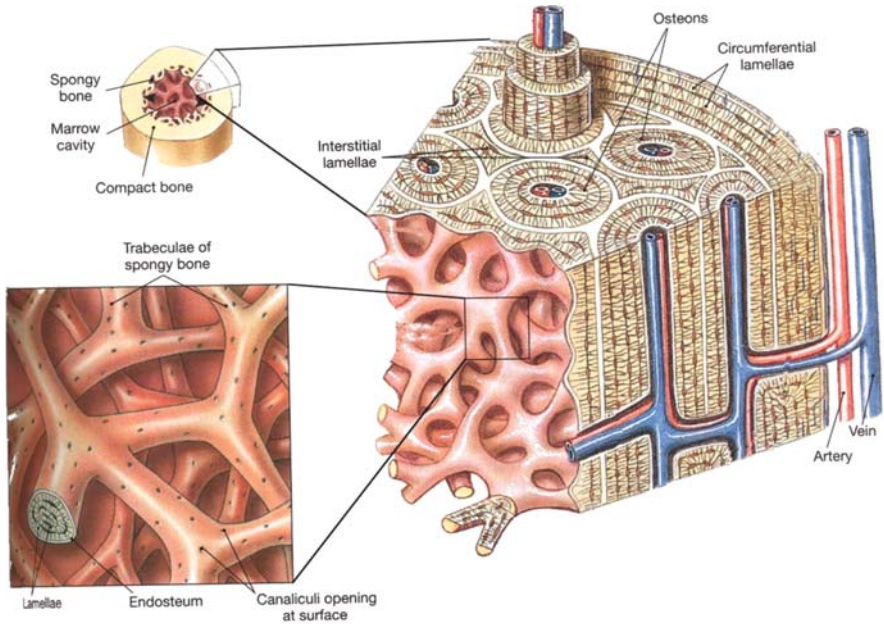
Fig. 4.1 Structure of osseous tissue (adapted from Martini 1989)

by lamellar bone at age 2 or 3 years. Lamellar bone is a mature bone that results from the remodeling of immature woven bone. Lamellar bone is highly organized; stress-oriented collagen of lamellar bone gives its anisotropic properties.

4.2.1.3 Bone Physical Properties

One of the fundamental difference between cortical and trabecular bone is its apparent porosity. Apparent porosity is the ratio of the mass bone tissue in a specimen to the bulk volume of the specimen. Typical apparent density for cortical bone and trabecular bone are $1.85 \, \text{g/cm}^3$ and $0.30 \, \text{g/cm}^3$ respectively with a much higher variability and standard deviation for trabecular bone. Physical properties vary from one bone to another depending on various parameters such as apparent density, ash density (total mineral content divided by bulk volume), histology (number of osteons, primary versus secondary bone), collagen composition and content, orientation of the collagen fibers and mineral, composition of the cement lines, bonding between the mineral and collagen phases, and accumulation of microcracks in the bone matrix and around osteons. Apparent density can be correlated to Young's modulus and ultimate strength using a power law with exponents for modulus ranging from 1.5 to 7.5. Volume fraction (proportional to apparent density) and mineral content (proportional to ash density) can also be correlated to Young's modulus and strength using a power law. Bone mineral content is the ratio between the mineral weight and the dry weight of the bone sample. The bone sample is burnt to

determine its mineral content or the ash fraction. Water content is also important in the mechanical properties of cortical bone. Wet bone, as found in situ, is less stiff, less strong, and less brittle than fully dried bone.

4.2.2 Biomechanical Properties of Bone

Bone has a major advantage over engineering structural materials in that it is self-repairing and can alter its properties and geometry in response to changes in mechanical and metabolic demand. Bone properties also change from species to species. Bone physical properties differ from one person to another but also within one individual from one location to another. Due to its different apparent porosity, mechanical properties of trabecular bone and cortical bone are clearly different.

4.2.2.1 Cortical Bone

Cortical bone accounts for approximately 80% of the skeletal mass. Cortical bone and relatively stiff trabecular bone have Young's moduli of about 17 GPa and 1 GPa respectively (it also depends on the specie and on the type of bone). Thus, most metals used in orthopedic applications are an order of magnitude stiffer than cortical bone. Human cortical bone is usually considered to be transversely isotropic with mechanical properties substantially different in the longitudinal direction (parallel to the axis of the osteons) than in the radial or circumferential directions but it has similar properties in the radial and circumferential directions. The modulus of cortical bone in the longitudinal direction is approximately 1.5 times its modulus in the transverse direction. Cortical bone has a higher strength in compression than in tension, and is stronger in the longitudinal direction than in the transverse direction. For longitudinal loading, cortical bone is a tough material because it can absorb substantial energy before fracture. Furthermore, cortical bone can be classified as a relatively ductile material for longitudinal loading since its ultimate strain for longitudinal loading is substantially larger than its yield strain. However, it is relatively brittle for transverse loading. Cortical bone exhibits a viscoelastic behavior, i.e., it is sensitive both to strain rate and the duration of the applied loads (Fig. 4.2). Elastic modulus and strength have a power law correlation with strain rate with a coefficient of 0.06 approximately. The in vivo strain rate for bone can vary from of 0.001 to 0.01 per second in the course of daily activities. Thus, in this range of strain rate, the Young's modulus increases by approximately 15% whereas it is about 20% stronger (Fig. 4.3). At very high strain rates representing high-impact trauma, cortical bone becomes more brittle. Thus, cortical bone exhibits a ductile to brittle transition as the strain rate increases. The three characteristic stages of creep behavior as described in many conventional engineering materials is also observed in cortical bone (Fig. 4.4). The mechanical properties of cortical bone progressively deteriorate with aging for both men and women. During aging there is an embrittlement of cortical bone with a reduction in energy absorption (area under stress-strain

Fig. 4.2 Strain rate dependence of cortical bone material behavior. Both modulus and strength increase for increased strain rates (adapted from McElhaney 1966)
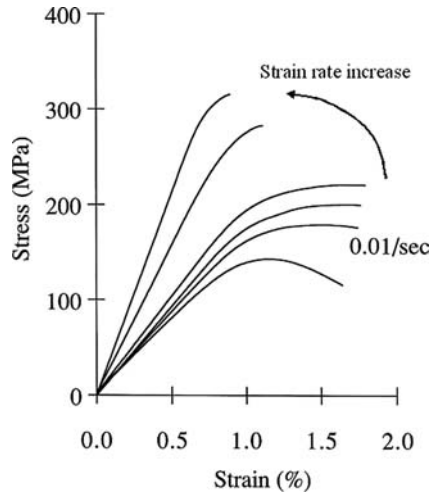
Fig. 4.3 Dependence of modulus and ultimate tensile strength of human cortical bone for longitudinal loading as a function of strain rate (adapted from Wright 1976)
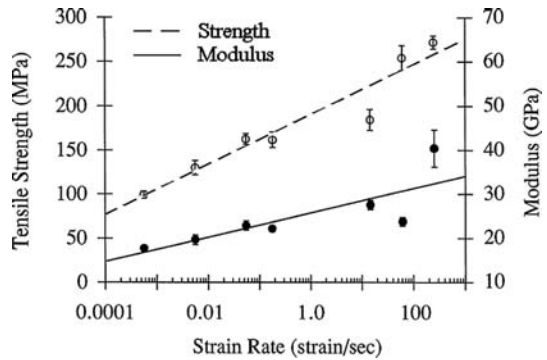
curve) due to a reduction in ultimate strain and therefore cortical bone becomes more brittle. As in any conventional engineering materials cortical bone is subjected to fatigue failure. However, due to its ability to remodel, fatigue failure can be reduced significantly. Cracks that have been formed during strenuous activities can be closed and repaired through bone remodeling activity. The fatigue mechanism for cortical bone can be described in three characteristic stages of fatigue fracture, corresponding to crack initiation, crack growth (propagation), and final fracture. In the primary stage, crack initiates due to local stress concentration given by haversian canals, lacunae, or canaliculi. In the secondary stage, crack propagation results in a slow but steady further decrease in stiffness and strength. These microcracks propagate and tend to join together once they progress beyond the initiation stage. In the tertiary stage, fracture is preceded by a rapid decrease in the ability to support load. The larger cracks run into weak material interfaces (cement lines) between the osteons and can lead to debonding of the osteons. However, imperfections in bone
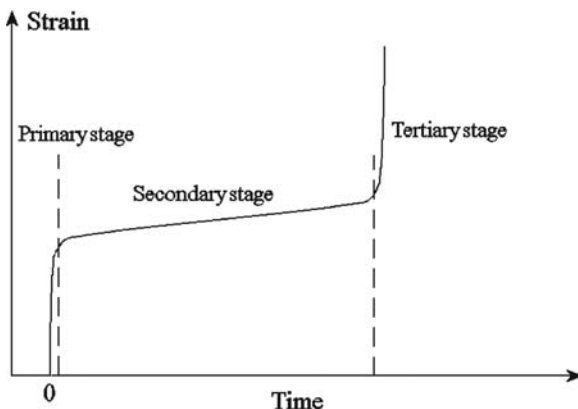
microstructure, such as the cement line, tend to stop the progression of cracks by changing the direction of crack propagation from perpendicular to the loading direction to parallel to the loading direction. The final stage of fatigue fracture occurs because cracks coalesce and become so large that the weak interfaces can no longer absorb them. The cracks can thus cross the bone and induce bone failure.

4.2.2.2 Trabecular Bone

Material properties of trabecular bone vary widely depending mainly on the anatomic location and age, for which the apparent density and the architecture can be markedly different. Trabecular bone is best described as an open-celled porous foam. Since trabecular bone is made up of a series of interconnecting trabeculae, it can be idealized as a combination of rod-rod, rod-plate, or plate-plate basic cellular structures where rods and plates represent thin and thick trabeculae respectively. The mechanical properties can vary by a factor of 10 depending on the type and orientation of these basic cellular structures. The architecture of trabecular bone depends on the thickness of individual trabeculae, and the spacing between trabeculae. The architecture of trabecular bone results in anisotropic elastic properties found in elderly lumbar spine. However, in contrast to cortical bone, trabecular bone is nearly isotropic at some anatomic sites like the proximal humerus. For isotropic bones, the elastic modulus and the ultimate strength of trabecular bone in any direction are related to its apparent density by a power-law relationship of the form $E = a + b\rho^c$ where a, b and c are constants that depend on the architecture of the tissue (Fig. 4.5). In general, the exponent c has a value of approximately 2.

Trabecular bone under compression has similar characteristics to porous foam materials such as wood. In the first stage, bone deforms in the linear elastic region, in which individual trabeculae bend and compress as the bulk tissue is compressed (Fig. 4.6). In the second stage, some trabeculae fail or buckle without increase of loading. As more and more trabeculae fail, the strain increases until broken

Fig. 4.5 Compressive modulus as a function of apparent density for trabecular bone (adapted from Keaveny and Hayes 1993)
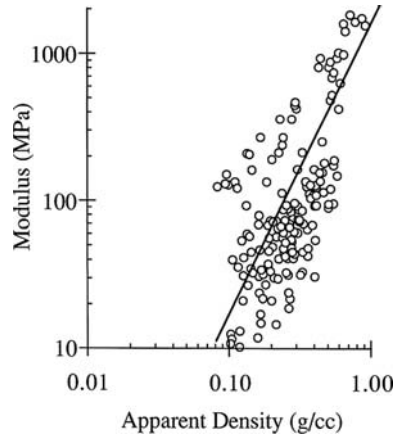
Fig. 4.6 Example of typical compressive stress-strain behaviors of trabecular and cortical bone for different apparent densities (adapted from Keaveny and Hayes 1993)
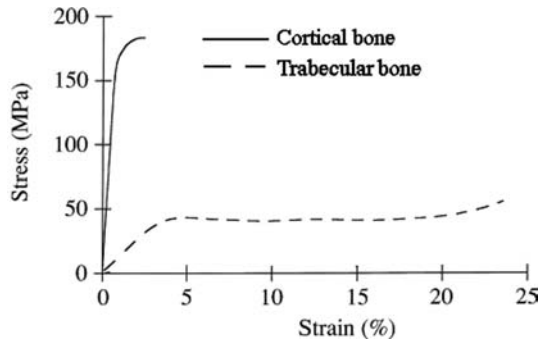
trabeculae begin to fill the pores, causing the specimen to stiffen. Thus, trabecular bone has a unique ability to resist a large compressive load for a minimal mass. This energy absorption allows compressive strains of over 50%. As in fiber-reinforced concrete, the tensile behavior of trabecular bone is poorer.

4.2.3 Bone Remodeling

Bone remodels throughout life to adapt its mechanical properties to the mechanical demands placed on it. Bone remodeling occurs on periosteal, endosteal, haversian canal, and trabecular surfaces. Metabolic turnover of trabecular bone is about eight times greater than that of cortical bone because of its greater surface area for cellular activity. Bone turnover is achieved through the joined activity of osteoclast, resorbing bone, and osteoblast, depositing bone.

Osteoclast binds to the bone surface through cell attachment proteins called integrins, and they resorb bone by isolating an area of bone under the region of cell
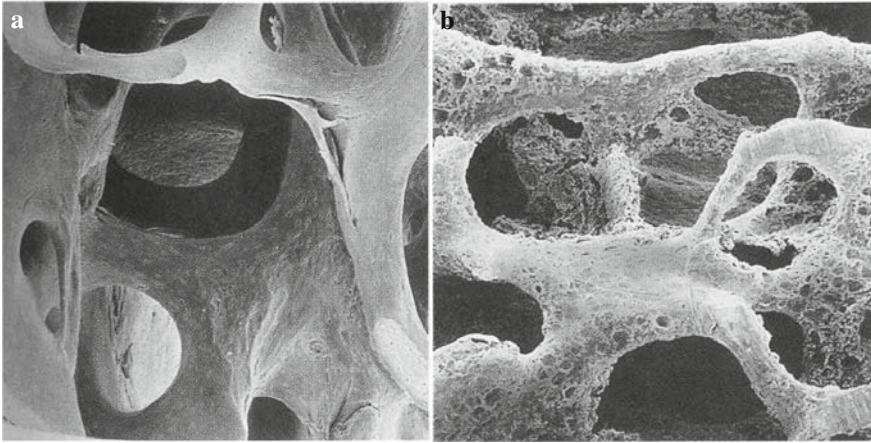
Fig. 4.7 Scanning electron microscopy pictures of (**a**) a normal spongy bone and (**b**) a spongy bone with osteoporosis (taken from Martini, 1989)

attachment. The osteoclast then makes the local environment more acidic by production of hydrogen ions through the carbonic anhydrase system. The lowered pH increases the solubility of the apatite crystals, and after the mineral is removed, the organic components of the matrix are hydrolyzed through acidic proteolytic digestion. Then osteoblasts produce bone matrix and osteoblasts become entrapped to differentiate into osteocytes.

With age, bone density diminishes and architecture changes. The reduction in density depends on a number of factors, including gender and anatomic site and is called osteoporosis (Fig. 4.7). In the lumbar spine, direct measurements have shown a decrease in trabecular bone density of approximately 50% from ages 20 to 80 years. With decreasing density, the number and thickness of the trabeculae decrease with preferential loss of vertical trabeculae, while the size of the intertrabecular spaces increases.

4.3 Cartilage Biomechanics

4.3.1 Cartilage Composition and Structure

4.3.1.1 Structure

Articular cartilage is a supporting connective tissue that consists primarily of a large extracellular matrix (ECM) with a sparse population of specialized cells (chondrocytes) distributed throughout the tissue. These cells live in small pockets called lacunae. The primary components of the matrix are water, proteoglycans, and collagens, with other proteins and glycoproteins present in lower amounts (Fig. 4.8). Cartilage

Fig. 4.8 Schematic representation of the distribution of proteoglycans, collagen and water through articular cartilage (adapted from Mow et al. 1984)
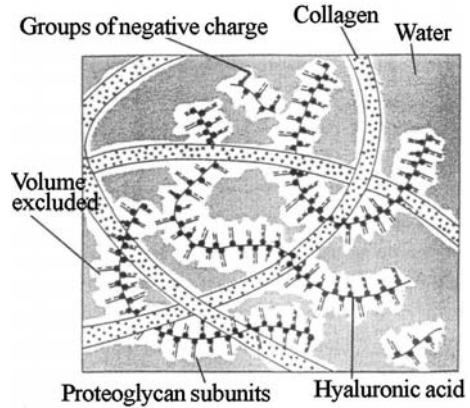
is avascular, and all nutrient and waste product exchange must occur by diffusion through the matrix.

There are three major types of cartilage in the body: hyaline cartilage, elastic cartilage, and fibrocartilage. Hyaline cartilage is the most common type of cartilage and is found mainly at bony joints. It provides stiff but somewhat flexible support because it contains closely packed collagen fibers. It also reduces friction between bony surfaces. Elastic cartilage contains numerous elastic fibers that make it extremely resilient and flexible. It is found in the tip of the nose or in the external ear. Fibrocartilage has little ground substance, and its matrix is dominated by collagen fibers. The collagen fibers are densely interwoven, making this tissue extremely durable and tough. It is found in the intervertebral disc between spinal vertebrae, between the pubic bones of the pelvis, and around or within a few joints and tendons. Fibrocartilage can resist compression, absorbs shocks, and prevents damaging bone-to-bone contact.

The articular cartilage can be divided into four zones from the articular cartilage to the subchondral bone: the superficial zone, the middle or transitional zone, the deep zone, and the zone of calcified cartilage (Fig. 4.9). The structure and composition of each zone is different. The superficial zone (10–20% of the total thickness) is the uppermost zone of the cartilage and forms the gliding surface. The thin collagen fibrils and the chondrocytes are aligned with the long axis parallel to the surface, the proteoglycan content is at its lowest level, and the water content is at its highest level. In the middle zone (40–60% of the total thickness), collagen fibers have a larger diameter and are organized more loosely, and the chondrocytes have a more rounded appearance. As opposed to the superficial zone, the deep zone (around 30% of the total thickness) has the highest concentration of proteoglycans and the lowest water content. The collagen fibers have a large diameter and are organized perpendicular to the joint surface. The chondrocytes are spherical and often are arranged in a columnar fashion. The deepest layer, the zone of calcified cartilage, separates the hyaline cartilage from the subchondral bone.

Fig. 4.9 Schematic diagram
of articular cartilage
organization (adapted from
Mow et al. 1989)
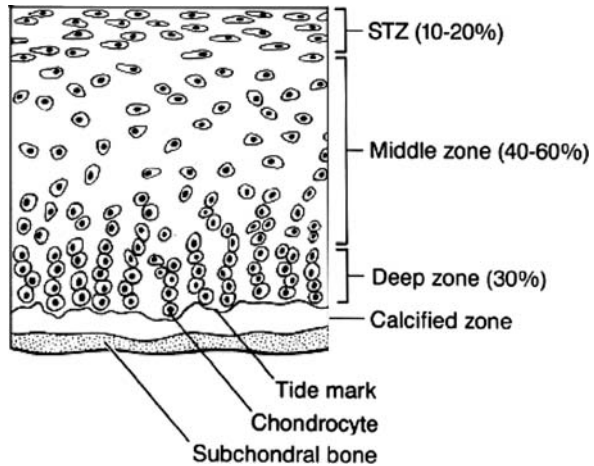
4.3.1.2 Composition

Within the solid matrix of articular cartilage, chondrocytes represent less than 10%
of the total volume tissue and are responsible of the maintenance of the ECM. Within
articular cartilage collagen types II, V, VI, IX, X, and XI are found of which col-
lagen type II represents 90–95% of the total. They are cross-linked in an intrafib-
ril network of proteoglycan gel to provide greater stability and strength. Collagens
provide the tissue's tensile and shear properties and immobilize the proteoglycans
within the ECM. Proteoglycans are complex macromolecules consisting in a protein
core with covalently bound polysaccharide (glycosaminoglycan) chains (Fig. 4.10).
Glycosaminoglycans (GAGs) consist of long-chain, flexible, unbranched, repeating
disaccharide units.

   Normal cartilage has water content ranging from 60% to 80% of its total weight.
Water content decreases from the superficial zone (80%) to the deep zone (60%).
The remaining wet weight of the tissues come from collagens and proteoglycans.
Several other classes of molecules, including lipids, phospholipids, proteins, and
glycoproteins, make up the remaining portion of the ECM. Inorganic salts, such as
sodium, calcium, chloride, and potassium, are dissolved in the tissue water. Water
content within cartilage is very important since it influences the metabolic, chem-
ical, and mechanical environments. Most of the water may be moved through the
ECM by applying a pressure gradient across the tissue, or by compressing the solid
matrix. Cartilage has the ability to withstand high compressive loading due to the
hydrophilic nature of proteoglycans. Proteoglycans have the ability to attract water
through two physiochemical mechanisms: (1) Donnan osmotic pressure and (2)
the entropy tendency of the proteoglycan to gain volume in solution. The Donnan
osmotic pressure is induced by the movement of counterions Ca^{2+} and Na^+ within
the interstitial fluid to neutralize the negative charges of the proteoglycans, or is
caused by the electrostatic repulsion of the proteoglycans. Thus, when water is in
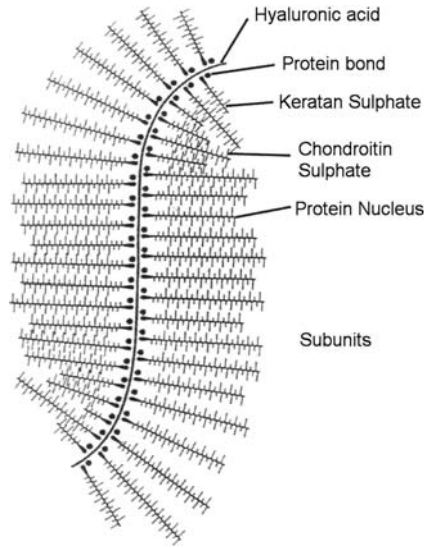
Fig. 4.10 Drawing of an aggregated proteoglycan (adapted from Rosenberg 1975)

contact with proteoglycans, a cohesive and strong solid matrix is formed that retains easily water.

4.3.2 Biomechanical Properties of Cartilage

4.3.2.1 Permeability

The articular cartilage found in articulations is subject to high repetitive loads for many decades. Thus, the microstructure of cartilage is organized into a strong and fatigue resistant matrix that can sustain the high stresses generated by these loads. This soft solid matrix is porous and permeable to the water contained within the microscopic pores. This water can flow through the matrix by the application of a pressure gradient or by compression of the matrix. Thus, the biomechanical properties of articular cartilage are those of a biphasic material similar to a sponge, composed of a solid phase and a fluid phase.

The permeability of a material is a property related to the ability of a fluid to flow through a solid matrix. When a pressure gradient ΔP is imposed across a height h or generated through compression, fluid flows at a volume rate Q across the permeating area A. Darcy's law is used to relate Q with the hydraulic permeability coefficient k as follows: $Q=kA(\Delta P/h)$. The permeation speed V is defined as $V=Q/A\phi^f$ where ϕ^f is the tissue porosity. Results from experiments show that for normal cartilage, the permeability coefficient k ranges from 10^{-15} to 10^{-16} m^4/Ns. The diffusive drag coefficient K defined as $K=(\phi^f)^2/k$, has a high value which indicates that large drag forces are generated within cartilage when fluid flows. Equivalently, this indicates
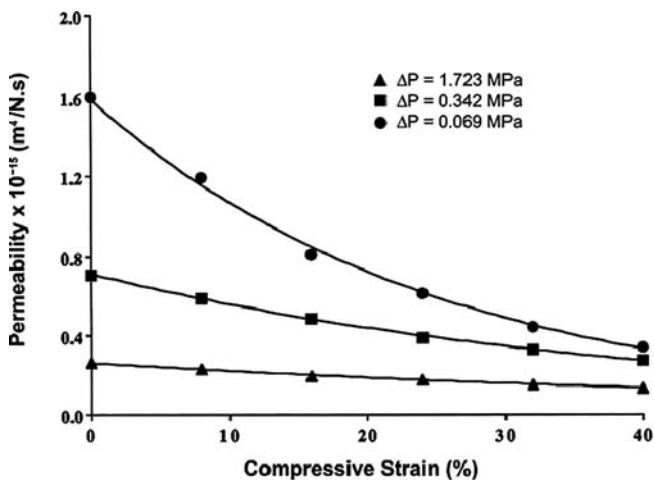
Fig. 4.11 Permeability of normal bovine articular cartilage showing nonlinear strain dependence and pressure dependence (taken from Mow et al. 1991)

that very high pressures are required to move water through the cartilage. Water will also flow through the porous-permeable solid matrix as the tissue is compressed. The compression of the solid matrix increases the fluid interstitial fluid pressure and therefore forces the fluid out of the tissue. Interstitial fluid pressure provides a significant component of total load support, thus minimizing the stress acting on the solid matrix. Articular cartilage permeability decreases nonlinearly with compression (Fig. 4.11). This strain-dependant permeability has two major advantages: (1) it provides a rapid initial exudation of fluid with loading and (2) for further compression, it is more difficult for the fluid to flow and therefore higher fluid pressure are generated which protect the matrix from stresses. This regulatory system has significant influence on aspects such as nutrition requirements of the normal tissue, the joint lubrication, the ability to bear loading and finally on tissue wear.

4.3.2.2 Viscoelastic Properties

Articular cartilage is viscoelastic; it will exhibit a time-dependent behavior when subjected to a constant load or constant deformation. In articular cartilage, there are two mechanisms responsible for viscoelasticity: a flow-independent and a flow-dependent mechanism. The flow-independent viscoelastic behavior is measured with a pure shear experiment in which no fluid flow occurs within the tissue and viscoelasticity is due to properties of the microstructure. Pure shear experiment provides a direct method to determine the flow-independent viscoelastic properties of articular cartilage since no pressure gradient or volume changed is developed. Dynamic pure shear experiments have indicated that articular cartilage is rather elastic (δ of 9° to 20°) and that its shear properties are provided by the proteoglycan-collagen structure and not by the proteoglycan-proteoglycan network.

Fig. 4.12 Creep displacement from an articular surface loaded with a rigid porous filter. The continuous displacement is followed by an exudation of fluid
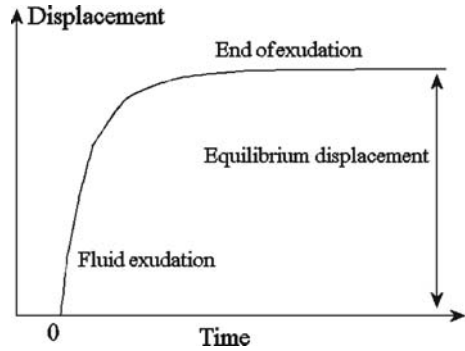
The flow-dependent mechanism depends on interstitial fluid flow and pressurization and corresponds to the movement of fluid as compression occurs as described above. The interaction between the interstitial fluid and the solid matrix is fundamental. Interstitial fluid pressure provides support when cartilage is loaded in compression. However, under constant load or slow motion load, the fluid exudes from the matrix and the load is gradually transferred from the fluid to the matrix. After sufficient time, equilibrium is reached, fluid pressure vanishes and load support is provided entirely by the compressed collagen-proteoglycan solid matrix (Fig. 4.12). This flow-dependant mechanism varies with loading rate. When the loading rate is high, the fluid does not have sufficient time to flow out of the matrix. In contrast, for a slow loading rate, tissue deformation increases as fluid flows.

4.3.2.3 Cartilage Swelling

Swelling is defined as the ability of a material to gain (or to lose) in size or weight when soaked in a solution. For articular cartilage, swelling is mainly derived from the negative charges of the proteoglycans that are packed in a space of 10–15 Å apart. To counteract this negative charge, positive ions are necessary to maintain electroneutrality and therefore the internal pressure is greater than the pressure in the external environment. For normal cartilage equilibrated in physiologic saline (0.15 M NaCl), the swelling pressure has been calculated to range from 0.1 to 0.25 MPa.

4.3.3 Cartilage Degeneration

The most apparent early changes in human osteoarthritic cartilage are increased water content and decreased proteoglycan content. The results of such changes is to increase tissue permeability and thus to decrease the ability for the fluid to bear the load. Therefore, the collagen-proteoglycan solid matrix is subjected to a higher load which is detrimental for the long term survival of cartilage, and may be an important

factor in the development and progression of cartilage degeneration in patients with osteoarthritis.

4.4 Skin Biomechanics

4.4.1 Skin Composition and Structure

The cutaneous membrane, or skin, has two components: the superficial epithelium or epidermis, and the underlying connective tissue of the dermis (Fig. 4.13). An extensive network of blood vessels branches through the dermis, and sensory receptors that monitor touch, pressure, temperature, and pain provide valuable information to the central nervous system about the state of the body. The main functions of the skin include (1) protection of underlying tissues and organs, (2) excretion of salts, water, and organic wastes, (3) maintenance of normal body temperature, (4) synthesis of a steroid, vitamin D_3, that is subsequently converted to the hormone calcitriol, important to normal calcium metabolism, (5) storage of nutrients, (6) detection of touch, pressure, pain, and temperature stimuli and the relay of information to the nervous system.

   The body is covered with two types of skin: thin skin and thick skin with layers of epithelial cells called keratinocytes. Most of the body is covered by thin skin
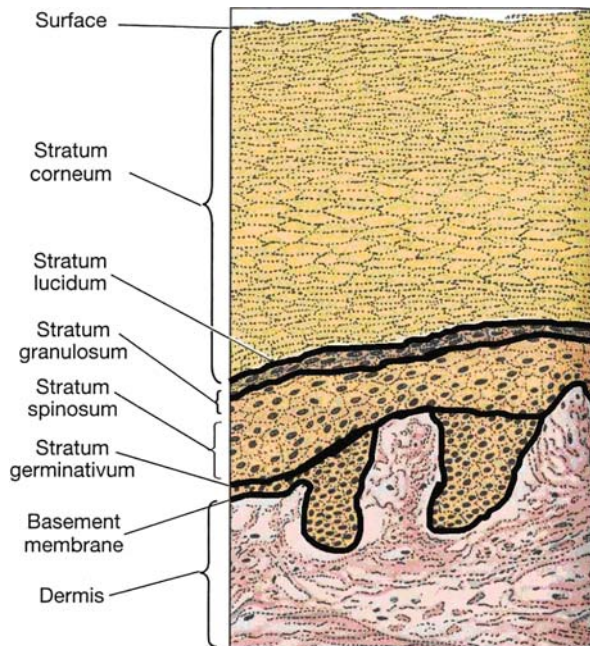
Fig. 4.13 Structure of the epidermis. A portion of the epidermis in thick skin, showing the major stratified layers of epidermal cells (adapted from Martini 1989)

of four layers with a total thickness of 0.08 mm. Thick skin such as that found in the sole and palm may be as much as six times thicker than thin skin and contains five layers (Fig. 4.13). Starting from the basement membrane and going towards the free surface, in thick skin we find the stratum germinativum, the stratum spinosum, the stratum granulosum, the stratum lucidum, and the stratum corneum. The stratum germinativum is firmly attached to the basement membrane. The contours of the skin surface follow the ridge patterns, which vary from small conical pegs (in thin skin) to the complex whorls seen on the thick skin of the palms and soles. Ridges on the palms and soles increase the surface area of the skin and increase friction to provide better grip. Through division of the progenitor cells in the stratum germinativum, cells migrate to the next layer, the stratum spinosum. The stratum spinosum consists of 8–10 layers of cells, where the keratinocytes are bound together by desmosomes. The layer of cells above the stratum spinosum is the stratum granulosum that consists of 3–5 layers of keratinocytes. In thick skin, a glassy stratum lucidum covers the stratum granulosum. The cells in this layer are flattened, densely packed, and filled with keratin. The stratum corneum is found at the interface of both thick and thin skin. There are normally 15–30 layers of keratinized cells in the stratum corneum. The dead cells within each layer of the stratum corneum remain tightly interconnected by desmosomes. It takes 15–30 days for a cell to move from the stratum germinativum to the stratum corneum.

4.4.2 Biomechanical Properties of Skin

As in other tissues, mechanical properties of skin can be obtained in vitro with conventional testing methods. However, control of humidity and temperature are a requisite. When skin is stretched at a slow rate (5 cm/min), the stress strain curve exhibits an initial short portion, where elongation occurs without much force. Then, elongation increases in an exponential manner until, at high extensions, it becomes almost a straight line indicating an elastic behavior immediately before rupture occurs. The tensile strength of skin ranges from 5 to 30 MPa, with the mean showing a maximum of about 21 MPa at 8 years, declining to about 17 MPa at 95 years. The ultimate strain varies from about 35% to 115% with a mean value of around 70%.

As in other soft tissue, skin exhibits a viscoelastic behavior. When the skin is stretched, it adapts its internal collagen network structure to minimize the strain. Thus relaxation and creep also occur in stretched skin. Some hysteresis is also observed when a skin sample is stretched at a given loading rate and unloaded at the same rate. The difference of area between the loading and unloading curves represents the energy lost in the system. The relative amount of energy loss is a measure of the viscosity of the sample. In general, this viscosity parameter decreases from youth to senescence. The elongation still present after each unloading cycle (the residual tension) indicates the plasticity of the skin and is not age dependent. However, the ability of the skin to relax or creep for a given strain or load decreases

with age. This indicates a loss of plasticity with age. Thus, skin has a complex viscoelastic behavior. In general, young skin is less protective against large strains than older skin, young skin is more viscous (or plastic) than older skin, older skin has a proportionately greater elastic region in its stress strain behavior than younger skin. It appears as well that strength and elastic properties of skin are determined by the collagen content, and not the elastin or ground substance components.

Skin displays anisotropic behavior; it was demonstrated by Langer that after having excised a circular patch of skin from a corpse, the circular shape changed into oval. This behavior known as the Langer lines is due to the anisotropic properties of skin as well as to residual stress. It is believed that skin anisotropy is strong with the ratio between the two main moduli being 5.

4.5 Tendon and Ligament Biomechanics

4.5.1 Structure and Composition

Tendons and ligaments are dense connective tissues. Tendon is a complex fibrocartilage material made of collagen fibrils embedded in a matrix of proteoglycans with relatively few cells. Ligaments are short bands of tough, flexible fibrous connective tissue that connect bones and support viscera. Tendons and ligaments are composed of fibroblast cells that are arranged in the spaces between the parallel collagen bundles. The function of tendon is to link muscle to bone or fascia and to transmit loads in tension from the muscle to the bone or fascia to induce movements of parts of the body through articular joints. The function of ligament is to stabilize articular joint and guide its movement in order to prevent instabilities.

The major constituent of tendon and ligament is type I collagen (86% fat-free dry weight for tendon and 70% fat-free dry weight for ligament). Tendons and ligaments are arranged in a hierarchy of four structures. The primary structure of the collagen chain consists of more than 60% of three amino acids: glycine (33%), praline (15%), and hydroxyproline (15%). The secondary structure of collagen relates to the arrangement of each chain in a left-handed configuration, and in its tertiary structure three collagen chains are combined into a collagen molecule. The quaternary structure of collagen consists in the organization of collagen molecules into a stable, low energy biologic unit based on a regular association of adjacent molecules' basic and acidic amino acids (Fig. 4.14).

Ligaments and tendons differ in a number of ways. Physically, ligaments are short and wide tissues that connect between bones to carry load, whereas tendons are long and narrow tissues that connect muscle to bone. Biochemically, ligaments contain a lower percentage of collagen and a higher percentage of ground substance than tendons. Biomechanically, the collagen fibers in tendons are more longitudinally organized than those in ligaments. Ligaments have a function of joint stability and therefore experience a more varied loading which induce a more varied distribution of fiber directions.
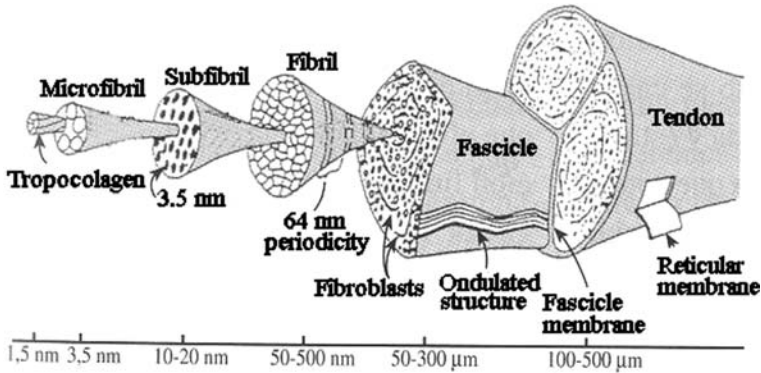
Fig. 4.14 Hierarchic structure of a tendon (adapted from Kastelic et al. 1978)

4.5.2 Biomechanical Properties of Tendons and Ligaments

Tendons and ligaments have very a high tensile strength property due to the arrangement and properties of its collagen fibers. The collagen fibers are arranged parallel to the direction of tensile force, and the mechanical properties of tendons and ligaments are influenced by collagen composition, fiber orientation, and the interaction between the collagen and surrounding ground substance.

The load-elongation curve of the bone-tendon-muscle structure begins with a toe region, in which the tendon can stretch easily at low load (Fig. 4.15). This is due to the straightening of the crimped fibrils and the reorientation of the fibers in the direction of loading. The toe region can extend up to a deformation of 4% and is generally smaller in tendon than in ligament because the initial orientation of collagen fibers in the traction axis of the tendon. The elastic modulus of the linear load–elongation curve ranges from 500 to 1200 MPa. The ultimate strength to traction of a tendon can vary between 50 and 100 MPa. The maximum elongation before rupture varies between 10% and 15%. Tendons exhibit viscoelastic properties and show hysteresis during loading cycles. For repeated loading and unloading cycles with short peri-
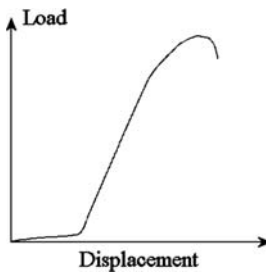
Fig. 4.15 Load-displacement curve from a rabbit tendon until rupture at a constant speed (adapted from Viidik 1967)
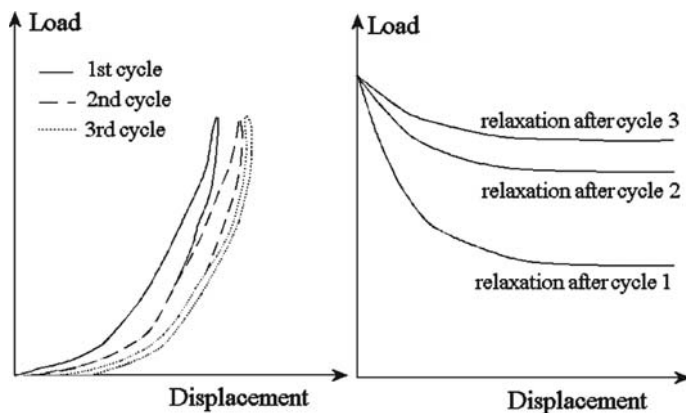
Fig. 4.16 Preconditioning of anterior cruciate ligament. Stress-strain curves and relaxation curves are represented for the first three cycles (adapted from Viidik 1967)

ods of relaxation between cycles, it is observed that the load-elongation curves are displaced to higher elongation for the same load indicating an increased in the toe region (Fig. 4.16). With repeated loadings, a decrease in hysteresis and an increase in the slope of the linear curve are also observed indicating respectively an energy loss and a stiffening of the tendon. Relaxation tests after these loadings show a decrease in the tissue ability to relax (Fig. 4.16). This additional viscoelastic effect in tendons and ligaments is called preconditioning. Following this preconditioning, the behavior of tissues becomes more repeatable.

Since tendons and ligaments are viscoelastic, their properties are strain-rate dependent: an increased elongation rate (and therefore a higher strain rate) will make them appear stiffer. Tensile properties of tendons and ligaments vary depending on the anatomic location and increase with exercise. Age has also a great influence. The modulus increases with age (up to skeletal maturity) and then remains relatively constant.

Rupture of tendons occurs when the maximum strain is overreached. In such cases, the disorganization of fibers happens and the fibers break. Fibers do not break all at once but sequentially. Strain rate is very important in the rupture of the ligament-bone complex. For high strain rates, the ligament usually fails, whereas at low strain rate the bony insertions usually fail. This may indicate that at high strain rate, bone is stronger than ligaments.

4.6 Muscle Biomechanics

4.6.1 Muscle Structure and Composition

Skeletal muscle is the largest tissue mass in the body, corresponding to 40–45% of the total body weight. It contains active or contractile elements and passive or inert elements. It is a composite material that consists of muscle cells, organized

Fig. 4.17 Distribution of the muscle fibers

Parallel   Unipennate   Bipennate   Multipennate
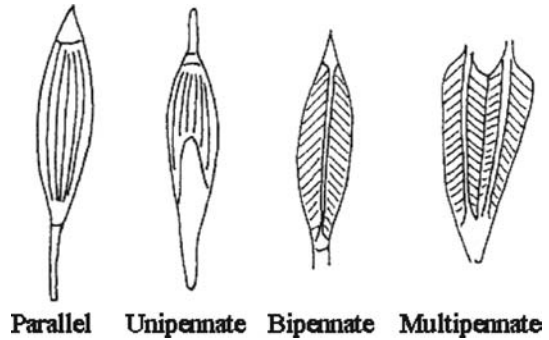
networks of nerves and blood vessels, and an extracellular connective tissue matrix. Nearly 80% of the muscular weight are muscular fibers or excitable cells, capable of contraction following a physiologic stimulus given by a nervous channel.

Muscle fiber is the main component of skeletal muscle. It is a syncytium of many cells fused together with multiple nuclei. Fiber arrangement can be parallel or oblique to the long axis of the muscle (Fig. 4.17). Muscle fibers are organized into contractile units embedded in the surrounding connective tissue to provide a joined motion of the fibers. Within the fibers are enclosed the myofibrils. Muscle fibers are arranged together as fascicles, which are large enough to be visible to the naked eye. The connective tissue surrounding fascicles is called the perimysium.

Since muscle function depends very much on its fiber properties, the main types of muscle fibers (I, II and other intermediate types) differ depending on its histochemical, ultrastructural, and physiologic properties. Fibers type I, or slow or red fibers, have an oxidized metabolism and a high content in oxyhemoglobin that allow specialization in strong and slow contractions. Fibers type II, or fast or pale fibers, have a glycolic metabolism and are used for quick and precise movements. From a functional point of view, muscle fibers are grouped together to form motor units controlled by neuronal signal. In one motor unit, all fibers are of the same type. Depending on the function of the muscle, there is a higher proportion of fibers of type I or II and therefore there are slow and quick muscles. However, depending on the training, the proportion of fibers can be altered and therefore the muscle can behave as one type or another.

The functional and contractile properties of muscles are mainly determined by fiber properties and fiber arrangement (Fig. 4.18). When muscle contracts, the fibers shorten and the tendon moves in the direction of pull. Since some fibers are pennated, the resultant force acting along the axis of pull is reduced by $\cos \alpha$ where α is the angle of pennation. Although this arrangement reduces the force transmitted to the tendon, it allows a larger number of fibers to be packed in a smaller cross-sectional area. The maximal force produced and the speed of shortening by a muscle is proportional to its physiologic cross-sectional area (PCSA) and the individual muscle fiber length. The architectural arrangement of muscle fibers is quite specific to the task of a muscle.
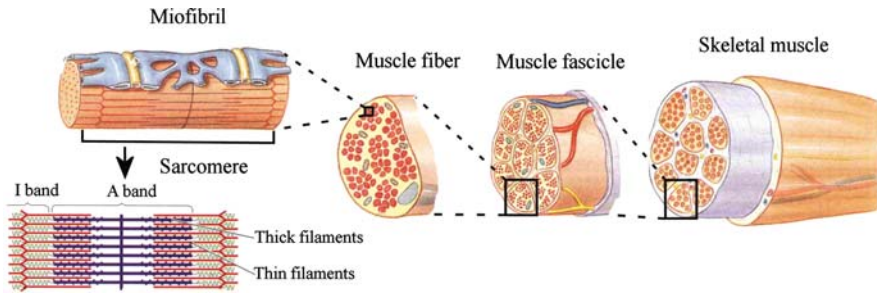
Fig. 4.18 Levels of functional organization in a skeletal muscle fiber (adapted from Martini 1989)

4.6.2 Biomechanical Properties of Muscles

Three types of axial mechanical tests can be performed to determine the mechanical behavior of muscles:

1. In isometric (same length) testing, muscle length is held constant, and the resultant force is measured
2. In isotonic (same load) testing, a muscle is activated to resist to a constant load, while length changes with time are assessed
3. In isokinetic (same speed) activation, the load accommodates to maintain a constant velocity of shortening or lengthening

Muscle activation is seen with a force generation within the muscle. A muscle shortens when the resisting load is less than that generated by the muscle, and this condition is called concentric action (or contraction). A muscle lengthens when the resisting force is greater than that generated by the muscle and it is called eccentric action (or elongation). In a parallel muscle, the sum of the forces induced for each fiber is the force measured in the tendon, whereas in a pennated muscle the force induced by the fibers is the force measured in the tendon divided by the cosine of the angle of pennation of the fibers.

Limb movement and human locomotion is produced through muscle force transmission to the tendon. The area responsible for this force transmission is the myotendinous junction, where there are highly folded membranes at the muscle-tendon interface. The folding increases junctional surface area by 10 to 20 times and therefore decreases the corresponding stresses.

Muscles demonstrate nonlinear, time-dependent viscoelastic responses similar to those of other biologic tissues. An initial model of muscle was proposed by Hill and consisted in two elements in parallel: an elastic component and in series an elastic component with a contractile element (Fig. 4.19). Hill's model can give a general picture of muscle behavior, but it is relatively simple and therefore needs improvement.

Fig. 4.19 Hill's three-element model. Classical representation for the Hill model of muscle are CE = contractile element, SEC = series elements, and PEC = parallel element (taken from Garrett and Best 2000)
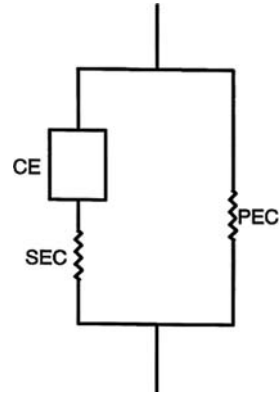
When a passive muscle is stretched, there is a 'toe' region with no tension due to the realignment of fibers as described for tendons and ligaments. Then there is a nonlinear increase in tension. In an activated muscle (by tetanically stimulating the motor neuron) there is almost no 'toe' region and the tension produced is higher. At larger stretches, the tension in activated muscle approaches that in passively stretched muscle. In addition to its non-linear behavior, muscle tissue exhibits viscoelastic properties with strain rate being a determinant parameter. Muscle stretched quickly is stiffer than muscle stretched slowly. Temperature effects are also important: cold muscle is stiffer than warm muscle. Muscle tissue is a tissue that adapts very rapidly to the mechanical environment such as observed in extreme cases in athletes or astronauts.

4.7 Blood Vessel and Arterial Biomechanics

4.7.1 Composition and Structure of Blood Vessels and Arteries

Blood vessels are tubes of variable diameters and for which the relation between the vessel diameter and the wall thickness is variable (Fig. 4.20). For example, the aorta lumen diameter is around 2.5 cm and the wall thickness is 2 mm. The middle size arteries have a diameter of 0.4 cm approximately and a wall thickness of 1 mm. The arterioles have a diameter of 30 μm and a wall thickness of 20 μm. Thus, as the arterial wall is reduced, the ratio of diameter against thickness increases.

The arterial wall consists of three layers: the intima, media, and adventia (Fig. 4.21). The intima consists of the endothelial cells, the basement membrane, and a layer composed of a diffuse aggregation of delicate collagen, elastin, reticular fibers, as well as smooth muscle and other cells. The transition to media is marked by an internal elastic membrane. The media is muscular and divided into concentric layers by elastic lamellae which have fenestrations through which collagen and
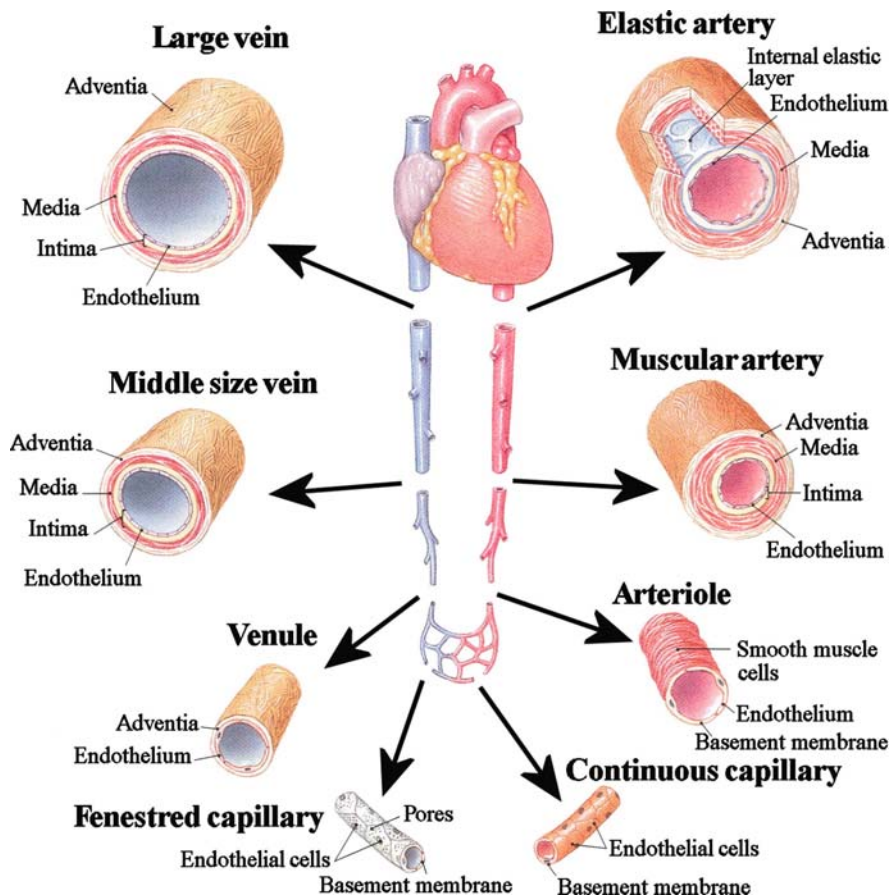
Fig. 4.20 Structure of blood vessels with views of the walls of arteries, veins and capillaries (adapted from Martini 1989)

elastin fibers pass to tie the structure together three-dimensionally. The adventia outside the external elastic membrane is a loosely organized connective tissue.

There are four types of vascular tissues: the endothelium, the elastic fibers, the smooth muscle, and the lax connective tissue of adventia. At the surface of the internal wall of all vessels, there is only one layer of endothelial cells which function is to give a smooth vascular wall as well as a water and solutes permeable surface. It requires little force to deform the endothelial cells and therefore they contribute little to the total elasticity of the vessels. Elastic fibers are numerous in all vessels except the capillaries. They form the internal elastic layer in the shape of a mesh between the intima and the media. They are also found sparsely in the media and adventia layers. Elastic fibers can extend easily, more than six times than rubber, and can increment many times its length before reaching its elastic limit. The function of the fibers, elastic and collagenous, is to produce automatic elastic tension in a passive
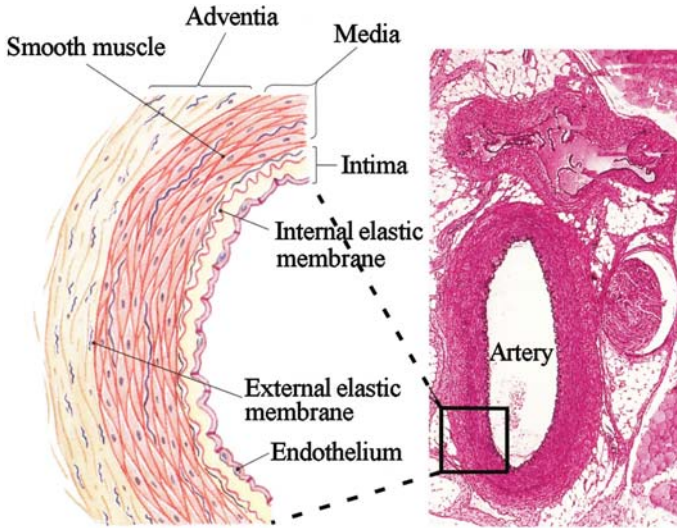
Fig. 4.21 Histologic structure of an artery (adapted from Martini 1989)

form and without energy loss, to resist the arterial pressure. Collagen fibers constitute networks in the media and adventia layers. Collagen fibers are much stiffer than elastic fibers (around 100 times more). But the configuration of the collagenous fibers does not bring any strength until a significant amount of deformation has been achieved. Thus, vessels have an internal surface softer than the external surface. The smooth muscle fibers are distributed circumferentially or helicoidally to the wall and probably contribute little to the total passive elastic stiffness. The function of the smooth vascular muscle is to produce an active stress through contraction. Smooth muscle is most present in arterioles and precapillaries. Thus, blood vessels and arteries have mechanical properties of composite materials and are difficult to obtain.

4.7.2 Biomechanical Properties

Since the composition and structure of the arterial wall change along the arterial tree, the mechanical properties also vary. Elastin is soft, elastic, and relaxes very little, while collagen is also elastic, but has a much larger Young's modulus than elastin, and relaxes more than elastin. As any other soft connective tissues, arteries are viscoelastic with an increased stiffness as deformation increases. Moreover they present anisotropic properties with the elastic modulus that is about three times higher in the radial direction than in the longitudinal direction. The stiffness of aorta and large arteries also increases with age with lower ultimate strain. Until the age of 20 years, changes are due to an increase in the arterial wall thickness. Later, it is the stiffness of the tissues that increases due to the increase in collagen fibers.

Elastic ultimate strain is reduced due an increase of the points of union between the collagen fibers and with other wall structures and therefore limit more easily the extension of the vessel.

Compliance is the property of the hollow organs to be distended and filled with a fluid. The reservoir function of the vein vessels are mainly related with the possibility of deformation of the wall, whereas in the arteries it is only related to the elasticity of the vascular wall. The elasticity in volume is defined as a relative increment of volume $\Delta V/V$ compared to an increment in pressure ΔP applied over a unity of volume $\Delta V/V = 1/E'^*\Delta P$ where $1/E'$ is defined as the vascular compliance.

Thus, $E' = V\frac{\Delta P}{\Delta V}$ is equal to the increment of pressure P that results from an increase of volume of 100%. In other words, this is the elastic modulus in volume and represents the hollow organ stiffness to distention, i.e., the inverse of compliance. Similar properties between modulus of elasticity and compliance are obtained. Vascular tissues have a markedly compliance at low pressure but decreases when the pressure increases. At year 20, the aorta has a maximum compliance.

When fluid flows within the vessels, they induce stresses on the wall surface. Theses stresses vary depending on the pressure exerted within the vessels and the diameter of the vessels (law of Laplace). There is an equilibrium between the intravascular pressure and the wall stress. If the arterial pressure increases, the radius of the vessel tends to increase causing an increase in wall stresses due to strain within the elastic fibers. In pathological vessels, the elastic tissue is almost non-existent and therefore an increase in pressure can rapidly induce damage within the tissues.

4.7.3 Critical Closing Pressure

In small vessels like arterioles the elastic tissue is minimum and the muscle tissue predominates. Thus, the wall tension has to be created strongly active for this tissue. If the arterial pressure increases, the radius increases, and if there is not enough wall tension, there will be some instability that can lead to rupture of the vessel. If the smooth muscle contracts and the wall tension increases progressively until values higher than the arterial pressure, the radius decreases. Following Laplace's law, lower wall tension would thus be needed to maintain the intravascular pressure, and therefore the situation would be unstable, tending to close the vessel. The combination of passive elastic fibers and active muscles in most of the vessels allow to maintain an equilibrium between wall tension and arterial pressure in a large range of radius. The elastic tissue gives a gradual control from the muscular fibers without which it might lead to some instability.

However, if instability occurs in arterioles, the vessel can close progressively at any level of arterial pressure. If the active tension remains constant but the arterial pressure decreases in a pathological manner below a limit called critical closing pressure, the vessel tends to close.

4.8 Joint Biomechanics

4.8.1 Description of Joint Biomechanics

There are three types of articulations depending on its function: synarthrosis, amphiarthrosis, and diarthrosis joints. A synarthrosis joint is an immovable joint that may be fibrous or cartilaginous depending on the nature of the connection, or the two bones may fuse. They are found between the bones of the skull or in the epophyseal plates for example. Amphiarthrosis joint is a slightly movable joint that may be fibrous or cartilaginous depending on the nature of the connection between the opposing bones. It is found, for example, between the tibia and fibula. A diarthrodial joint is a freely movable joint and is also called a synovial joint. Diarthrodial joints are the most common and are all enclosed by a strong fibrous capsule lined with the metabolically active synovium. These joints are typically found at the ends of long bones, such as those of the upper and lower limbs. The load-supporting bony ends of these joints are lined with a thin layer of articular cartilage. These linings form the joint cavity, which contains the synovial fluid.

Under normal conditions, the bony surfaces at a synovial joint cannot contact one another due to the pressure exerted in the synovial fluid. Microscopically, articular surfaces are relatively rough compared to machined bearing surfaces (R_a of 1–$6\,\mu$m, whereas metal joint replacement are on the order of $0.025\,\mu$m). However, in arthritic cartilage a large degree of surface irregularity is usually observed. These surface irregularities have profound effects on the lubrication mechanism involved, and thus greatly affect the friction and the rate of degradation of the articular cartilage.

4.8.2 Function of Joint Biomechanics

The articulation of joint biomechanics consists in the movement of two articular cartilages between a synovial fluid, giving a nearly frictionless bearing system (Table 4.1). The synovial fluid has three primary functions: lubrication, nutrient distribution, and shock absorption. When part of an articular cartilage is compressed, some of the synovial fluid is squeezed out of the cartilage and into the space between the opposing surfaces. This thin layer of fluid markedly reduces friction between moving surfaces. When the compression stops, synovial fluid is sucked back into the articular cartilages. The synovial fluid circulates continuously to provide nutrients and a waste-disposal route for the chondrocytes of the articular cartilages. Synovial fluid also provides shock absorption when joints are compressed. When the pressure across a joint suddenly increases, the synovial fluid lessens the shock by distributing it evenly across the articular surface.

Synovial fluid is a highly viscous fluid that exhibits non-Newtonian fluid flow properties, which include viscosity and viscoelastic effects that are shear rate dependent. The viscosity of synovial fluid from normal knee joints decreases from 500 to 0.5 Pa as the shear rate increases from 0.001 to $1\,\text{s}^{-1}$. The shear rate dependent

Table 4.1 Friction coefficients of typical engineering materials and articular cartilage in synovial joints (from Mow et al. 2000, Chapter 5)

Material combination	Friction coefficient
Gold on gold	2.8
Aluminum on aluminum	1.9
Silver on silver	1.5
Steel on steel	0.6–0.8
Brass on steel	0.35
Glass on glass	0.9
Wood on wood	0.25–0.5
Nylon on nylon	0.2
Graphite on steel	0.1
Synovial joints	**Friction coefficient**
Human knee	0.005–0.02
Porcine shoulder	0.020–0.35
Canine ankle	0.005–0.01
Human hip	0.010–0.04
Bovine shoulder	0.002–0.03

viscosity derives from the alignment of the long-chain hyaluronate molecules as the synovial fluid is sheared.

4.8.3 Mechanical Stresses of Joints

Diarthrodial joints undergo heavy mechanical conditions due to the high loading that needs to be sustained repetitively. For example the physiological joint force in the human hip joint is around four times body weight with a minimum of 2 million cycles per year. Nonetheless this joint functions well with little wear during decades. This demands an efficient lubrication process to minimize friction and to prevent wear of cartilage in the joint. However, over time wear can occur and these large cyclical stresses and strains may cause fatigue failure within the bulk material and may grow by an accumulation of microscopic damage within the material. Interfacial wear occurs as a result of solid-solid contact at the surface of bearing materials. Adhesive wear occurs when a junction is formed between the two opposing surface as they come into contact, leading to fragmentation of articular cartilage. This can lead to the creation of a third body. This type of wear occurs frequently in orthopedics, for example, when PMMA cement particles are trapped in the joint space between an endoprosthesis and cartilage.

4.9 Conclusion

A description of most of the tissues within the human body has been given in this chapter and the relation between composition, function and biomechanical response was studied. It has become quite clear that natural tissues are complex material

structures in constant changes and therefore are more difficult to study compared to conventional engineering materials. However, it was seen that the mechanical properties of soft and hard tissues can be similar to some synthetic materials and thus the knowledge acquired on tissues is useful for the design of novel concept of biomaterials that can mimic as much as possible the natural tissues.

Bibliography

Books

Fung YC, ed. Biomechanics: Mechanical Properties of Living Tissues. Springer-Verlag: New York, NY, 1981.

Mow VC, Ratcliffe A, and Woo SL-Y, eds. Biomechanics of Diarthrodial Joints, Springer-Verlag: New York, NY, 1990.

Cowin, SC, ed. Bone Mechanics Handbook. 2nd edn, CRC Press, Boca Raton, 2001

Odgaard A and Weinans H, eds. Bone Structure and Remodeling, World Scientific, Singapore, 1995.

Martini, FH. Fundamentals of Anatomy and Physiology, 4th edn, Prentice Hall International: New Jersey, 1989.

Daniel DM, Akeson WH, and O'Connor JJ, eds. Knee Ligaments: Structure, Function, Injury, and Repair, Raven Press: New York, NY, 1990.

Buckwalter JA, Einhorn TA, and Simon SR, eds. Orthopaedic Basic Science: Biology and Biomechanics of the Musculoskeletal System. 2nd edn,, Am Acad Orthop Surg: Rosemont, Il, 2000, pp. 567–580.

Bilezikian JP, Raisz LG, and Rodan GA, eds. Principles of Bone Biology, Academic Press: San Diego, 1996

Martin RB and Burr DB, eds. Structure, Function, and Adaptation of Compact Bone, Raven Press: New York, NY, 1989.

Mow VC and Ratcliffe A, eds. Structure and Function of Articular Cartilage, Boca Raton, FL, CRC Press, 1993.

Mow VC and Hayes WC, eds. Basic Orthopaedic Biomechanics, 2 edn, Lippincott-Raven: Philadelphia, PA, 1997.

Mow VC, Flatow EL, Atheshian GA. Biomechanics. In Buckwalter JA, Einhorn TA, Simon SR, eds. Orthopaedic Basic Science: Biology and Biomechanics of the Musculoskeletal System, 2nd edn, Am Acad Orthop Surg, 2000, pp. 133–180.

Bone

Ascenzi A. The micromechanics versus the macromechanics of cortical bone – a comprehensive presentation, J Biomech Eng, 1988, 8: 143.

Ashman RB, Corin JD, and Turner CH. Elastic properties of cancellous bone: measurement by an ultrasonic technique. J Biomech, 1987, 20: 979.

Bostrom MPG, Boskey A, Kaufman JJ, and Einhorn TA. Form and function of bone. in Buckwalter JA, Einhorn TA, and Simon SR, eds. Orthopaedic Basic Science: Biology and Biomechanics of the Musculoskeletal System, 2nd edn,, Am Acad Orthop Surg: Rosemont, Il, 2000, Chapter 13, pp. 319–370.

Burr DB, Martin RB, Schaffler MB, and Radin EL. Bone remodeling in response to in vivo fatigue damage. J Biomech, 1985, 18: 189.

Burr DB. Muscle strength, bone mass and age related bone loss. J Bone Miner Res, 1997, 12, 1547.

Burstein AH, Reilly DT, Martens M. Aging of bone tissue: Mechanical properties. J Bone Joint Surg, 1976, 58B: 82–86.

Burstein AH, Zika JM, Heiple KG, and Klein L. Contribution of collagen and mineral to the elastic-plastic properties of bone. J Bone Joint Surg, 1975, A57: 956.

Bonfield W and Tully AE. Ultrasonic analysis of the Young's modulus of cortical bone. J Biomed Eng, 1982, 4: 23–27

Caler WE, Carter DR. Bone creep-fatigue damage accumulation. J Biomech, 1989, 22: 625–635.

Carter DR, Caler WE. A cumulative damage model for bone fracture. J Orthop Res, 1985, 3:84–90.

Carter DR and Hayes WC. Bone compressive strength: The influence of density and strain rate. Science, 1976, 194: 1174.

Carter DR and Hayes WC. The compressive behaviour of bone as a two-phase porous structure. J Bone Joint Surg, 1977, 59A: 954–962.

Courtney AC, Wachtel EF, Myers ER, and Hayes WC. Effects of loading rate on strength of the proximal femur. Calcif Tissue Int, 1994, 55: 53–58.

Currey JD. The effects of drying and re-wetting on some mechanical properties of cortical bone. J Biomech, 1988, 21: 439–441.

Ding M, Dalstra M, Danielsen CC, Kabel J, Hvid I, and Linde F. Age variations in the properties of human tibial trabecular bone. J Bone Joint Surg (Br), 1997, 79: 995–1002.

Donahue HJ. Gap junctional intercellular communication in bone: A cellular basis for bone mechanostat set point. Calcif Tissue Int, 1998, 62: 85.

Evans GP, Behiri JC, Currey JD, and Bonfield W. Microhardness and Young's modulus in cortical bone exhibiting a wide range of mineral volume fractions and in a bone analogue. J Mater Sci Mater Med, 1990, 1: 38.

Frost HM. A determinant of bone architecture: the minimum effective strain. Clin Orthop Rel Res, 1983, 175: 286

Frost HM. Bone "mass" and the "mechanostat", a proposal. Anat Rec, 1987, 219: 1.

Gibson LJ. The mechanical behaviour of cancellous bone. J Biomech, 1985, 18: 317.

Hasegawa K, Turner CH, and Burr DB. Contribution of collagen and mineral to the elastic anisotropy of bone. Calcif Tissue Int, 1994, 55: 381.

Holden JL, Clement JG, and Phakey PP. Age and temperature related changes to the ultrastructure and composition of human bone mineral. J Bone Miner Res, 1995, 10: 1400.

Keaveny TM and Hayes WC. Mechanical properties of cortical and trabecular bone, in Hall BK, ed. Bone, vol 7, CRC Press: Boca Raton, FL, 1993, pp. 285–344.

Keaveny TM, Wachtle EF, Ford CM, and Hayes WC. Differences between the tensile and compressive strengths of bovine tibial trabecular bone depend on modulus. J Biomech, 1994, 227: 1137–1146.

Klein-Nulend J, van der Plas A, Semeins CM, Ajubi NE, Frangos JA, Nijweide, PJ, and Burger EH. Sensitivity of osteocytes to biomechanical stress in vitro. FASEB J, 1995, 9: 441–445.

Lanyon LE. Osteocytes, strain detection, bone modeling and remodeling. Calcif Tissue Int, 1993, 53(Suppl 1): 102.

Linde F, Hvid I, and Pongsoipetch B. Energy absorptive properties of human trabecular bone specimens during axial compression. J Orthop Res, 1989, 7: 432–439.

Linde F, Nørgaard P, Hvid I, Odgaard A, and Søballe K. Mechanical properties of trabecular bone: Dependency on strain rate. J Biomech, 1991, 24: 803.

Linde F and Sorensen HC. The effect of different storage methods on the mechanical properties of trabecular bone. J Biomech, 1993, 26: 1249.

Martin RB. Toward a unifying theory of bone remodeling. Bone, 2000, 26: 1.

Martin RB and Boardman DL. The effects of collagen fiber orientation, porosity, density and mineralization on bovine cortical bone bending properties. J Biomech, 1993, 26: 1047–1054.

McElhaney JH. Dynamic response of bone and muscle tissue. J Appl Physiol, 1966, 21: 1231–1236.

Melvin JW. Fracture mechanics of bone. J Biomech Eng, 1993, 115: 549–554.

Owan I, Burr DB, Turner CH, Qiu J, Tu Y, Onyia JE, and Duncan RL. Mechanotransduction in bone: osteoblasts are more responsive to fluid forces than mechanical strain, 1997, Am J Physiol, 273: C810.

Peaock M. Effects of calcium and vitamin D insufficiency on the skeleton. Osteoporosis Int, 1998, 8: S45.

Reilly DT, Burstein AH, and Frankel VH. The elastic modulus for bone. J Biomech, 1974, 7: 271–275.

Rho J-Y, Ashman RB, and Turner CH. Young's modulus of trabecular and cortical bone material: ultrasonic and microtensile measurements. J Biomech, 1993, 26: 111–119.

Rho J-Y, Tsui TY, and Pharr GM. Elastic properties of human cortical and trabecular lamellar bone. Bone, 1996, 18: 417–428.

Rice JC, Cowin SC, and Bowman JA. On the dependence of the elasticity and strength of cancellous bone on apparent density. J Biomech, 1988, 21: 155–168

Roodman GD. Cell biology of the osteoclast. Exp Hematol, 1999, 27: 1229–1241

Schaffler MB and Burr DB. Stiffness of compact bone: effects of porosity and density. J Biomech, 1988, 21: 13–16

Schaffler MB, Radin EL, and Burr DB. Mechanical and morphological effects of strain rate on fatigue of compact bone. Bone, 1989, 10: 207–214

Turner CH. Yield behaviour of cancellous bone. J Biomech Eng, 1989, 111: 1–5

Van der Perre G and Lowet G. Physical meaning of bone mineral content parameters and their relation to mechanical properties. Clin Rheumatol, 1994, 13(Suppl 1): 33.

Wright TM and Hayes WC. Tensile testing of bone over a wide range of strain rates: effect of strain rate, micro-structure and density. Med Biol Eng Comput, 1976, 14:671–680.

Cartilage

Anderson DD, Brown TD, and Radin EL. The influence of basal cartilage calcification on dynamic juxtaarticular stress transmission. Clin Orthop, 1993, 286:298–307.

Armstrong CG and Mow VC. Variations in the intrinsic mechanical properties of human articular cartilage with age, degeneration, and water content. J Bone Joint Surg, 1982, 64A: 88–94.

Bachrach NM, Valhmu WB, Stazzone E, Ratcliffe A, Lai WM, and Mow VC. Changes in proteoglycan synthesis of chondrocytes in articular cartilage are associated with the time-dependent changes in their mechanical environment. J Biomech, 1995, 28: 1561–1569.

Buckwalter JA and Mankin HJ. Articular cartilage I: Tissue design and chondrocyte-matrix interactions. J Bone Joint Surg, 1997, 79A: 600–611.

Buckwalter JA, Kuettner KE, and Thonar EJ. Age-related changes in articular cartilage proteoglycans: electron microscopic studies. J Orthop Res, 1985, 3: 251–257.

Edwards J. Physical characteristics of articular cartilage. Proc Inst Mech Eng, 1967, 181: 16.

Freeman MAR. The fatigue of cartilage in the pathogenesis of osteoarthrosis. Acta Orthop Scand, 1975, 46: 323.

Front P, Aprile F, Mitrovic DR, and Swann DA. Age-related changes in the synthesis of matrix macromolecules by bovine articular cartilage. Connect Tissue Res, 1989,19: 121–133.

Guilak F. Compression-induced changes in the shape and volume of the chondrocyte nucleus. J Biomech, 1995, 28: 1529–1541.

Lai WM and Mow VC. Drag-induced compression of articular cartilage during a permeation experiment. J Biorheol, 1980, 17: 111.

Lohmander S. Proteoglycans of joint cartilage: structure, function, turnover and role as markers of joint disease. Baillieres Clin Rheumatol, 1988, 2: 37–62.

Mankin HJ, Mow VC, Buckwalter JA, Iannotti JP, and Ratcliffe A. Articular cartilage structure, composition, and function, in Buckwalter JA, Einhorn TA, and Simon SR, eds. Orthopaedic

Basic Science: Biology and Biomechanics of the Musculoskeletal System, 2nd edn, Am Acad Orthop Surg: Rosemont, Il, 2000, Chapter 17, pp. 443–470.

Mankin HJ and Thrasher AZ. Water content and binding in normal and osteoarthritic human cartilage. J Bone Joint Surg, 1975, 57A: 76–80.

Mow VC, Homes MH, and Lai WM. Fluid transport and mechanical properties of articular cartilage. A review, J Biomech, 1984, 17: 377.

Mow VC, Proctor CS, and Kelly MA. Biomechanics of articular cartilage, in Nordin M and Frankel VH, eds. Basic Biomechanics of the Musculoskeletal System, 2nd edn, Lea & Febiger: Philadelphia, PA, 1989, pp. 31–57.

Mow VC and Ratcliffe A. Structure and function of articular cartilage and meniscus, in Mow VC and Hayes WC, eds. Basic Orthopaedic Biomechanics, 2nd edn, Lippincott-Raven: Philadelphia, PA, 1997, pp. 113–177

Mow VC, Ratcliffe A, and Poole AR. Cartilage and diarthrodial joints as paradigms for hierarchical materials and structures. Biomaterials, 1992, 13: 67–97.

Mow VC, Zhu W, and Ratcliffe A. Structure and function of articular cartilage and meniscus, in Mow VC and Hayes WC, eds. Basic Orthopaedic Biomechanics, Raven Press, New York, NY, 1991, pp 143–198.

Roughley PJ. Structural changes in the proteoglycans of human articular cartilage during aging. J Rheumatol, 1987, 14: 14–15.

Rosenberg LC. Structure of cartilage proteoglycans. in Burleigh PMC, Poole AR, eds. Dynamics of connective tissue macromolecules, North Holland Publisher: Amsterdam, 1975, pp. 105–128

Scott JE. Proteoglycan: collagen interactions in connective tissues. Ultrastructural, biochemical, functional and evolutionary aspects. Int J Biol Macromol, 1991, 13: 157–161.

Skin

Barbenel JC and Evans JH. The time-dependent mechanical properties of skin. J Invest Dermatol, 1977, 69: 31820.

Edwards C and Marks R. Evaluation of biomechanical properties of human skin. Clin Derm, 1995, 13: 375–380.

Finlay B. The torsional characteristics of human skin in vivo, J Biomed Eng, 1984, 6: 567–73.

Langer K. On the anatomy and physiology of the skin, III (1862). Translated by Gibson T. Br J Plast Surg, 1978, 31: 185–99.

Siguhara T, Ohura T, Homma K, and Igawa HH. The extensibility in human skin: variations according to age and site. Br J Plast Surg, 1991, 44: 418–22.

Vogel HG. Age dependence of mechanical and biochemical properties of human skin, Part I: Stress-strain experiments, skin thickness and biochemical analysis. Bioeng Skin, 1987, 3: 67–91.

Vogel HG. Age dependence of mechanical and biochemical properties of human skin. Part II: Hysteresis, relaxation, creep and repeated strain experiments. Bioeng Skin, 1987, 3: 141–76.

Wijn P, Brakkee AJM, Kuiper JP, and Vendrik AJH. The alinear viscoelastic properties of human skin in viva related to sex and age, in Marks R and Payne PA, eds. Bioengineering and the skin, MTP Press: Lancaster, 1981, pp. 135–45.

Tendon and Ligament

Amiel D, Frank C, Harwood F, Fronek J, and Akeson W. Tendons and ligaments: a morphological and biochemical comparison. J Orthop Res, 1984, 1: 257–265.

Butler DL, Kay MD and Stouffer DC. Comparison of material properties in fascicle-bone units from human patellar tendon and knee ligaments. J Biomech, 1986, 19: 425–432.

Cribb AM and Scott JE. Tendon response to tensile stress: An ultrastructural investigation of collagen: proteoglycan interactions in stressed tendon. J Anat, 1995, 187: 423–428.

Haut RC. The influence of specimen length on the tensile failure properties of tendon collagen. J Biomech, 1986, 19: 951–955

Haut RC. Age-dependent influence of strain rate on the tensile failure of rattail tendon. J Biomech Eng, 1983, 105: 296–299.

Hubbard RP and Chun KJ. Mechanical responses of tendons to repeated extensions and wait periods. J Biomech Eng, 1988, 110: 11–19.

Kastelic J, Galeski A, and Baer E. The multicomposite structure of tendons. Connect Tissue Res, 1978, 6: 11.

Kennedy JC, Hawkins RJ, Willis RB, and Danylchuck KD. Tension studies of human knee ligaments: yield point, ultimate failure, and disruption of the cruciate and tibial collateral ligaments. J Bone Joint Surg, 1976, 58A: 350–355.

Shadwick RE. Elastic energy storage in tendons: Mechanical differences related to function and age. J Appl Physiol, 1990, 68: 1033–1040.

Viidik A. Experimental evaluation of the tensile strength of isolated rabbit tendons. Scand J Plast Reconstr Surg, 1967, 1: 141.

Woo SL. Mechanical properties of tendons and ligaments: I. Quasistatic and nonlinear viscoelastic properties. Biorheology, 1982, 19: 385–396.

Woo SL, An KA, Frank CB, Livesay GA, Ma CB, Zeminski J, Wayne JS, and Myers BS. Anatomy, biology and biomechanics of tendon and ligament. in Buckwalter JA, Einhorn TA, and Simon SR, eds. Orthopaedic Basic Science: Biology and Biomechanics of the Musculoskeletal System, 2nd edn, Am Acad Orthop Surg: Rosemont, Il, 2000, Chapter 24, pp. 581–616.

Woo SL, Chan SS, and Yamaji T. Biomechanics of knee ligament healing, repair and reconstruction. J Biomech, 1997,30: 431–439.

Woo SL, Gomez MA, Woo YK, and Akeson WH. Mechanical properties of tendons and ligaments: II. The relationships of immobilization and exercise on tissue remodeling. Biorheology, 1982, 19: 397–408.

Woo SL, Ohland KJ, Weiss JA. Aging and sex-related changes in the biomechanical properties of the rabbit medial collateral ligament. Mech Ageing Dev, 1990, 56: 129–142.

Muscle

Alexander R and McN Vernon A. The dimension of knee and ankle muscles and the forces they exert. J Hum Mov Stud, 1975, 1: 115–123.

Baldwin KM, Winder WW, and Holloszy JO. Adaptation of actomyosin ATPase in different types of muscle to endurance exercise. Am J Physiol, 1975, 229: 422–426.

Best TM, McElhaney J, Garrett WE Jr, and Myers BS. Characterization of the passive responses of live skeletal muscle using the quasi-linear theory of viscoelasticity. J Biomech, 1994, 27: 413–419.

Close RI. Dynamic properties of mammalian skeletal muscles. Physiolo Rev, 1972, 52: 129–197.

Elftman H. Biomechanics of muscle. J Bone Joint Surg, 1966, 48A: 363–373.

Garrett WE and Best TM. Anatomy, physiology and mechanics of skeletal muscle, in Buckwalter JA, Einhorn TA, Simon SR, eds. Orthopaedic Basic Science: Biology and Biomechanics of the Musculoskeletal System, 2nd edn, Am Acad Orthop Surg, 2000, Chapter 26, pp. 683–716.

Josephson RK. Contraction dynamics and power output of skeletal muscle. Ann Rev Physiol, 1993, 55: 527–546

Hill AV. The heat of shortening and the dynamic constants of muscle. Proc R Soc London Ser B, 1938, 126: 136–195.

Hill AV. The dimensions of animals and their muscular dynamics. Sci Prog, 1950, 38: 209–230.

Huxley AF and Simmons RM. Proposed mechanism of force generation in striated muscle. Nature, 1971, 233: 533–538.

Huxley AF. Muscle structure and theories of contraction. Prog Biophys Chem, 1957, 7: 255–318.

Murray M, Gardner G, Mollinger L, and Sepic S. Strength of isometric and isokinetic contractions: knee muscles of men aged 20 to 86. Phys Therapy, 1980, 60: 412.

Taylor DC, Dalton JD Jr, Seaber AV, and Garrett WE Jr. Viscoelastic properties of muscle-tendon units: the biomechanical effects of stretching. Am J Sports Med, 1990, 18: 300–309.

Woittiez RD, Huijing PA, Boom HB, et al. A three-dimensional muscle model: quantified relation between form and function of skeletal muscles. J Morphol, 1984, 182: 95–113.

Blood Vessel and Artery

Burton AC. Relation of structure to function of the tissues of the wall of blood vessels. Physiol Rev, 1944, 34: 619–642.

Burton AC. Properties of smooth muscle and regulation of circulation. Physiol Rev, 1962, 42: 1–6.

Nichol J, Girling F, Jerrard W, Claxton EB, and Burton AC. Fundamental instability of small blood vessels and critical closing pressures in vascular beds. Am J Physiol, 1951, 164: 330–344.

Roach MR, Burton AC. The reason for the shape of the distensibility curves of arteries. Canad J Biomech & Physiol, 1957, 35: 681–690.

Roach MR, Burton AC. The effect of age on the elasticity of human arteries. Canad J Biomech & Physiol, 1959, 37: 557–569.

Chapter 5
Metal Corrosion

Miroslav Marek

5.1 Interaction of Metallic Biomaterials with the Human Body Environment

Degradation of metals by interaction with the environment is called corrosion. In most industrial corrosion situations the damage to the object is of main concern, while the effect of the corrosion process on the environment is either negligible, or of secondary importance. For devices inserted in the human body, on the other hand, both the degradation of the device and the effect of the interaction on the physiological system, local or systemic, are of equal importance. Corrosion degradation may result in a loss of mass, mechanical integrity, functionality and, sometimes, esthetic quality, while the release of corrosion products and the flow of the corrosion currents may cause inflammation, allergic reactions, local necrosis and many other health problems. Although the intensity of the interaction is an important parameter, an equally important factor is the identity of the corrosion products and their toxicity, carcinogenicity, allergenicity and other potentially hazardous properties.

The corrosion interaction of electronic conductors, such as metals, with ionically conducting liquids, such as human fluids, is almost always an electrochemical process, involving a change from the metallically bonded solid to dissolved ions or ionically bonded products. There are also physical and chemical rather than electrochemical interactions with components of body fluids and tissues, but they do not fall into the generally accepted meaning of the term corrosion. The only notable exception in the area of biomaterials may be the release of mercury from dental amalgam fillings. Because of the high nobility of mercury but low bond strength, mercury dissolves much more easily in the atomic, non-ionized form.

When a foreign object, such as a medical implant, is inserted in the tissues, it causes irritation not only because of the chemical or electrochemical interactions, but also by its mere presence. The irritation induces foreign body reactions and tendency to remove the foreign body. The type of the physiological response depends on the size and shape of the object as well as the reactivity of the material. While particulates and fluids may be ingested by macrophages, larger objects are usually either extruded or walled-off (encapsulated). Encapsulation first involves accumulation of epithelium cells, which may develop into a membrane and a fibrous capsule.

The presence of a membrane on metallic implants has an effect on the corrosion interaction, which has not been fully examined. There is evidence that the thickness of the membrane increases with increasing reactivity of the material, the reactivity being a combination of the intensity of the corrosion reactions and the toxicity of the released elements. On the other hand, the membrane slows down the transport of the corrosion products into surrounding tissues or fluids, and also affects the chemistry of the environment at the metal surface and thus the corrosion process itself.

While the released metal ions in a sufficiently high concentration may be toxic enough to cause inflammation and necrosis of the cells and tissues, and accumulation of metal ions in various organs may be a health hazard, another common problem are allergic reactions, which require only a small amount of the released ions for a response of the immune system in susceptible individuals. Although allergic reactions are often described as dose-independent, this term is inherently inaccurate, because one could always identify a quantity that would not provoke the reaction. However, the threshold dose is difficult to determine. One of the allergens among components of implant alloys, known for a high percentage of susceptible individuals, is nickel, which is contained in such common alloys as stainless steel and Ni-Ti ("Nitinol") alloys.

For the most common materials for implants and dental restorations and appliances the most important environmental variables from the corrosion viewpoint are the concentrations of chloride and hydrogen ions and dissolved oxygen. The chloride ion concentration varies in a relatively narrow range close to 0.1 mol/L for most body fluids, such as blood plasma, interstitial fluid, bile and gastric juice, but is quite variable in urine (depending on the salt intake) and considerably lower in saliva. The hydrogen ion concentration (usually expressed as pH value) is close to neutral (pH 7) for plasma, interstitial fluid, bile and saliva, but more variable in urine (pH 4–8) and very low (acidic) in gastric juice (pH 1–3).[1] The dissolved oxygen concentration varies, but for most common implant conditions the partial pressure of oxygen is about 40 mmHg (interstitial fluid, venous blood) or 100 mmHg (arterial blood). More detailed data on body fluid compositions can be found in various sources, such as Burke [2] and Lentner [3, 4].

5.2 Electrochemical Reactions on Metallic Biomaterials

Spontaneous electrochemical corrosion of metals is a process that lowers the free energy of the system. With the exception of a few very noble metals, most elements in the metallic condition and in natural environments are in a metastable state, and their energy can be lowered by conversion to an ionic form, thus reversing the

---

[1]The pH value is defined as a function of the activity of hydrogen ions $[H^+]$, pH = $-\log [H^+]$. Solution with pH<7 have a higher concentration of H^+ than when neutral (pH 7) and are acidic; those with pH>7 have the H^+ concentration lower than in neutral water and are alkaline.

process of producing the metal from the ore. During this conversion some of the metal atoms, which in the metallic state share their free electrons with all the other atoms in the given solid, are released from the solid, leave the free electrons in the solid and become positive ions, or cations. Since this would make the solid negatively charged and electrostatic attraction would make further release of positively charged cations difficult, continuation of the process requires another reaction that consumes the superfluous electrons. In the electrochemical terminology the process of creating positive ions is called oxidation, and that of consumption of electrons is called reduction.

The above conversion can be achieved by a chemical reaction of a metal with a nonconductive environment, such as a reactive gas, for instance oxygen, a process called "dry oxidation." Even this outwardly chemical reaction involves transfer of electrons, i.e., in principle, electrochemical reactions. In a conductive environment, however, true electrochemical corrosion occurs, which involves a flow of electricity through the environment between the sites of oxidation and reduction. The electronically conducting solid metal is called an electrode, and an ionically conductive liquid environment is an electrolyte. In the conductive human body fluids electrochemical corrosion is the dominant form of interaction with metals.

The electrochemical degradation of metals caused by dissolution or formation of nonmetallic corrosion products is directly caused by the oxidation part of the process, while reduction plays an enabling role. The formation of cations by stripping the metal atom of some of its electrons does not by itself lowers the energy; as a matter of fact, some energy must be supplied, resulting in a positive energy change, and an "energy barrier". The energy drop, which allows the process to proceed spontaneously, is due to the interaction of the cations with molecules, called "ligands." In aqueous environments, such as body fluids, water molecules almost exclusively serve the role of ligands. In the conventional writing of an oxidation reaction, such as

$$Ni \rightarrow Ni^{++} + 2e^- \tag{5.1}$$

and in any associated calculation of energies, if (5.1) describes a reaction in an aqueous environment, it is assumed that the nickel cation, Ni^{++}, is in this "hydrated" form, i.e., surrounded by attached water molecules. If the reaction proceeds in the opposite direction, i.e., as reduction, the ligands must first be stripped from the cation before the ion can become a neutral atom. For both oxidation and reduction the species involved must overcome an energy barrier called "activation energy" associated with the "activated state" in between the conditions as neutral atoms and hydrated ions.

It is of fundamental interest to know how strong is the tendency of different metals to form these hydrated cations, i.e., to dissolve in water. It is possible to express this tendency by showing the energy change involved in the reaction. However, since the charge of the ions also plays a role, because more highly charged positive ions are more strongly attracted to the negatively charged metal they have left, a more convenient way is to introduce a concept of electrode potential difference $\Delta\Phi$ at

equilibrium, which is proportional to the energy change ΔG and inversely proportional to the ionic charge z,

$$\Delta \Phi = -\Delta G/zF \tag{5.2}$$

where F is a constant, called Faraday,[2] which serves as a conversion between the energy units for ΔG [J/mol] and electric units for the potential [V]. The physical meaning of $\Delta \Phi$ is that at equilibrium it is such a difference between the potential of the hydrated metal ions in the electrolyte just outside the metal surface and the potential of the same ions within the metal surface that the electric work of moving the ions of charge z across this potential difference balances the energy change ΔG involved in the reaction. The negative sign in reaction (5.2) is due to the fact that the potential difference and the energy change are measured in opposite directions.

Since (5.2) is applicable to reactions written in either direction (oxidation or reduction), and since most available data for energies of reactions are relative to the energy involved in the reaction of hydrogen, an international convention has been accepted for expressing the fundamental dissolution tendency of elements using (5.2). According to this convention the reactions are considered at equilibrium (i.e., proceeding at the same rate in both directions), are written as reduction, rather than oxidation, and the values of the energy change ΔG are relative to the energy change in the reaction

$$2H^+ + 2e^- = H_2 \tag{5.3}$$

The potential difference with respect to the potential of a standard reference reaction (5.3) is then called "the single electrode potential" for the given reaction, and a simple symbol, such as e, is used for this parameter. The true meaning of e is thus a tendency for a reaction to proceed spontaneously in the direction of reduction relative to the tendency of hydrogen ions to be reduced in reaction (5.3).

The ΔG values and thus also the values of e generally vary with temperature, pressure, and activity of the ions. When all the reactants and products are in the standard state[3] (ΔG^o) the potential (e^o) is called "the standard single electrode potential." A listing of such standard reduction potentials is called an electromotive or electrochemical series. Because of the way e is defined, the higher the tendency for dissolution, the more negative is the single electrode potential, while noble elements, such as gold, have a positive value of e^o. Selected standard single electrode potentials for some elements common in implant and dental alloys are shown in Table 5.1. The potentials are listed on a standard hydrogen scale (SHE), i.e., with

---

[2]F = 96,485 C/mole

[3]For electrochemical reactions the standard state is defined by temperature $25\,^\circ C$ (298.16 K), pressure 101325 Pa (1 atm, 760 torr), and activity of ions equal to one.

Table 5.1 Standard single
electrode potentials for
reduction reactions of
hydrated ions of selected
metals

Metal	Reaction	e^o (V, SHE)
Gold	$Au^{3+} + 3e^- = Au$	1.498
Platinum	$Pt^{2+} + 2e^- = Pt$	1.188
Palladium	$Pd^{2+} + 2e^- = Pd$	0.987
Silver	$Ag^+ + e^- = Ag$	0.799
Mercury	$Hg_2^{2+} + 2e^- = 2Hg$	0.788
Copper	$Cu^{2+} + 2e^- = Cu$	0.337
Tin	$Sn^{2+} + 2e^- = Sn$	-0.136
Molybdenum	$Mo^{3+} + 3e^- = Mo$	-0.200
Nickel	$Ni^{2+} + 2e^- = Ni$	-0.250
Cobalt	$Co^{2+} + 2e^- = Co$	-0.277
Iron	$Fe^{2+} + 2e^- = Fe$	-0.440
Chromium	$Cr^{3+} + 3e^- = Cr$	-0.744
Zirconium	$Zr^{4+} + 4e^- = Zr$	-1.539
Titanium	$Ti^{2+} + 2e^- = Ti$	-1.630

Based on data in Pourbaix, 1974

respect to the standard potential of the reference reaction (5.3), which is considered
zero.[4]

It is important to keep in mind that standard single electrode potentials of
elements show only the fundamental thermodynamic tendency to form hydrated
cations and thus to dissolve in water under standard conditions, and not necessarily
the susceptibility to electrochemical corrosion. In many cases the corrosion suscep-
tibility is more a function of the rate of the reactions, i.e., the kinetics, than of the
thermodynamic "driving force." On the other hand, since thermodynamics predict
what cannot happen more decisively than what will happen, a lack of thermody-
namic tendency for dissolution is a safe indication than the metal will not corrode.
The choice of noble metals, such as gold and platinum in dentistry and in some
special cases of implants is based on this consideration.

An important factor that must be considered in evaluation of the thermodynamic
tendency (driving force) for a reaction is the difference between the actual and stan-
dard conditions. The effect of the temperature, pressure and activities on the single
electrode potential of a reaction is given by the Nernst equation (5.4). For an elec-
trochemical reaction at equilibrium $aA + bB + z\,e^- = cC + dD$, where [A] and [B]

---

[4]A standard hydrogen electrode (SHE) can be constructed using a platinized platinum electrode
immersed in an acid of unit activity of H^+ ions (pH 0), saturated with hydrogen gas at 1 atm pressure
and maintained at 25 °C. More commonly other, more convenient secondary standard reference
electrodes are used in experimental work, such as mercury/mercury chloride (calomel electrode),
silver/silver chloride and copper/copper sulfate couples in appropriate electrolytes, which have
accurately known and stable potential differences from SHE. In any reporting of potential values
the potential scale used must be specified.

are activities[5] of the oxidized reactants A and B, [C] and [D] are activities of the reduced products C and D, and a, b, c and d are the numbers of the molecules in the reaction

$$e = e^0 + \frac{RT}{zF} \ln \frac{[A]^a \cdot [B]^b}{[C]^c \cdot [D]^d} \qquad (5.4)$$

where R is the universal gas constant, T is temperature (K), and z is the number of electrons.

In the human body the pressure and temperature are not substantially different from those defining the standard state; the chemical species, especially the metallic ions dissolved in the liquids, however, usually are in concentrations (and activities) much lower than one. For a simple reaction showing equilibrium between a metal and its hydrated ions with a charge z, and for human body temperature (37 °C), the Nernst equation simplifies to

$$e = e^o + (0.061/z) \log[Me^{z+}] \qquad (5.5)$$

where the activity of metal ions $[Me^{z+}]$, for diluted solutions is approximately equal to the concentration of the dissolved metal ions at the electrode.

Since the value of e is a measure of the thermodynamic tendency for dissolution in water, the importance of the metal ion concentration for dissolution of metals in the human body now can be appreciated. For example, although the standard single electrode potential for iron ($e_{Fe} = -0.44$ V, SHE), is considerably lower than for cobalt ($e_{Co} = -0.277$ V, SHE), making cobalt significantly more noble than iron, the difference is much smaller at the mean concentrations of the elements in blood plasma, where the concentration of iron normally is much higher than that of cobalt [4]. Also, although noble metals, such as gold, do not have any tendency to dissolve under standard conditions (and are found in nature in the metallic state), as the concentration of their ions approaches zero, even their single electrode potentials, according to (5.4), approach very high negative values, indicating an extremely high driving force for dissolution. The practical significance of this result is that there is no "zero dissolution," even for the most noble metals. On the other hand, for very noble metals, such as Au and Pt, the tendency for dissolution quickly drops to zero at extremely low concentrations at the practical potentials that get established in the human body environment.

Another factor that may substantially affect the tendency for dissolution is the propensity of metal ions to form complexes with components of the environment. In the complicated human body fluids there are many substances, inorganic and organic, that may form complexes with ions of metallic biomaterials. When a

---

[5] Activity of ions a is proportional to the concentration c [mol/L], $a = \gamma . c$, where γ is the "activity coefficient." For dilute solutions γ is close to one and activity may be replaced by concentration. For concentrated solutions, however, the difference between concentration and activity must be considered. For gaseous species partial pressure is used in place of activity in (5.4).

complex has a lower energy than a simple hydrated cation, ΔG in (5.2) is higher (for reduction), and the potential e becomes lower, indicating a higher tendency for dissolution than in pure water.

It is important to keep in mind that a single reaction at equilibrium, i.e., when the actual potential of the electrode E is equal to e of the reaction, does not produce any significant corrosion damage, because the oxidation part of the reaction is balanced by the reduction process of the same reaction, both occurring at the same rate. Even at equilibrium some of the ions entering the liquid may drift away and possibly cause allergic reaction, but the rate of such a release normally is too low to cause a noticeable loss of mass.

True corrosion occurs when the oxidation part of a reaction between metal and its ions significantly exceeds the rate of the same reaction going in the opposite direction, i.e., reduction. For oxidation or reduction to exceed the rate of the other process the actual electrode potential E must deviate from the equilibrium potential e. In spontaneous electrochemical corrosion of a single electrode this deviation is due to some other reaction occurring simultaneously. For a reaction to proceed in the oxidation direction, as shown in (5.1), another, different reaction must be available to consume the electrons liberated in the oxidation process and thus proceed in the direction of reduction. In the body fluids and many other aqueous electrolytes the two dominant reactions that may serve this role are the reduction of hydrogen ions, shown in (5.2), and reduction of dissolved oxygen:[6]

$$O_2 + 4H^+ + 4e^- = 2H_2O \qquad (5.6)$$

Both reactions involve as reactants only species always present in body fluids, hydrogen ions formed by dissociation of water molecules and dissolved oxygen, and thus are, in principle, capable of serving as reduction reactions for electrochemical corrosion.

When the equilibrium potentials of two or more possible reactions on the electronically conducting surface are different, this potential difference causes a flow of electric current I. This current may be the sum of many small current flowing between atomically close sites of the reactions, or a current between quite distant areas of the electrode surface, depending on the conditions, as discussed later. Within the metal electrode electrons carry the current, and the circuit is completed by ionic conduction in the electrolyte. The electric current per unit area (current density i) associated with an electrochemical reaction can be taken as a measure of the rate of that reaction, because it is proportional to the amount of the substances reacted or produced. Faraday's law gives the conversion between the mass of the substance involved and the amount of electricity:

$$m = \frac{QA}{z\mathrm{F}} = \frac{ItA}{z\mathrm{F}} \qquad (5.7)$$

---

[6]Standard potential $e^o = 1.228$ V (SHE) [5].

where Q is the electric charge [C], A is atomic weight [mol^{-1}], z is the oxidation state change, F is the Faraday′s constant, I is the electric current [A], and t is time [s].

The flow of the current causes each reaction to be "polarized," i.e., its potential changes from the equilibrium value. When an electrochemical reaction is polarized the rates of oxidation and reduction are not the same. Polarization in the positive direction of the potential causes the oxidation rate to exceed the reduction rate, because to achieve a positive potential change some electrons must be removed, lowering the force of attraction and thus making the loss of positive cations easier. A potential change in the negative direction results in the rate of reduction exceeding the rate of oxidation. (More accurately, the potential change increases or decreases the activation energy barrier for the transfer of charges between the metal and the electrolyte.)

For only two corrosion reactions on an electrode, the potential of each reaction shifts toward the potential of the other one. If there is no electric resistance in the current circuit the two potentials become identical and the sites are completely "short-circuited." This condition is closely approximated when the oxidation and reduction sites are atomically close. For oxidation and reduction sites macroscopically distant the resistance, usually in the electrolyte, may be sufficient to result in a substantial potential difference between the two reactions.

Regardless of whether the potentials of the two or more reactions are almost identical or different, they are always established in such a way that the total oxidation current is equal to the total reduction current to keep electric neutrality. For two simultaneous reactions the reaction with the lower (less positive or more negative) equilibrium potential is polarized in the positive direction until the net oxidation current equals the net reduction current of the other reaction, which is polarized from a more positive (higher) potential in the negative direction. When the resistance in the current path is low and the two or more potentials are virtually equal, the common potential is called "mixed potential."

A very characteristic feature of the kinetics of electrochemical reactions is that the relationship between a potential change and the current flow is nonlinear. Fundamentally, for a simple electrochemical reaction involving an activation energy barrier the oxidation current exponentially increases and the reduction current exponentially decreases with a positive potential change. In the field of electrochemical corrosion it is traditional to switch the variables and express the potential change ΔE as a function of the current densities in the form of the "Tafel equation,"

$$\Delta E = E_2 - E_1 = \beta \log i_2 / i_1 \tag{5.8}$$

where β is a Tafel constant. The Tafel constant β is positive for oxidation and negative for reduction. The units of a Tafel constant are volts (or mV) per 10-fold current density change (decade), and usual values range from about 30 to 200 mV/decade. In other words, for an electrochemical reaction controlled by the energy barrier (activation energy), the rate of the reaction increases or decreases by a factor of ten for each 30 to 200 mV of potential change.

When an electrochemical reaction is at equilibrium, the rates of oxidation and reduction currents are equal, and the value per unit area is called exchange current density i_o. A deviation of the potential from the equilibrium value e (single electrode potential) is called overpotential $\eta = E - e$. The Tafel equation as applied to the deviation from equilibrium then can be written as follows:

$$\eta = e - e_o = \beta \log i / i_o \tag{5.9}$$

The Tafel relationship can be conveniently shown in a semi-logarithmic plot called a polarization diagram, where current density is plotted on a logarithmic scale, and potential on a linear scale, as shown in Fig. 5.1. The variations of current densities for oxidation and reduction with the potential in such a diagram are straight lines, intersecting at e and i_o, where the reaction is at equilibrium and the absolute values of the current densities for oxidation and reduction are equal. The slopes of the lines are equal to the values of the Tafel constants, β_a for oxidation and β_c for reduction. The value of the exchange current density i_o depends on both the reaction and the material of the electrode and varies widely for different reactions and materials from about 10^{-14} A/cm^2 to about 10^{-2} A/cm^2.

For a corrosion process with different reactions providing the main oxidation and reduction currents, the balance of the oxidation and reduction currents at the mixed potential can be shown in a polarization diagram similar to the one for a single reaction. The lines intersecting at the mixed potential are those for total oxidation and total reduction currents, which are obtained by summing the currents of all the oxidation reaction and the currents of all the reduction reactions for all electrically connected electrodes surfaces exposed to the electrolyte. Because of the rapid exponential decrease of the reduction current with increasing (more positive or less negative) potential and similar decrease of the oxidation current with decreasing (less positive or more negative) potential, in many cases some of the currents are negligible at certain potentials and can be ignored. Figure 5.2 presents a polariza-
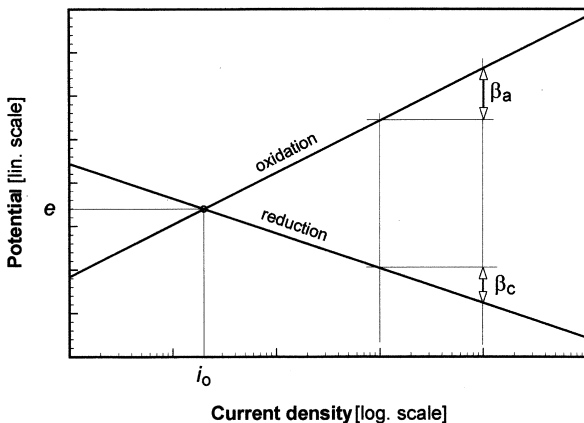

Fig. 5.1 Polarization diagram for an electrochemical reaction under activation energy control, showing the equilibrium potential e, exchange current density i_o, and Tafel slopes β_a and β_c

Fig. 5.2 Evans diagram for metal corrosion showing polarization curves for metal oxidation under activation energy control (anodic line) and reduction of some electrolyte species under activation and diffusion control (cathodic line)
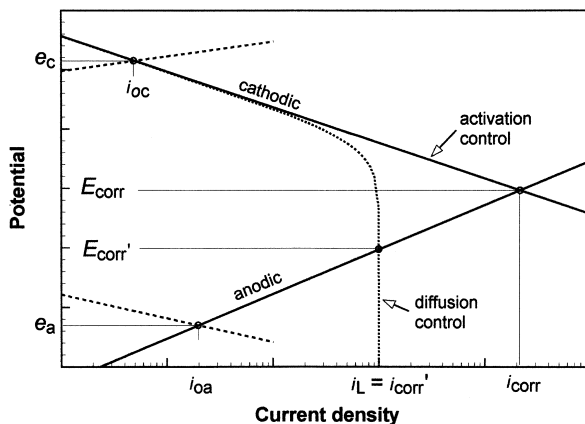
tion diagram, called Evans diagram, for corrosion of a metal where only oxidation of the metal (reaction with e_a and i_{oa}) and reduction of one species in the electrolyte, such as hydrogen ions or dissolved oxygen (reaction with e_c and i_{oc}), play significant roles. The mixed potential, at which the oxidation and reduction currents are equal is called the corrosion potential E_{corr}, and the oxidation current density at this potential is the corrosion current density i_{corr}; the corrosion rate is proportional to i_{corr} according to Faraday's law (5.7).

The use of the corrosion current density i instead of current I in Fig.5.2 suggests that both oxidation and reduction take place uniformly on the same electrode surface. Such a condition would occur exactly only for ideally pure metals and perfect surface conditions in an ideally homogeneous electrolyte. In such a case, although oxidation and reduction take place on atomically different sites at any instant, both reactions are distributed uniformly on a microscopic and macroscopic scale. In more common practical situations the surface chemistry and conditions are more favorable for one or the other reaction on microscopically or macroscopically distinct areas, and the current density values in the Evans diagram are averages over the whole surface area of the electrode. For technically pure metals, for example, impurities may be preferential sites for reduction, and oxidation may occur at a higher rate on some surface irregularities. Even for heterogeneous alloys consisting of chemically different phases it is common to plot average current density values because the total current can be directly measured, while individual current components for different phases would be difficult to determine. It is important to keep in mind, however, that actual current density may be considerable higher than average on very active but small areas of the surface.

When oxidation dominates on one of two dissimilar areas of the metal surface (or two dissimilar electrodes are electrically connected), while reduction dominates on

the other, the former is called the anode, while the latter is the cathode.[7] In this case the electric current flows between two identifiable macroscopic surfaces, as opposed to many local currents flowing between many small sites on the surface, which may have only atomic dimensions. Because of the use of the terms anode and cathode for different electrodes, it is also common to use the terms anodic sites (where oxidation occurs) and cathodic sites (where reduction occurs), to call reactions proceeding in the direction of oxidation "anodic reactions," and those proceeding in the direction of reduction "cathodic reactions," and to use terms "anodic Tafel constant" and "cathodic Tafel constant" for the constants for oxidation and reduction, respectively.[8]

The straight, "Tafel" polarization lines shown in Fig. 5.2 represent the current-potential relationships controlled by the activation energy barriers. A metal or alloy corroding under the conditions of this control is said to be under "active corrosion condition," in "active state" or "actively corroding." Since the oxidation process generates metal ions, which have to be transported away from the electrode for the process to continue, and reduction requires a supply of species transported to the electrode, either process can become controlled by transport of the respective species in the electrolyte rather than by activation energy as the reaction rate increases. This condition occurs more likely for slower moving cathodic species, such as dissolved oxygen than, for instance, hydrogen ions or hydrated metal ions. Diffusion is usually the most important transport mechanism for dissolved species at the electrode surface, and oxygen diffusion control of corrosion of metals with substantial thermodynamic dissolution tendency is quite common in aqueous environments containing dissolved air. When the cathodic reaction is completely diffusion-limited, corrosion is under full cathodic control, the corrosion rate becomes potential-independent and equal to the diffusional flow, and the maximum rate can be expressed as a limiting current density i_L. The effect of mass transport on

---

[7]Formally, the anode is defined as the electrode, at which the direction of the current is from the electrode into the electrolyte, while for the cathode the direction is from the electrolyte into the metal. Since the positive charges move from the metal into the electrolyte during oxidation while electrons flow through the metal to the cathode where they are consumed by the reduction reaction, this definition is in accord with the convention of current flowing in the direction opposite to the flow of electrons.

[8]The use of the terms "anodic" and "cathodic" can be confusing. Since polarization in the positive direction of the potential increases the rate of oxidation, this direction of the potential change is called "anodic direction," and "cathodic direction" is the potential change toward less positive or more negative values. Consequently, scanning the potential in the positive direction is "anodic polarization," and scanning in the negative direction is "cathodic polarization." On the other hand, when two electrodes have different potentials, the one with the more positive potential is "cathodic" or "more noble" with respect to the one with the less positive or more negative potential, and the other one is "anodic" or "less noble" or "more active." It is also important to keep in mind that the terms "higher potential" and "lower potential" always mean more positive (or less negative) and more negative (or less positive) values, respectively.

the cathodic reaction is shown by the dotted line in Fig. 5.2. The illustrated condition shows corrosion under full cathodic diffusion control at E_{corr}' and i_{corr}' $= i_L$.

Figure 5.2 illustrates the types of interaction of oxidation and reduction processes common in corrosion of base metals that are in the state of activity, i.e., when products of oxidation are soluble or not protective. For biomaterials in the human body this condition is uncommon, because it usually results in a relatively high corrosion rate, which is unacceptable for this type of application. The diagram in Fig. 5.2 is important, however, as an illustration of the fundamental principles of electrochemical interaction, allowing prediction of the effects of some variables. It shows, for instance, that an increase in the oxidation rate, perhaps as a result of a change in the surface condition of the metal while the cathodic kinetics remains the same, causes a shift of the oxidation line in the diagram to the right, and thus an increase in i_{corr} and a shift of the corrosion potential toward a lower value.[9] An increase in the rate of reduction under activation energy control (a shift of the cathodic line to the right) also increases i_{corr} but shifts E_{corr} to a higher value. Under full cathodic diffusion control an increase in i_L (perhaps due to an increase in the flow rate of the electrolyte) is associated with an increase in both i_{corr} and E_{corr}, while a change in the oxidation kinetics under cathodic diffusion control changes E_{corr} but i_{corr} remains constant at i_L. Thus an observed change in the corrosion potential may or may not indicate of a change in the rate of corrosion, and a correct interpretation requires some knowledge of the type of kinetics involved.

The successful use of base metals, such as chromium, cobalt, nickel, titanium, niobium and zirconium in the human body is due to the phenomenon of passivity. In spite of some conflicting theories of the passivation process, there is little doubt that in a passive state the above metals in an aqueous electrolyte are covered with a thin and very protective film of a stable, poorly soluble oxide or hydroxide. The stability of oxides and hydroxides is typically a function of the pH. In water the oxides form electrochemically and their existence is thus also potential dependent. A convenient illustration of a region of stability of an oxide as well as of other potential or pH dependent species is a potential-pH (E-pH) diagram, commonly called a Pourbaix diagram.[10] Figure 5.3 shows a simplified E-pH diagram for titanium, one of the metals of special significance for medical implants. A collection of E-pH diagrams for most metal/water systems can be found in an atlas [5].

Pourbaix diagrams are constructed based on the thermodynamic data for various chemical and electrochemical reactions possible in the given metal-electrolyte system. They can be interpreted as showing regions of thermodynamic stability of various species. If the stable specie is a non-ionized metal, the region is one of immunity, because under those pH and potential conditions the metal is thermodynamically stable and thus cannot corrode. Regions of stability of soluble species

---

[9]It is assumed that such a change would result in an increase in the effective exchange current density, while the anodic Tafel constant remained the same.

[10]Prof. Marcel Pourbaix of Belgium devoted his life to the construction and application of potential-pH diagrams.

Fig. 5.3 Potential-pH
(Pourbaix) diagram for the
system titanium/water at
37 °C. Based on data in [5]
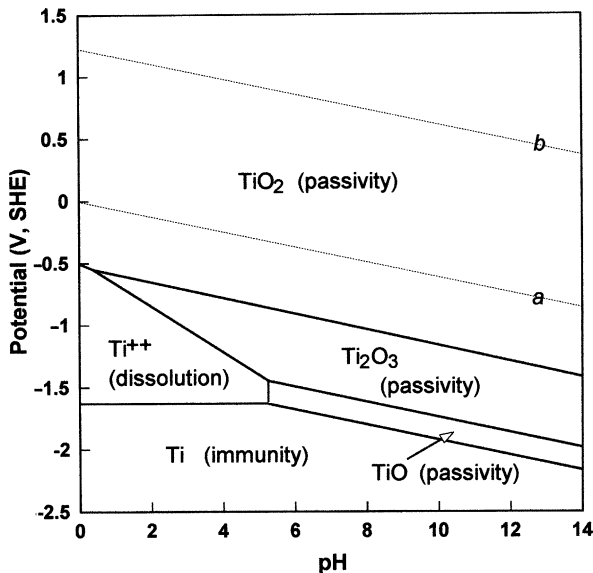
generally are regions of corrosion, because there is a thermodynamic tendency for
dissolution. In the regions of stability of an oxide (or hydroxide) passivity is possi-
ble, assuming that the oxide is poorly soluble and non-porous and thus protective.

Although a Pourbaix diagram thus provides a "map" of regions of immunity,
corrosion and passivity at various pH and potential conditions, it must be kept in
mind that for a specific corrosion condition the pH of the electrolyte is probably
more or less constant, and the above regions are reduced to ranges of potential values
along a vertical line in the diagram. The actual corrosion condition is then the result
of a combination of thermodynamic and kinetic factors that determine the actual
corrosion potential, the diagram showing only the former and not the latter factors.

Traditionally, each Pourbaix diagram for metal-water systems also includes two
lines, marked (a) and (b) and representing the equilibria for reactions (5.3) and (5.5),
respectively. Both reactions are pH-dependent with the same slopes in the diagram,
e = e^o − 0.061 pH at 37 °C. These two reactions, proceeding in the direction of
reduction, are the two most important cathodic reactions for corrosion in aqueous
electrolytes, such as body fluids. Since the corrosion potential must establish itself
below the equilibrium potential of the cathodic reaction (so that the reaction pro-
ceeds in the direction of reduction), the lines (a) and (b) show the theoretical max-
ima for the corrosion potentials in oxygen-free water (line a) and oxygen-saturated
water (line b).

When a metal without a surface oxide is exposed to water not containing its
ions there is a very high tendency for oxidation, and the potential rapidly increases
from the very low value toward the potential of the available reaction providing
reduction. If the mixed potential of the balance of oxidation and reduction stabilizes

Fig. 5.4 Evans diagram for metal corrosion showing polarization curves for a passivating metal and activation energy controlled reduction of electrolyte species (lines c_1, c_2 and c_3)
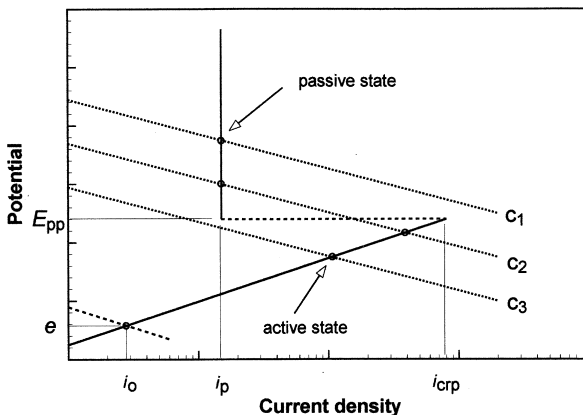
in the field of dissolution in the Pourbaix diagram, active corrosion occurs. If the potential reaches the field of a stable, protective oxide, the metal becomes passive.

The variations of the corrosion current density with potential again can be depicted in a polarization or Evans diagram. If the potential change along a vertical line in the E-pH diagram for the given pH value passes through the field of dissolution before reaching the field of passivity, the diagram shows a classical "textbook type" anodic polarization curve for passivation, illustrated in Fig. 5.4. The anodic line, in principle starting from the equilibrium conditions, shows a segment of Tafel behavior through the field of dissolution (active region) where the current density for oxidation increases exponentially with the potential. It then reaches a peak current density called critical current density for passivation (i_{crp}) at the boundary between the regions of dissolution and stability of the oxide at the primary passivation potential E_{pp}, and drops rapidly to the current density in the passive state (i_p). The current density remains more or less constant in the region of passivity. This type of polarization behavior is experimentally observed, for instance, for iron in acidic solutions, and theoretically should also be observed for titanium (see Fig. 5.3).

Alternatively, and more commonly for metallic biomaterials with a high passivation tendency, as the potential increases the metal passivates without any significant active dissolution, the potential-pH conditions moving from the region of immunity to passivity. The result is an experimental anodic polarization curve lacking the typical "passivation" peak, but showing the potential-independent region of passivity with a low value of i_p. Moreover, since passivating metals have a high tendency for oxidation they react easily with atmospheric oxygen, so that the surface usually becomes covered with an oxide film even before immersion in the electrolyte. This film may dissolve if the kinetics of the anodic and cathodic reactions maintain the E-pH condition in the region of dissolution. If, however, after immersion the potential quickly passes through the active region to passivity, the characteristics of the film may change but active corrosion does not necessarily occur.

Similarly to the behavior of an actively corroding electrode illustrated in Fig. 5.2, the corrosion potential for a passivating metal establishes itself at a value resulting in the equality of oxidation and reduction currents. This is shown in Fig. 5.4 as the intersection of the cathodic polarization line with the anodic line. When a true passivation peak exists, the reduction current must exceed the oxidation current at the primary passivation potential E_{pp} for spontaneous passivation to take place. For a theoretically uniform distribution of the current this means $i_{red} > i_{crp}$ at E_{pp} for the electrode to exist in a state of stable passivity (line c_1 in Fig. 5.4). If $i_{red} < i_{crp}$ at E_{pp} (the cathodic polarization line intersects the anodic line in both the dissolution segment and the region of passivation, line c_2 in Fig. 5.4), the electrode is in a state of unstable passivity and may be either active or passive, or even switch from one state to the other.[11]

In the region of passivity the current density is very low (typically less than $1 \, \mu A/cm^2$ for implant alloys in actual or substitute body fluids) and often approximately potential-independent. Since the driving force for the oxidation reaction increases with increasing potential, the potential independence of i_p indicates that the passive film becomes more protective with increasing potential, mostly by increasing thickness. However, there are many cases where i_p varies with the potential in the passive region, either increasing or decreasing. Still, this potential dependence is always much weaker than the exponential variation for Tafel behavior in the active state. The oxide films, especially on solid solutions or other multicomponent phases can be quite complex in structure and chemistry, and the changes in the properties also may be potential-dependent.

The existence of measurable anodic current density in the passive state even under steady state conditions at a constant corrosion potential indicates that oxidation continues. The rate of oxidation equals the rate of chemical dissolution of the oxide, thus keeping the oxide thickness constant. Therefore, implants in passive condition continue to release metal ions, albeit at a low rate. Because of the variation of the solubility of the oxide with the pH of the solution, i_p shows a similar dependence.[12]

For some polyvalent metals (e.g., Cr), soluble species (CrO_4^{2-}) become thermodynamically stable as the valence changes at potentials above those for a stable oxide, (from three to six for Cr). This change results in another region of active dissolution, a phenomenon called transpassivity.[13]

---

[11]In this case the cathodic line also intersects the anodic line in the segment of a current density drop from i_{crp} to i_p. Since this is a region of "negative resistance" (positive current decreasing with positive potential increase), this condition cannot be sustained.

[12]Most oxides providing passivity for metallic biomaterials exhibit increasing solubility with increasing acidity (decreasing pH) below neutral conditions, and some also under alkaline conditions. While alkaline conditions are physiologically of little importance, the increase in solubility with increasing acidity plays a major role in localized forms of corrosion.

[13]Lately, the term "transpassivity" frequently has been applied to any increase in activity for passivating metals at potentials above the passive region, such as when localized breakdown of passivity occurs, resulting in pitting.

5.3 Forms of Corrosion of Metallic Biomaterials

The principles of electrochemical corrosion discussed in Sect. 5.2 apply universally in all cases of corrosion of metallic biomaterials. In the individual forms of corrosion, however, specific conditions may make the corrosion process much more complex, for instance because of the interaction of areas of dissimilar electrochemical properties, localized variations in the solution chemistry, interaction with mechanical effects, etc. Following are the main forms of electrochemical corrosion relevant to biomaterials. In practical situations, several forms may occur at the same site, e.g., pitting combined with galvanic corrosion, crevice corrosion with wear corrosion, etc.

5.3.1 Uniform Dissolution

The corrosion process is most easily analyzed when the oxidation and reduction reactions take place uniformly on a homogeneous metal surface. Completely uniform corrosion, however, is an ideal condition that does not occur in practice. Ideally, oxidation and reduction would occur randomly on surface atoms that were identical in electrochemical properties so that the oxidation and reduction sites were interchangeable and allowed instantaneous and continuous redistribution. In reality, defects of the crystal structure (vacancies and interstitials, dislocations, impurities) and segregation of elements result in an inhomogeneous distribution of electrochemical activity on the surface even for very pure metals and solid solutions. In many cases, impurities precipitating as second phases play an important role as preferential sites for reduction, oxidation or both. The conditions are even more complex for multiphase alloys, where some phases may be clearly anodic and other cathodic. In many cases averaging of the corrosion parameters over the nominal surface provides a reasonable picture of the electrochemical behavior at the corrosion potential or under polarization and yields useful values of corrosion parameters. However, surfaces of previously smooth metals usually show distinct roughness after active state corrosion. Uniform corrosion of single phase materials in passive state usually preserves smoothness of the surface, both because of the low corrosion rates and because the surface reaction is chemical dissolution of a relatively uniform film.

An aspect of uniform corrosion that usually receives little attention is selectivity of dissolution in alloys.[14] Electrochemically more active components of a solid solution have a tendency to dissolve preferentially, resulting in their depletion in the surface layer and enrichment in more stable elements. The commonly observed increase in the corrosion potential with time is usually attributed to the growth of an

---

[14]Selective leaching as a specific form of corrosion, often called by a generic term "dezincification" (leaching of zinc from brass) is a sometimes serious form of industrial corrosion, resulting in loss of mass and a porous layer on the surface of alloys containing elements of very different electrochemically activity. Such a severe degradation is normally not observed in much more corrosion resistant biomedical alloys.

oxide film or formation of other corrosion products, but selective dissolution also may contribute to the effect.

From the practical viewpoint the most important corrosion parameter in uniform corrosion is the (nominally uniform) corrosion rate. For biomaterials the corrosion rate must be very low, and is usually determined by sensitive electrochemical techniques as corrosion current density. Physiologically, on the other hand, the more important value is the release rate of specific elements in view of the diversity in toxicity and generally dissimilar human body tolerance for different elements. Although the corrosion current density can, in principle, be converted to the mass loss using Faraday's law (5.7), with the exception of pure metals the determination is considerably less straightforward because the elements do not necessarily dissolve at relative rates equal to the atomic ratios in the alloy.

There are several methods of determination of the average corrosion current density, most of them based on calculation from the value of the polarization resistance R_p. The polarization resistance is defined as the slope of the polarization curve at zero current density, $R_p = (dE/di_n)_{in = 0}$, where i_n is the net current density, which is the difference between the anodic (oxidation) and cathodic (reduction) current densities.[15] The corrosion current density is then calculated using (5.10):

$$i_{corr} = \frac{\beta_a \cdot \beta_c}{2.3 R_p (\beta_a + \beta_c)} \tag{5.10}$$

5.3.2 Galvanic Corrosion

In galvanic corrosion the degradation is enhanced by a combination of two or more dissimilar metals that touch each other or are otherwise in electronic contact. "Galvanic interaction" also occurs between dissimilar phases in multiphase materials, but the term "galvanic corrosion" usually is used for structures consisting of two or more components made of different materials, such as ball and stem of a total hip implant.

Galvanic corrosion occurs when the corrosion potentials (E_{corr}) of the individual materials (electrodes), when not connected with each other, are different. When an electric connection is established the potential difference causes a current to flow, and the current polarizes the individual electrodes toward a common "galvanic potential" (E_g). Electrodes with the lower individual potential ($E_{corr} < E_g$) are polarized in the positive direction (anodically), their oxidation rates exceed the reduction rates (they become anodes) and they are likely to corrode more than in the absence of the galvanic contact. Those with the higher individual potential ($E_{corr} > E_g$) are

---

[15]The measured net current $I_n = I_{oxidation} - I_{reduction}$ assuming both oxidation and reduction currents to be positive values. It is more common to take the reduction current as negative, in which case the net current is the sum of oxidation and reduction currents.

polarized in the negative direction, become cathodes, and may to some extent be protected.

The electrodes of the galvanic system would reach the same galvanic potential E_g only if the total resistance in the path of the current between them were zero. Since the current must flow not only through the metal contact but also through the ionically conducting solution, the resistance is not necessarily negligible, except near the interface between the contacting metals. The result is that there is often a significant variation in the potential from the interface to the more distant areas. For the anode(s) of the system it means that the increase in the corrosion rate tends to be highest near the interface and decrease with a distance from it.

Overall, for the whole galvanic system, the oxidation current must still equal the reduction current, and the anodic current increase on the anode(s) must equal the cathodic current increase on the cathode(s). For electrodes of unequal area this means a higher current density increase for a small electrode than for a large one. Thus a combination of a large cathode with a small anode is potentially more damaging than the opposite case.

The actual severity of galvanic corrosion also depends, however, on the polarization behavior of the electrodes, especially of the anode(s). The essential effect of galvanic interaction is an increase in the potential of the anode, and if this increase does not result in an increase in the oxidation rate, there is no galvanic corrosion. This is true for a passivating metal in the region of potential-independence of the current density. Coupling with a more noble electrode may increase the corrosion potential, but if E_{corr} remains in the region of passivity, the dissolution rate increase little or not at all. Since the corrosion potential cannot exceed the equilibrium potential of the reduction reaction (or the highest one if several reactions participate), galvanic interaction has no detrimental effect if there is no current density increase on the anode at up to this potential.

A special case of galvanic interaction, called "oral galvanism," may occur in the oral cavity if restorations or appliances made of dissimilar metals are in contact continuously or intermittently (during biting). The electric current circuit includes a path through the tissues and the current flow may affect nerves, causing pain. The conditions for galvanic interaction are much more common for metals in the oral cavity than for implants because of the frequent use of both noble and base metals, as in a contact between a gold crown and a dental amalgam filling. Still, severe continuing galvanic corrosion is relatively rare because the galvanic effect rapidly diminishes when an electric resistance is placed in the current path, causing an *IR* (current times resistance) potential drop and thus reducing the potential increase on the anode. When two dissimilar restorations are placed in contact the initial galvanic current may be high (and cause pain), but the corrosion reactions often result in the formation of nonmetallic corrosion products at the interface that increase the resistance. Intermittent contacts of upper and lower jaw restorations are more likely to cause high current spikes, but even there the oxide formation on the anode is likely in time to reduce the effect.

Galvanic interaction also may have beneficial effects. In industrial practice one type of corrosion protection is achieved by using a "sacrificial anode" made of a

cheap base metal as "cathodic protection" (the protected part is made the cathode of the cell) to protect the more valuable structure. This type of protection is normally not applied to biomaterials because the intensive corrosion of the anode would be unacceptable within the human body. However, a mild galvanic interaction that increases slightly corrosion of one part and decreases dissolution of the other part may be beneficial if the elements dissolving from the protected part (cathode) are more harmful (such as Ni) than those from the anode (such as Ti). Galvanic effect may also be used to ensure stable passivity by increasing the corrosion potential well above the primary passivation potential (E_{pp}), but this must be done carefully to avoid the danger of passivity breakdown (see Sect. 5.3.4). Conversely, the danger of passivity breakdown my be lessened by coupling with a material with lower E_{corr} (anode of the galvanic cell), on the condition that this will not result in unacceptable corrosion of the anode or shift the potential of the cathode into the region of unstable passivity or activity.

5.3.3 Concentration Cell Corrosion

Similarly to galvanic corrosion, concentration cell corrosion involves a potential difference between macroscopically separate electrode areas causing a flow of the current and changes in the polarization of the areas. While in galvanic corrosion the potential difference results from a difference in the electrochemical properties of the electrode surfaces, in concentration cell corrosion there is a difference in the concentration of species in the electrolyte involved in the corrosion reactions on the electrode areas. The equilibrium potential of an electrochemical reaction is a function of the ionic concentration according to the Nernst equation (5.4), and a concentration difference generates a potential difference. Differences in concentration of both anions and cations can create a potential difference; since hydrogen ions, dissolved oxygen and metal ions are the most important reactants or products of corrosion reactions in aqueous solutions, hydrogen ion, oxygen and metal ion concentration cells are most common.

In a metal ion concentration cell the electrode area exposed to a lower concentration of the ions has a higher tendency to dissolve and lower individual corrosion potential. When connected with the area of higher ionic concentration it is polarized toward the higher potential of the other area, becomes the anode and its corrosion rate increases. Differences in the concentration of cathodic reactants, such as hydrogen ions or dissolved oxygen, on the other hand, do not result directly in a difference in the corrosion rate of the regions, because under short-circuit conditions the corrosion potential is the same and thus both areas are polarized to the same potential, resulting in the same anodic overpotential for the same oxidation reaction. Such a concentration cell may cause corrosion acceleration of one of the areas indirectly, as a result of chemistry changes at the electrode surface. Both reduction of hydrogen ions and reduction of dissolved oxygen involve consumption of hydrogen ions, and the areas where these reactions are more intensive exhibit a pH increase. In the

industrially important case of an oxygen concentration cell (situation called "differential aeration"), corrosion of steel is accelerated in the oxygen depleted region next to the air saturated region because of the passivation of the latter due to the pH increase, in spite of the fact that an oxygen concentration increase by itself normally results in a higher corrosion rate.

Although some concentration differences are common, concentration cell corrosion of metallic biomaterials on unshielded surfaces is rare and usually mild. As seen from the Nernst equation, a difference in metal ion concentrations by a factor of 2 and 10, for the most common divalent ions, results in potential differences of only about 10 and 30 mV, respectively. Moreover, similarly to galvanic corrosion, the strength of the interaction diminishes with distance because of the electric resistance in the current path, but unlike the galvanic situation at the interface between different solid phases, a sharp transition of one concentration to another within the electrolyte cannot be sustained. Body fluids are well buffered against large pH changes, but even for the larger pH variations due to post-surgery trauma the concentration effect is rather mild for the common highly passivating alloys. Concentration cells exists in the oral cavity because of the exposure of restorations to both the oral and tissue fluids, but the corrosion effect is again relatively small. Concentration cells, however, play a important roles in the forms of occluded cell corrosion, which are discussed next.

5.3.4 Pitting and Crevice Corrosion

Except for degradations combining corrosion with mechanical effects, pitting and crevice corrosion are the most destructive forms of corrosion of metallic biomaterials. Both are types of localized attack occurring in what is called "occluded corrosion cells," i.e., where solution at the site of the attack becomes different and much more aggressive than the solution outside those sites because of the restricted transport of ionic species in and out of the site. Both pitting and crevice corrosion are most common and severe for passivating metals exposed to solutions containing chloride ions, where small regions of actively corroding metal become anodes, surrounded by large cathode areas of slowly corroding passive surface.

The mechanism of corrosion in active pits and crevices is virtually the same, and the main difference between the two forms is in the initiation, pitting requiring a breakdown of the passive film to establish the occluded condition. In the absence of a mechanical action, such as scratching, the breakdown is usually electrochemical. Such a breakdown of passivity is potential dependent and occurs when the electrode potential exceeds a critical value, called breakdown potential (E_b) or pitting potential (E_{pit}). In spontaneous corrosion breakdown occurs when the corrosion potential, which, for passive metals, usually increases with time, reaches the breakdown potential. Thus a material with a lower breakdown potential is more susceptible to pitting initiation, assuming otherwise the same anodic and cathodic kinetics. Since the corrosion potential cannot exceed the equilibrium potential of the cathodic

reaction, this latter potential can be considered the minimum value of E_b for immunity against pit initiation.

There are conflicting theories for the mechanism of the breakdown, but there is evidence that chloride ions are involved (other halogens and some other aggressive ions may play this role in some solutions), either by making the oxide film locally soluble, slowing down repassivation of breaks in the film, or some other mechanism. Pitting may also be initiated by surface impurities, such as sulfide inclusions in stainless steels. Once the metal becomes locally active the corrosion reactions cause local changes in the electrolyte chemistry, which prevent repassivation. The most important of these changes is acidification, which results from a chemical reaction of metal ions with water, called hydrolysis. For example, for chromium the hydrolysis reaction may be as follows:

$$Cr^{3+} + 3H_2O = Cr(OH)_3 + 3H^+ \tag{5.11}$$
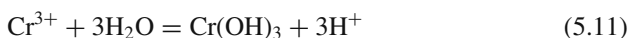
The reaction consumes the dissolving metal ions, a solid corrosion product precipitates, and hydrogen ions are generated, thus causing the pH to drop. The high acidity generated by the hydrolysis reaction prevents repassivation because of the higher solubility of the oxide at lower pH. The precipitated solid is non-protective, and may impede transport of ions from and to the active surface.

The hydrolysis reaction starts when the concentration of the metal reaches a critical value, which depends on pH and on the value of the solubility product K_s. For metals with poorly soluble oxides and hydroxides, which also form stable passive films, acidification thus starts generally earlier than for metals with less effective passivation. The pH drop continues and metal ion concentration increases (because of higher oxide solubility at lower pH) until either a balance of transport in and out of the pit is reached, or the metal concentration becomes high enough for precipitation of a salt. Values of pH as low as 1.5 have been found experimentally in active occluded corrosion cells.

Contributing to acidification is local depletion in dissolved oxygen, which normally consumes hydrogen ions (5.6), because oxygen diffuses slowly and cannot be replaced sufficiently quickly at the site of dissolution. On the other hand, negative chloride ions from outside the pit are attracted by the positive charge of the metal ions within the pit, further impeding passivation.

If the whole surface of the electrode became active, the corrosion potential would drop from the region of passivity to the region of activity. Since initiated pits are only a small fraction of a total area of the electrode and are surrounded by a large passive surface, the potential does not change that much, resulting in high anodic polarization for the active area within the pits and a very high corrosion current density, often three or more orders of magnitude higher than outside the pits. The consequence is a high rate of growth of the pits. The growth rate may decrease with time, however, as the resistance in the current path between the outside and the bottom of the pits increases and results in a potential drop, which reduces the anodic polarization of the active surface.

The maintenance of an active pit requires a dissolution rate generating enough ions to keep the pH continuously low by hydrolysis, while some hydrogen and metal ions escape from the pit and some oxygen diffuses inside. Since the dissolution rate in the active state is a function of the potential, lowering the potential decreases the dissolution rate, until eventually the active surface within the pit cannot be sustained and the surface repassivates. The potential value at which this happens has been called "protection potential" (E_{prot}) or "repassivation potential" (E_{rp}).

Crevice corrosion involves the same mechanism as pitting, except that the occlusion results from shielding of a part of the surface. Crevice corrosion cells are thus common where two metal parts or a metal and a nonmetal part are in contact that allows access of the electrolyte between them. For medical implants some common sites are between the head of a bone screw and a bone plate, or between a ball and stem of a two-component hip implant. For a newly formed crevice, or a crevice newly exposed to the solution, there is a period of inactivity, during which the occluded cell conditions are getting established, i.e., depletion of dissolved oxygen, accumulation of metal ions by a slow dissolution of the oxide film, followed by metal ion hydrolysis, pH drop, accelerated dissolution of the oxide, and increasing concentration of chloride ions. The active surface within the crevice is coupled with the usually larger external passive surface and is thus anodically polarized, which results in a high corrosion rate. Similarly to pitting, lowering the potential below a critical value reduces the dissolution to a rate, at which the active corrosion conditions within the crevice are not sustained.

Crevice corrosion is generally more severe for narrower crevice gaps, because the transport restriction is more effective. On the other hand, since the important polarization effect of the external surface requires a current flow, a too narrow crevice, or a crevice filled with corrosion products may involve a potential drop that reduces the crevice effect. For the same reason, dissolution is usually most severe inside but near the mouth of the crevice. The crevice corrosion rate may be increased or decreased by a galvanic interaction if the two contacting components are made of dissimilar materials, resulting in a potential change.

Ranking of the most common orthopedic implant alloys based on the values of the breakdown potential in vitro shows increasing corrosion resistance in the order Type 316 LVM stainless steel, Co-Cr alloys and Ti alloys, which is in agreement with the clinical experience. In the area of dental restorative materials, high-copper dental amalgams, which do not exhibit breakdown of passivity in vitro, also show less corrosion than low-copper amalgams, which do.

5.3.5 Environment Induced Cracking

When a metal part is subjected to a mechanical load in a corrosive environment, cracking may occur as a result of a combination of those conditions. If the load is mostly static the failure is usually called stress corrosion cracking (SCC), while fluctuating loading may result in corrosion fatigue (CF). The term hydrogen induced

cracking (HIC) is used when the failure can be attributed to the entry of hydrogen atoms into the metal.

These forms of failure are of great industrial importance and there are many conflicting theories and models of the mechanisms. While some researchers seek and favor a single mechanism for different alloy/environment combinations, others find evidence for different mechanisms in different systems. Dissolution models of SCC and CF attribute the crack extension to anodic dissolution of a strained metal at the crack tip, while mechanical models describe the process as a mechanical failure facilitated in various ways by the environment, such as by weakening of the atomic bonds at the crack tip due to adsorption of species from the electrolyte, formation of brittle films or other brittle phases, various effects of the electrolyte on dislocation generation, interaction and movements, etc. Hydrogen induced cracking may be due to absorbed hydrogen affecting the metal lattice properties or dislocations, or formation of brittle hydrides. An increase in the electrode potential (anodic polarization) often is observed to increase the rate of crack propagation, which is predicted by dissolution models but is also in accord with some mechanical models. When cracking is accelerated by cathodic polarization, which increases the supply of reduced hydrogen ions at the metal surface, it suggests that HIC is involved.

Stress corrosion cracking has been observed in specific metal/environment systems, is mostly confined to alloys, and most common metallic biomaterials are resistant to SCC. Stainless steels are well known to suffer from SCC in hot chloride solutions, but there is no conclusive evidence that SCC occurs for implant-grade stainless steel in body fluids at body temperature. Some titanium alloys are highly susceptible to SCC, but commercially pure Ti and implant Ti alloys, such as Ti-6Al-4 V, do not appear to be susceptible in body fluids. On the other hand, there is some concern regarding possible HIC of Ti and its alloys.

In contrast to SCC, corrosion fatigue is not confined to specific systems, and acceleration of fatigue crack propagation by corrosive environments is common. Moreover, since metal fatigue is highly sensitive to surface conditions, initiation of fatigue (and CF) failure may be facilitated by localized corrosion, such as pitting. Thus the corrosive environment may decrease the threshold stress intensity or the fatigue crack propagation rate, or both.

5.3.6 Intergranular Corrosion

In intergranular corrosion the dissolution is confined to a narrow region along the grain boundaries, and the attack may not be clearly visible until failure occurs due to the loss of mechanical integrity. A severe susceptibility of the grain boundary regions may be due to the precipitation of corrosion susceptible phases, or to depletion of elements providing corrosion protection along the boundaries due to precipitation of phases rich in those elements. The latter is the cause of intergranular corrosion of "sensitized" stainless steels (and some Ni-Cr alloys), sensitization being due to precipitation of chromium-rich carbides along grain boundaries, and

may occur during heating into a specific temperature range, for instance during welding. This dangerous form of corrosion of stainless steel may be avoided by proper heat treatment (dissolving existing carbides at high temperature and rapid cooling to prevent precipitation) or by alloying with other strong carbide forming elements, but in stainless steel for implants normally it is prevented by using steels with a very low carbon content.

5.3.7 Wear-Corrosion, Abrasion-Corrosion, Erosion-Corrosion, Fretting

In these forms of corrosion the mechanical action destroys the protective films, allowing active corrosion of the unprotected metal surface. If only part of the surface is active the coupling of the active surface with neighboring passive areas intensifies corrosion. Wear-corrosion involves parts in a friction contact with relative movement, while abrasion-corrosion involves abrasive particles between the contacting surfaces, which may be the debris from the wear process. In erosion-corrosion the destruction of the surface film is due to the flow of a liquid, mostly one containing abrasive particles. Fretting is a special case of wear-corrosion, in which there are only small relative movements between contacting parts that are essentially in a static mutual relationship. Thus wear- and abrasion-corrosion may occur in articulating implants, such as on a metal ball of a hip joint in contact with the polyethylene cup, and fretting may take place between the ball and stem of multi-component hip implant. Wear- and abrasion-corrosion occur on the mastication surfaces of dental restorations. Erosion-corrosion may cause degradation of metallic heart valves.

Wear- and abrasion-corrosion and fretting are often electrochemically quite complex. The narrow gap between contacting surfaces creates crevice conditions, so that the destructive effect of friction and abrasion on the protective surface film is superimposed on the corrosion mechanism in the occluded cell. Galvanic interaction may be involved for contacting metal parts, resulting in either acceleration or mitigation of the attack. For intermittent wear and abrasion events the repassivation ability of the metal is an important factor.

5.4 Corrosion Testing of Metallic Biomaterials

In corrosion research a large number of tests that can be used are often designed for the purpose of investigation of corrosion mechanism and determination of parameters for specific materials or classes of materials. Most corrosion tests are performed in the laboratory (in vitro); tests in animals or humans (in vivo) are mostly limited to examination of retrieved implants and dental restorations or surrounding tissues, and only rarely to determine details of the corrosion behavior and parameters. The discussion here is limited to laboratory tests used for general evaluation of metallic

implant and dental materials with the purpose of predicting their performance in the human body.

The choice and design of such laboratory corrosion tests is based on a number of scientific and practical considerations, including the following: (1) the type of corrosion in the human body environments is almost exclusively electrochemical; (2) metallic biomaterials are mostly highly corrosion resistant, resulting in very low corrosion rates requiring very sensitive testing techniques; (3) the chemistry of the human body fluids is relatively stable for specific application sites; (4) the rate of release of metallic ions and their nature is an important factor in the suitability of materials for use in the human body. As a result of these considerations the evaluations tests have the following general characteristics. First, electrochemical tests, which are highly sensitive, are used to determine the major parameters of the corrosion behavior, and immersion tests, which include solution analysis for metal ions, are used to determine qualitatively and quantitatively the released metal ions. Second, tests are performed using relatively simple solutions, omitting components of minor corrosion significance, but controlling the most important characteristics of the target environment. The conditions and procedures for the most common tests have been standardized [1].

Because of the simplification involved in the choice of the environment and some of the test conditions, such as short exposure time, laboratory corrosion tests of biomaterials are mostly ranking tests, predicting a relative, rather than absolute performance. Their usefulness is based on the clinical experience with some traditional medical implant and dental restorative materials. If a tested material yields corrosion results similar or better than materials with proven, acceptable performance in the human body, there is reason to believe that the new material will perform equally well or better clinically. Eventually, this assumption must be confirmed in animal and human tests, but the laboratory tests are an important way of eliminating unsuitable materials before the clinical test phase is reached.

Most metallic biomaterials are passivating alloys and electrochemical tests are capable of yielding a substantial number of parameters related to the passivation and passivation breakdown. For the purpose of predicting the corrosion performance, however, in vitro electrochemical corrosion tests focus mainly on the resistance to the loss of passivity. When active corrosion occurs locally on otherwise passive surface it results in a severe corrosion attack, usually in the localized form of pitting or crevice corrosion. The fundamental electrochemical resistance to the localized loss of passivity is related to the passivity breakdown that occurs when the potential exceeds a critical value, the breakdown potential E_b. The breakdown potential is commonly determined using an electronic potentiostat[16] and subjecting the tested sample to increasingly higher potentials until a sharp increase in the anodic current

---

[16]Electronic potentiostat is a device for controlling the potential of a tested electrode by flowing the necessary current through it. It is usually used with three-electrode corrosion test cells, in which the tested sample is a "working electrode," the potential of which is measured with respect to a reference electrode, and a relatively inert auxiliary "counter electrode" is used to complete the circuit for the current flow through the electrolyte.

signals initiation of active corrosion. This potential scanning technique is described, for instance, in the standard test method ASTM 2129, "Conducting Cyclic Potentiodynamic Polarization Measurements to Determine the Corrosion Susceptibility of Small Implant Devices." The same procedure can be used for testing laboratory specimens of any passivating biomaterial. The standard test solution according to ASTM 2129 is a phosphate-buffered saline at pH 7.2, and is used at 37 °C and saturated with pure nitrogen (deaerated). This solution is appropriate for testing of biomaterials that are to be exposed to blood or tissue (interstitial) fluids. The standard also describes a synthetic bile solution for testing of biliary implants, and the same procedure may be used with other body fluid substitutes. The results are somewhat dependent on the choice of the potential scanning speed, the most common values being 0.1667 or 1 mV/s.

The ASTM standard F2129 also describes the determination of the "protection" or "repassivation" potential E_{prot}, which is obtained by reversing the scan after the passivity breakdown occurs, and is defined as the potential at which the current density drops to the value previously observed during the forward scan. Although this determination is theoretically valid, the results must be treated with caution. The reversal of the solution chemistry changes that occur during initiation of pitting depends on the extent of the damage that occurs after the breakdown, and E_{prot} generally decreases with increasing current at scan reversal. Even at the same current value at reversal, however, E_{prot} also depends on the size and number of initiated pits, since many small pits repassivate more easily than few large ones. The observed value of E_{prot} also depends on the breakdown potential E_b, because pits initiated at a higher potential grow faster than those initiated at lower potentials. Still, cyclic polarization (potential scan with a scan reversal) is useful because the presence of a hysteresis loop, i.e., the current density returning to the pre-breakdown value at a potential $E_{prot} < E_b$, is evidence that the current increase was due to localized passivity breakdown rather than some oxidation process in the electrolyte or general reactivation.

Another way to determine the repassivation potential of alloy specimens (rather than actual devices) is described in ASTM F746 "Standard Test Method for Pitting or Crevice Corrosion of Metallic Surgical Implant Materials." A cylindrical specimen is equipped with a plastic collar creating an artificial crevice, breakdown of passivity is induced by application of a high potential, and the potential is lowered in steps until repassivation occurs. Localization of the attack inside the artificial crevice is intended to lessen the problem of differences in active cell size, but the exposure time may not be always sufficient for creation of crevice conditions.

Corrosion current density is an important parameter both with respect to the damage to the device and release of metal ions into the body, but its determination is often problematic. Typically i_{corr} is substantially time-dependent, for passivating alloys mostly decreasing with time of exposure, and the test results thus vary with timing of the test. Since the directly measurable current at E_{corr} is zero, indirect methods, such as extrapolation of the Tafel lines to E_{corr} or calculation from polarization resistance, must be used. A standard test method for conducting potentiodynamic polarization resistance measurements is described in ASTM G59.

Most serious degradation of metallic implants often is due to a combination of corrosion with mechanical forces. Devices subjected to fluctuating loads, such as many orthopedic implants, may fail by corrosion fatigue. A standard practice for corrosion fatigue testing of metallic implant materials is described in ASTM F 1801. Modular devices, such as hip implants consisting of a separate femoral head and stem may suffer fretting corrosion and are tested as per ASTM F 1875. Fretting corrosion also may occur at the contact between bone screws and plates, and the susceptibility to this degradation is tested as per ASTM F 897.

References

1. ASTM. Annual Book of ASTM Standards.ASTM International: West Conshohocken, PA, 2008.
2. Burke DR. The Composition and Function of Body Fluids. 3rd edn, C.V. Mosby: St. Louis, 1980.
3. Lentner C, ed. Geigy Scientific Tables, Vol. 1: Units of Measurement, Body Fluids, Composition of the Body, Nutrition, CIBA-Geigy: Basle, 1981.
4. Lentner C, ed. Geigy Scientific Tables, Vol. 3: Physical chemistry, Composition of Blood, Hematology, Somatometric Data. CIBA-Geigy: Basle, 1984.
5. Pourbaix M, ed. Atlas of Electrochemical Equilibria in Aqueous Solutions. NACE: Houston, 1974.
6. Stephen D, Cramer SD, and Covino BS Jr, eds. ASM International. Metals Handbook, Vol. 13A: Corrosion, ASM International: Materials Park, OH, 1987.
7. Black J. Biological Performance of Materials, 3rd edn, Marcel Dekker: New York, NY, 1999.
8. Davis JR, ed. Handbook of Materials for Medical Devices, Chapter 4: Corrosion of Metallic Implants and Prosthetic Devices. ASM International: Materials Park, OH, 2003, pp. 51–74.
9. Fontana MG. Corrosion Engineering, 3rd edn, McGraw Hill: London, 1986.
10. Jones DA. Principles and Prevention of Corrosion, 2nd edn, Prentice-Hall: Upper Saddle River, 1996.
11. Scully JC. The Fundamentals of Corrosion, 3rd edn, Pergamon Press: Oxford, 1990.

Chapter 6
Wear

Chunming Jin and Wei Wei

6.1 Introduction

Wear is a critical issue for prostheses, implants, and other medical devices. Wear may lead to significant loss of material and/or failure of a medical device. For example, wear and wear-related damage commonly cause failure of hip, knee, and other orthopedic prostheses [1]. Even a relatively small amount of wear can lead to significant degradation of function for some medical devices. For example, wear debris generated from degradation of a joint prosthesis can result in a biological process known as osteolysis (bone resorption), which can cause loosening of the prosthesis [2, 3], Wear may also lead to failure of artificial heart valves and other medical devices that enable critical physiologic activities [4]. In this chapter, the wear mechanisms that are commonly encountered in biomedical materials and medical devices are discussed.

6.2 Friction, Lubrication, and Wear

Friction, lubrication, and wear are three concepts that are encountered when examining the relative motion of two surfaces that are placed in contact. The significance of friction, lubricant and wear in the function of medical devices has been investigated for several decades. The scientific examination of friction, lubrication, and wear emerged as an independent field in the 1960 s. For example, the term "tribology" originated in a 1966 British Department of Education and Science report to describe the study of friction, lubrication, and wear. Soon afterward, Dowson coined the phrase "biotribology" to describe the examination of wear, friction, and lubrication in biological systems [5].

The term friction refers to the resisting force that acts along the surfaces of two interacting bodies. Friction force occurs under the following conditions: (1) there is a force acting perpendicularly to the surfaces, (2) the surfaces have a predisposition towards relative movement or are actively moving relative to each other. Friction force F_f is related to the normal force $F_\perp$ through the following equation:

$$F_f = \mu F_\perp \tag{6.1}$$

in which μ is known as the coefficient of friction and is a dimensionless quantity. The friction force between surfaces that are moving relative to each other is referred to as kinetic friction. When there is no relative motion between the two surfaces, a friction force known as static friction is observed; the static friction coefficient is higher than kinetic friction coefficient. Movement of objects leads to energy loss since kinetic energy is converted to thermal energy. Materials known as lubricants may be used to separate the surfaces of the two interacting bodies in order to minimize friction and wear. Gases, liquids, solid films, or particles can serve as lubricants.

Friction is divided into dry friction and viscous friction. The term dry friction refers to friction behavior between two dry sliding solid bodies. Viscous friction refers to friction behavior of solid bodies that are separated by gases or fluids. In ideal situations, classical laws are valid for dry friction. For example, the frictional force is proportional to the normal load in ideal situations. In addition, the coefficient of friction is independent of surface area and sliding speed. Viscous or lubricated friction can be classified into different categories using the Stribeck curve [6, 7], which is shown in Fig. 6.1. The Stribeck curve is a plot of the coefficient of friction as a function of the parameter ZN/P, where Z is the viscosity of the lubricant, N is the velocity, and P is the load. When the value of ZN/P is low, a thin lubricant film coats the sliding surfaces; this situation is referred to as boundary lubrication. In boundary lubrication, the thickness of the lubricant film is similar to the height of the uneven features (asperities) on the surface of the material. In this case, the coefficient of friction is determined by the chemical properties of the lubricant film, the physical properties of the lubricant film, and the lubricant film thickness; this lubricant film (known as a boundary film) serves to limit wear. In general, boundary lubrication is characterized by relatively high coefficient of friction values and wear rates. A high ZN/P values corresponds to a situation in which high velocities, high lubricant viscosities, and low loads are observed. In this case, which is known as hydrodynamic lubrication, a thick, continuous lubricant film is observed
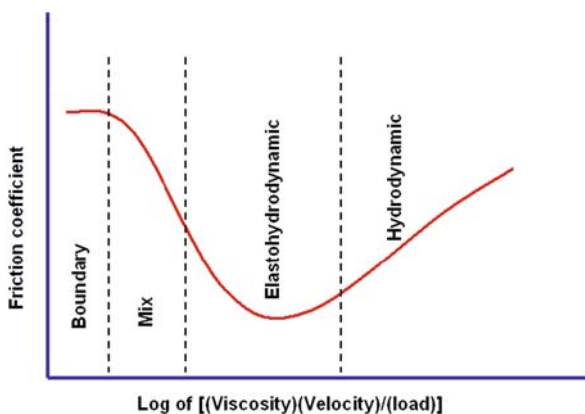

Fig. 6.1 Schematic illustration of Stribeck curve, which shows different lubrication regimes

between the two sliding surfaces. Friction in this regime is determined by the rheologic properties of the lubricant; wear of the sliding surfaces is limited since the two surfaces are not in direct contact. In elasto-hydrodynamic lubrication, the lubricant film is thinner than in hydrodynamic lubrication but it is sufficiently thick to prevent the two surfaces from directly contacting one another. Elastic deformation of the sliding surfaces may result from transfer of pressure through the separation film. In the mixed lubrication regime, the lubricant film is thin and discontinuous. Direct interaction between the two sliding surfaces occurs, resulting in high wear rates. In this regime, elastohydrodynamic and boundary mechanisms may be active at the same time. In large joints, the two sliding surfaces are bones that are covered at the ends with articular cartilage. In a healthy joint, viscous synovial fluid provides a lubricant that is released into the joint space under pressure in what is commonly referred to as "weeping" lubrication. The articulating motion for healthy human joint has very low coefficient of friction, which is in the range 0.005–0.025 [26]. In the human body, different lubrication modes may be observed at various stages of tissue motion. For example, boundary lubrication is the dominant mechanism when joint motion is initiated [8].

The term "wear" covers a wide range of phenomena related to surface damage; it is commonly defined as "damage to a solid surface, generally involving progressive loss of material, due to relative motion between that surface and a contacting substance or substances" [9–11]. Wear damage may also occur by means of plastic deformation near the surface of a material. In this case, no material removal is observed, but the shape of the material is changed due to the relative motion of the two surfaces. Wear damage to the surface of material may include any manner of surface degradation, including material removal, material displacement due to plastic deformation, topology changes (e.g., fracture), and surface chemical changes (e.g., oxidation) [10, 11].

6.3 Wear Classifications and Fundamental Wear Mechanisms

Several classification schemes have been developed to understand and compare wear processes. One or more fundamental wear mechanisms may play a role in any given real world wear process. Wear is classified based on several characteristics, including (1) the physical mechanism by which wear damage occurs, (2) the appearance of the wear damage, and (3) the condition of the wear process [11]. Bayer classified the physical mechanism of the wear damage into eight distinct categories, including adhesive wear, single-cycle deformation wear, repeated-cycle deformation wear, corrosive or oxidation wear, thermal wear, tribofilm wear, abrasive wear, and atomic wear [11]. In many biomedical devices (e.g., artificial joints), damage generally occurs by means of adhesive wear, abrasive wear, corrosive wear, and fatigue wear [1].
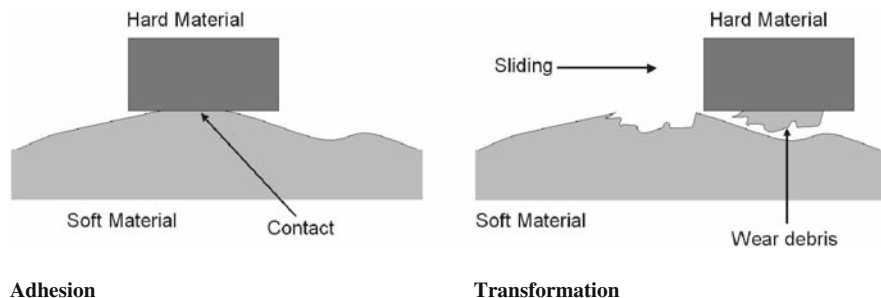
Adhesion   **Transformation**

Fig. 6.2 Schematic illustrations of adhesive wear behavior

6.3.1 Adhesive Wear

Adhesive wear originates from adhesion between two surfaces that are placed in contact. When two surfaces are brought into contact, asperities of the two surfaces make physical contact. This "true" contact area is significantly smaller than the apparent surface area of the two contact surfaces. The contact area between the two surfaces is localized to the small regions known as asperities; these asperity-asperity contact regions are referred to as junctions [11]. The size of a junction is usually in the range 1–100 μm; the typical size of a junction is 10 μm in diameter. The number of junctions is dependent on the surface roughness and the amount of load that is applied. Under load, bonding between asperities on the two contact surfaces may occur. The amount of deformation at these junctions is also dependent on the number of junctions and the size of the junctions. Under sliding motion, plastic deformation, cracking and fracture can occur in the "true" contact area. Adhesive wear is largely due to fracture of material and transfer of material at the asperity-asperity contact regions (Fig. 6.2). Prior to fracture, plastic deformation and crack formation may cause damage to the contact surfaces.

The formation of wear debris is dependent on the mechanical properties of the contact materials and the geometry of the asperities on the contact surfaces. For example, ductile materials generally produce a higher number of wear particles than brittle materials. Wear debris is usually generated from the material with lower hardness values. For example, in an orthopedic prosthesis that contains metallic and polymeric components, the adhesive wear debris is usually generated from the polymeric component [12]. Adhesive wear is associated with a high wear rate and a variable coefficient of friction [7, 11, 13].

The adhesion between two surfaces is also dependent on the chemical properties of the contact surfaces [13]. Materials with similar chemical properties generally exhibit higher adhesive forces because they are able to form chemical bonds more readily. Strong metal-to-metal adhesion may occur as a result of electron exchange and bond formation. For example, when adhesion between iron and other metals was examined under vacuum, the iron-iron contact had the highest adhesive force and the iron-aluminum contact had the second highest adhesive force [13]. In certain

cases, the thin oxide layers on many metallic materials may limit chemical bonding between metal surfaces. Strong adhesion between metallic and polymeric materials is also commonly observed. Adhesive wear is also a significant wear mechanism for joint prostheses that contain metallic and polymeric components. Chemical interaction and adhesive wear has been also observed between metal and ceramic contact surfaces [14].

An equation first proposed by Archard in 1953 has been developed to describe the volumetric wear rate for adhesive wear [11, 15]:

$$R_V = \frac{K P}{3 h} \tag{6.2}$$

In this equation, K is the probability that a given junction will produce adhesive wear, P is the normal force that holds the two contact surfaces together, and h is the penetration hardness of the softer material. This equation indicates that the adhesive wear rate is directly proportional to the load and inversely proportional to the hardness of the softer material. The equation also indicates that adhesive wear rate is proportional to the value of K. The value of K is determined by the adhesive behavior of the junction interface; weaker adhesion is correlated with smaller K values. For self-mated metals, typical values for K are in range of 2×10^{-4} to 0.2 for dry wear and 9×10^{-7} to 9×10^{-4} for lubricated wear [16]. This equation provides good agreement with experimental data in most cases.

6.3.2 Fatigue Wear

Fatigue wear is a wear mechanism that occurs when a material undergoes cyclic loading. Other wear mechanisms such as delamination wear and flow wear also fall in the general category of repeated-cycle wear processes [11]. Fatigue wear can be observed in sliding, rolling, or impact wear processes. In repeated rolling processes, wear damage may be referred to as surface fatigue or contact fatigue [13].

In repeated sliding situations, fatigue wear may occur even if the coefficient of friction is small and lubricant is present. In this mechanism, shear forces during sliding processes cause strain near the surface. As the plastic strain accumulates, the movement of dislocations leads to the formation of microcracks on the surface or below the surface (Fig. 6.3). Cracks are initially produced on the weak, imperfect, or otherwise damaged regions of the contact surface. These cracks grow and propagate through the material to form crack networks under repeated sliding motion. As the two surfaces continue to slide against one another, material on the contact surface is fractured and wear debris is formed. In rolling contacts, fatigue wear mechanisms depend mostly on the material properties and testing/operational conditions.

Fig. 6.3 Schematic
illustration of fatigue wear
behavior
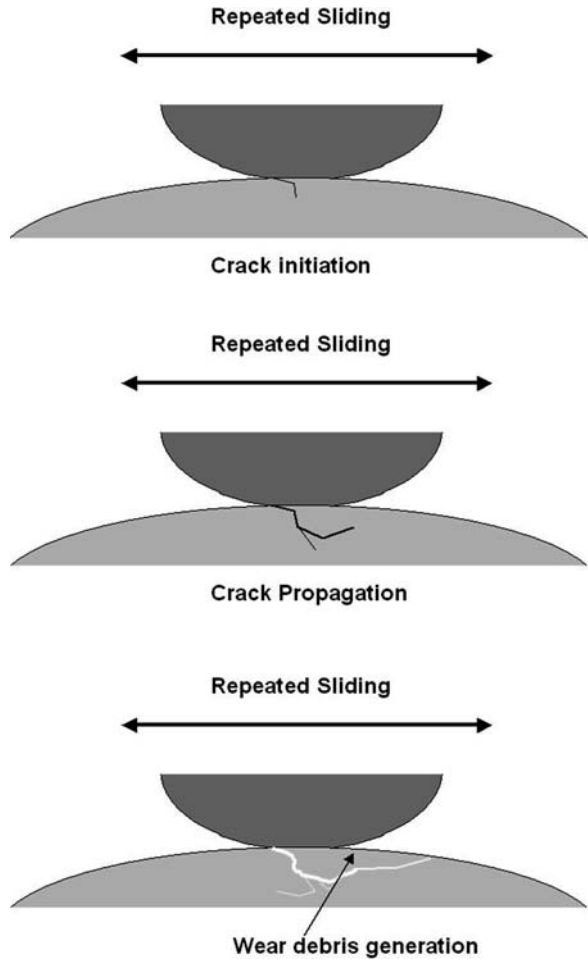
6.3.3 Abrasive Wear and Third-Body Wear

Abrasive wear is caused by plowing of particles and asperities [7, 11, 13]. This wear
process is commonly observed when the surface is placed in contact with either par-
ticles or surfaces that exhibit equal or greater hardness values. The term "abrasion"
is commonly used to describe wear processes involving two wearing surfaces and
the term "erosion" is commonly used to describe wear processes in which only
one surface is involved. Abrasion processes may be subdivided into two-body and
three-body abrasion processes [11]. Two-body abrasion involves damage caused by
the particles and asperities that are attached to another surface. On the other hand,
three-body abrasion wear process involves damage caused by hard particles that are
not fixed on a surface but instead move between the two contact surfaces.

Fig. 6.4 Schematic illustration of oxidative wear behavior
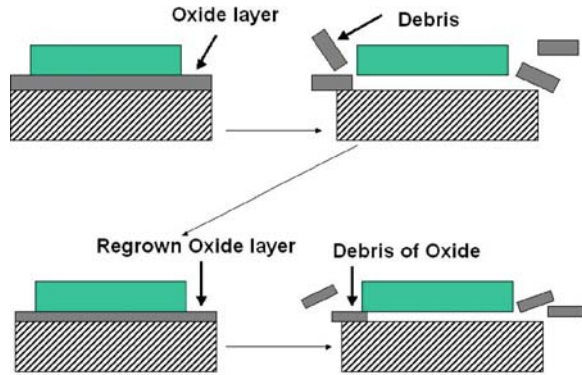
Abrasive wear can occur through four types of mechanisms: cutting, fracture, fatigue, and grain pull-out [13]. The cutting mechanism occurs when a hard sharp particle or surface asperity ploughs into the surface of the softer surface. Surface damage can also occur by fracture of brittle material, fatigue under repeated deformation, and grain detachment due to sliding of abrasive particles.

For ductile materials, abrasive wear volume is proportional to the normal load and the sliding distance and is inversely proportional to the hardness of the material [8]. For brittle materials, wear debris is produced primarily as a result of fracture. In this case, the abrasive wear rate is inversely proportional to both the hardness and the fracture toughness of the material.

6.3.4 Chemical (Corrosive) Wear

In chemical (corrosive) wear, chemical or electrochemical reactions accelerate the wear process [11, 17, 18]. Oxidative wear is a chemical wear process that is observed in many metals. Oxidative wear involves the continuous removal of oxide layers from the contact surface [11]. The removal of material commonly results from sliding contact. After the oxide layer is removed, the denuded metal surface is exposed to air or biological fluids and is quickly re-oxidized. As a result, a metal oxide layer re-forms on the surface of the material. These re-grown oxide layers are removed by sliding cycles and are subsequently regenerated by exposure to air or biological fluids. It is interesting to note that in many cases the metal oxide may exhibit lower wear rates than the corresponding metal [19]. In other cases, the oxide layer may delaminate due to sliding movement if adhesion between the oxide layer and the underlying metal is poor (Fig. 6.4) [17, 18]. Oxidative wear is commonly observed in dry wear processes; it may also be observed in lubricated wear processes. The oxidative wear rate is proportional to $\exp[-Q/(R_gT)]$, where T is temperature, Q is the activation energy, and R_g, is the gas constant. This relationship is valid for both metallic and ceramic materials [7].

6.4 Wear in Biomedical Devices and Biomaterials

6.4.1 Wear in Prostheses and Biomedical Devices

Over the past 40 years, biotribologists have made several notable efforts to increase the wear resistance of prostheses and medical devices. As mentioned earlier, wear may result in the failure of knee joint prostheses, hip joint prostheses, and other orthopedic prostheses [1–3]. Wear may also cause failure of cardiovascular devices, including artificial heart valves [4].

Many efforts have been made to improve the wear resistance of small and large joint prostheses, including shoulder, hip, knee, and finger joints. For example, hip prostheses are utilized for reduction of pain and re-establishment of function in patients who suffer from osteoarthritis, rheumatoid arthritis, osteonecrosis, post-traumatic arthritis, bone tumors, and hip fractures. A large number of hip prosthesis failures are due to osteolysis and aseptic loosening (loosening of the prosthesis-bone interface in the absence of infection). According to a recent National Institutes of Health report, the biological response due to wear debris, along with fixation of the acetabular component and problems associated with revision surgery, are critical issues that affect life span of hip prostheses [1]. In recent years, the materials and processes used for acetabular fixation have been significantly improved. Unfortunately, osteolysis and other reactions to wear debris remain critical issues that significantly limit the lifespan of hip prostheses.

A total hip joint prosthesis is composed of a femoral component and an acetabular component. Metals, ceramics, and polymers may be used to fabricate the acetabular component; metals and ceramics may be used to fabricate the femoral components. In a joint prosthesis, four different modes of wear behavior may be observed [12]. Mode 1 wear is associated with articulation between the anticipated bearing surfaces; for example, mode 1 wear in a total hip replacement prosthesis occurs between the femoral head and the acetabular cup. Mode 2 involves articulation between a primary bearing surface and an unintended surface. In a total hip replacement prosthesis, mode 2 wear may take place between the femoral head and the metal backing of the acetabular cup. Mode 3 is a three-body wear process that involves the intended bearing surfaces. Interaction between metal debris, the femoral head, and the acetabular cup is an example of three-body wear. Mode 4 is an articulation between two nonbearing secondary surfaces in a prosthesis.

In a joint prosthesis, the wear rate is dependent on various factors, including the type of motion that takes place when the prosthesis components interact and the number of cycles of motion that the prosthesis undergoes. In addition, wear rate in the artificial joint is also related to the clinical practices, design considerations, patient-specific factors, materials parameters, surface preparation, and tissue-material interaction [5].

The major wear mechanisms for joint prostheses include adhesive, abrasive, fatigue, and corrosive mechanisms [1]. In most situations, wear damage of an implanted device is caused by a combination of these mechanisms, with one or

more of these mechanisms being dominant. In metal-polyethylene joint implants, adhesive wear is a common primary wear mechanism.

In joint prostheses that contain metallic and polymeric components, most adhesive wear debris is made up of the softer polymeric material. In many cases, adhesive wear may be correlated with the types of asperities or regions of unevenness on the surface of metal component. Adhesion or cold welding may occur at the asperity-asperity junction regions due to large local stresses. As a result, small pieces of polyethylene are transferred from the polymer to the metal due to junction formation-destruction-reformation process. The wear debris may be either temporarily or permanently attached to the counterface material. The wear volume is proportional to both the sliding distance and the load acting on the device; the wear volume is also inversely proportional to the hardness of the material. The removal of polyethylene results in pitting and void formation on the polyethylene surface.

Asperities on the surface of the metal component can cause abrasive wear of polyethylene. In general, materials that exhibit higher hardness values generally demonstrate higher resistance to abrasive wear. In joint prostheses that contain metallic and polymeric components, abrasive wear damage mostly occurs on the polymer; some damage may also be observed on the metal surface. Two-body abrasive wear in metal-polymer joint prostheses is related to the surface roughness of the metallic component. Scratches on the metal surface may significantly increase the wear rate of the polyethylene surface. Hard particles (e.g., bone cement) may contribute to three-body wear of joint prostheses. These loose particles come from the prosthesis or from the immediate implant-tissue environment. Particles are either trapped between the contact surfaces (three-body wear) or attached to one of the contact surfaces (two-body wear). Metal, polymer, or bone particles embedded in a polyethylene-bearing surface may induce three-body wear in artificial joints. The rate of abrasive wear is determined by the surface roughness of the metal component and the presence of third-body particles.

Fatigue wear primarily takes place on the softer polyethylene surface. Subsurface delamination and cracking of the polyethylene component may result in the generation of polyethylene debris. Deep cracks in the polymeric biomaterial may result in the formation of particles. Corrosive wear may result from chemical or electrochemical interactions on the prosthesis surface; for example, metals may react with oxygen to form metal oxides on a metallic implant surface. This metal oxide surface may have lower shear strength than the underlying metal, and may exhibit a more rapid wear rate. The corrosive wear rate depends on the reactivity of the component materials and the biological environment.

6.4.2 Wear Resistance of Biomedical Materials

The materials currently used in total hip joint prostheses include metals, ceramics, and polymers. Two types of polymers are widely used for manufacturing joint devices [20]. For example, ultra high molecular weight polyethylene is commonly

used for fabricating the liner of the acetabular cup in total hip joint prostheses. Poly-methylmethacrylate is a bone cement that is used for attaching (fixing) the hip prosthesis to the surrounding bone. Degradation of both of these polymers may lead to failure of hip prostheses.

Polyethylene contains long chains of ethylene molecules (C_2H_4). Ultra-high molecular weight polyethylene exhibits a molecular weight of 2–6 million g/mole and a melting point of 140–145 °C. The density of ultra high molecular weight polyethylene is ∼0.930–0.945 g/cm^3, which is similar to the density of high density polyethylene (0.952–0.965 g/cm^3) [21, 22]. The mechanical properties of ultra high molecular weight polyethylene are dependent on its chemical structure, molecular weight, crystallinity, and thermal history [21].

Ultra high molecular weight polyethylene is a semi-crystalline polymer that exhibits crystallinity values in range of 45–75 % [21–23]. Ultra high molecular weight polyethylene consists of 10–50 nm thick, 10–50 μm long crystalline lamellae, which are surrounded by amorphous regions [21]. The amorphous regions consist of randomly oriented polymer chains that join the lamellae together and impart mechanical strength. The high crystallinity of ultra high molecular weight polyethylene provides good resistance to fatigue crack propagation [20]. The crystallinity of ultra high molecular weight polyethylene decreases when the material is heated above its melting point; this loss of crystallinity is irreversible [22].

The coefficient of friction for ultra high molecular weight polyethylene depends on the counterpart and contact conditions. For example, McKellop et al. reported friction coefficient values from 0.07 to 0.2 for ultra high molecular weight polyethylene lubricated with bovine serum [24]. Similar coefficient of friction results were reported by Klapperich et al.; they reported coefficient of friction values in the range 0.08–0.23 [23]. The in vivo performance of ultra high molecular weight polyethylene has also been investigated; average wear rates of ∼0.1–0.6 mm/year and wear volumes of 50–100 mm^3/year were reported by Chiesa et al. [25]. Wear rates for ultra high molecular weight polyethylene differ based on the surface properties of the counterpart material. For example, the wear volume of ultra high molecular weight polyethylene against cobalt-chromium alloy is ∼65 mm^3/million cycles. On the other hand, the wear rate of ultra high molecular weight polyethylene against alumina is 18 mm^3/million cycles [26, 27].

Every joint prosthesis and implantable medical device must be sterilized prior to implantation. Due to the low thermal tolerance of polyethylene, sterilization is usually performed by exposing the polyethylene component to gamma irradiation at doses between 25 and 40 kGy using a ^{60}Co source. Irradiation may cause two effects in ultra high molecular weight polyethylene, degradation and crosslinking [21, 28–34]. Degradation of ultra high molecular weight polyethylene may be caused by oxidation of the material either during irradiation or after device implantation [20, 28, 29]. Oxidation induces the chain scission of ultra high molecular weight polyethylene. As a result, the molecular weight of the polymer is reduced and the mechanical properties of the polymer are altered, resulting in reduced wear resistance. The interaction between high-energy radiation and ultra high molecular weight polyethylene results in the formation of free radical species through bond cleavage [35]. The

hydrogen radicals produced by C–H bond cleavage have high mobility; on the other hand, the mobility of radicals that result from cleavage of C–C bonds is low. If oxygen is present, free radicals can also form peroxyl radicals [28]. Free and peroxyl radicals can react with polyethylene chains, resulting in chain scission and additional free radical formation. Newly-generated free radical species maintain these reactions and contribute to the degradation of polyethylene. Chain scission reduces the length of polymer chain, and increases the crystallinity of the polymer. The reduction of chain length and molecular weight alters the physical properties of ultra high molecular weight polyethylene. The density and elastic modulus values of ultra high molecular weight polyethylene increase due to the chain scission; as a result, the material becomes brittle. The fatigue, fracture and wear resistance values for oxidized ultra high molecular weight polyethylene are lower than those for as-prepared material. Free radical species demonstrate very long lifetimes. Even if radiation sterilization of ultra high molecular weight polyethylene is performed in an oxygen-free environment, free radical species may be formed that cause embrittlement of the material. Oxidative degradation may continue after the prostheses are implanted in the body; this degradation process may result in poor long-term wear resistance.

Free radicals created during high-energy irradiation may be minimized using thermal processes, including annealing or melting processes [34]. Unfortunately, these treatments are not practical for total joint prostheses and other medical devices. Several attempts have been made to increase the wear resistance of ultra high molecular weight polyethylene [21]. One possibility involves the use of carbon-fiber reinforced ultra high molecular weight polyethylene (Poly II^{TM}). This composite material contains short carbon fibers in an ultra high molecular weight polyethylene matrix. Although laboratory studies provided promising results for this composite, in vivo implantation studies demonstrated osteolysis and mechanical failure [21, 36–38]. It was found that the carbon-fiber reinforced composite exhibited very low crack resistance compared with ultra high molecular weight polyethylene, since the carbon fiber did not bond to the polymer matrix [21].

Crosslinking is an effective method for improving the abrasive wear resistance of polyethylene [21, 39–41]. Crosslinking of polyethylene chains may be performed using three different methods: irradiation, peroxide chemistry, and silane chemistry [21]. In the irradiation method, gamma ray or other radiation is used to cleave C–H and C–C bonds in polyethylene in order to produce free radicals. The molecular weight of the polymer is reduced when carbon-carbon bonds are cleaved (chain scission). Crosslinking is achieved when inter-chain covalent bonds are formed by the reaction of free radicals from different chains. If cross-linking involves radicals from C–H bonds, it is referred to as H-type cross-linking. Y-type cross-linking results from reactions between the free radicals generated by the cleavage of C–C bonds. Crosslinking is most significant in amorphous regions of ultra-high molecular weight polyethylene. Improvements in the physical properties of ultra high molecular weight polyethylene are dependent on several crosslinking parameters, including technique, dose, radiation source, process temperature, and process time. Crosslinking may significantly improve the wear resistance of ultra high molecular

weight polyethylene; however, the fracture toughness of crosslinked ultra high molecular weight polyethylene may be reduced [34].

Metal-on-metal joint prostheses have been developed in an attempt to reduce the number of total hip joint prosthesis failures associated with wear of ultra high molecular weight polyethylene. Early metal-on-metal hip prosthesis designs were fabricated using stainless steel; however, most current metal-on-metal hip prostheses contain two components that are fabricated out of cobalt-chromium alloys [42, 43]. First-generation metal-on-metal prostheses were replaced with ultra-high molecular weight polyethylene/cobalt-chromium alloy prostheses in the 1970 s for several reasons, including seizure of the cast metal surfaces [42, 43]. Second-generation cobalt-chromium alloy/cobalt-chromium alloy prostheses were developed in the 1980 s. During in vivo studies, these implants demonstrated a one year "running-in" period, in which the wear rate of 25 μm/year was observed. After the "running-in" period, a more impressive in vivo wear rate of 5 μm/year was noted. Abrasive wear is the dominant wear mechanism for these devices; however, adhesive wear and fatigue wear have also been observed.

Ceramic materials have been used in joint prostheses as an alternative to the polyethylene-metal design since the 1970 s [44]. The two most common ceramic materials used in prosthesis bearing surfaces are alumina and zirconia. Alumina exhibits very high hardness (Vickers hardness 400–450 GPa) and elastic modulus (380 GPa) values [45]. Prostheses containing alumina components generally produce a low number of wear particles. A wear rate of $0.1 \, mm^3$/million cycles has been reported for alumina-on-alumina devices; this value is significantly lower than the wear rate that has been reported for cobalt-chromium alloy–polyethylene devices ($65 \, mm^3$/million cycles) [26, 27]. Several factors may contribute to the enhanced wear resistance of alumina. In addition to the high hardness, alumina components can be manufactured with very low surface roughness ($<0.005 \, \mu m$); as a result, abrasive wear is minimized. The hydrophilic nature of alumina may also contribute to good wear resistance properties, since it enables full film lubrication [46]. One major drawback of alumina is low fracture toughness values (brittleness), which can cause failure of the prosthesis or medical device [44].

Zirconia (zirconium oxide, ZrO_2) exhibits higher fracture toughness values than alumina. Pure zirconia has three different crystalline phases: cubic, tetragonal and monoclinic. At room temperature, pure zirconia is monoclinic. The transformation from the monoclinic phase to the tetragonal phase occurs at 1000–1100 °C. At 2000 °C, transformation from the tetragonal phase to the cubic phase takes place. Shape change and volume expansion is associated with each phase transformation. The stresses created by these phase transformations result in the formation of cracks. For this reason, zirconia is stabilized with yttrium oxide (Y_2O_3) or magnesium oxide (MgO); these doped structures exhibit a stable tetragonal phase at room temperature. Doped zirconia materials include partially stabilized zirconia (PSZ), tetragonal zirconia polycrystals (TZP), and zirconia toughened ceramics. For example, yttrium stabilized tetragonal zirconia polycrystal material is a fine-grained material that exhibits high fracture toughness values (6–12 MPa $m^{1/2}$); these values are approximately twice as high as that of alumina (4–5 MPa $m^{1/2}$) [44]. However,

yttrium stabilized tetragonal zirconia polycrystals undergo a slow transformation from tetragonal to monoclinic phase. Sato et al. have shown that water molecules in the environment promote the tetragonal-monoclinic transformation on the surface of zirconia prostheses [47]. This volume change leads to surface microcrack formation, which can result in an increase in surface roughness.

Surface modification has been considered for optimizing the wear resistance and biocompatibility of the total hip joint prostheses and other medical devices [48]. For example, superhard biocompatible coatings, including titanium nitride, silicon carbide, tungsten carbide, and diamond-like carbon, have been utilized for increasing the wear and corrosion resistance of metallic biomaterials [49]. Diamond-like carbon is a metastable amorphous material that contains both sp^3-hybridized carbon atoms and sp^2-hybridized carbon atoms; hydrogenated and nonhydrogenated forms of this material may be prepared. Diamond-like carbon is a superhard material. The elastic modulus and hardness value of diamond-like carbon depends on the fraction of sp^3-hybridized atoms in the films [50–53]. An sp^3 fraction of 10% corresponds to a hardness value of 2,000–3,000 Hv. As the fraction of sp^3-hybridized carbon atoms increases to 50%, a hardness value of 7,000–8,000 Hv can be achieved. Diamond-like carbon films with 100% sp^3 hybridized carbon atoms have demonstrated hardness values of 10,000 Hv [54]. The elastic modulus values for diamond-like carbon films with 0% to 90% sp^3-hybridized carbon atoms can vary between 300 GPa and 800 GPa [55].

The coefficients of friction values for diamond-like carbon coatings depend on the amount of hydrogen incorporated in the film, ambient humidity, topology, and sliding partner [56]. For example, humidity plays a significant role in determining the coefficient of friction for hydrogenated diamond-like carbon coatings. The coefficient for hydrogenated diamond-like carbon coatings is in the range of 0.01–0.3 in vacuum conditions, but it greatly increases under humid conditions. On the other hand, hydrogen-free diamond-like carbon coatings demonstrate low coefficient of friction values (<0.1) under both low and high humidity conditions [57].

Diamond-like carbon coatings can significantly improve wear resistance of metallic biomaterials. For example, Affatato et al. performed an in vitro investigation of femoral heads coated with diamond-like carbon [48]. In their study, diamond-like carbon was coated on a titanium alloy (Ti6Al4V) head using chemical vapor deposition. They found that wear of the polyethylene-diamond-like carbon coated Ti6Al4V couple was comparable to that of the polyethylene-alumina couple. Similar results have been reported by other groups for diamond-like carbon coated cobalt-chromium-molybdenum alloy-polyethylene and diamond-like carbon-coated stainless steel-polyethylene devices [58, 59]. However, there is some controversy in the reported results. For example, Sheeja et al. indicated there was only a slight difference in wear rates between cobalt-chromium-molybdenum alloy-ultra high molecular weight polyethylene and multilayered diamond-like carbon-coated cobalt-chromium-molybdenum alloy-ultra high molecular weight polyethylene wear couples [60].

The major problem associated with diamond-like carbon coatings on metal surfaces is poor adhesion, which is related to compressive stress in the film and poor

chemical bonding between the film and the substrate [61, 62]. The internal compressive stress within diamond-like carbon coatings can be as high as 10 GPa, which limits the coating thickness in range 100–200 μm. To reduce film stress and increase film thickness, diamond-like carbon-metal composite coatings have been utilized. Diamond-like carbon-metal composite coatings, including diamond-like carbon-silver composite coatings and diamond-like carbon-titanium composite coatings, maintain hardness and wear properties similar to those of unalloyed diamond-like carbon films. In addition, they exhibit excellent adhesion to medical alloy substrates [63, 64].

6.5 Summary

The wear properties of biomedical materials play a major role in determining the overall success of medical prostheses and other implantable medical devices. Adhesive wear, abrasive wear, fatigue wear and corrosive wear play a key role in degradation of medical devices and prostheses. Several factors determine the relationship between the in vitro properties of the component materials and the in vivo wear performance of the medical device or prosthesis. Significant improvements in the wear properties of biomedical materials may be achieved through intensive examination of material-, device-, surgical-, and patient-specific parameters that determine in vivo wear behavior. Minimization of wear in biomedical materials may only be achieved through effective interaction among clinicians, biotribologists, and biologists.

References

 1. Wright TM and Goodman SB, eds. Implant Wear in Total Joint Replacement: Clinical and Biologic Issues, Material and Design Considerations, American Academy of Orthopaedic Surgeons: Rosemont, IL, 2001.
 2. Zhu YH, Chiu KY and Tang WM. Polyethylene wear and osteolysis in total hip arthroplasty, J Orthop Surg, 2001, 9: 91–99.
 3. Teoh SH. Fatigue of biomaterials: a review, Int J Fatigue, 2000, 22: 825–837.
 4. Kelpetko V, Moritz A, Schurawitzki H, Domanig E and Wolner E. Leaflet fracture in Edwards–Duromedics bileaflet valves. J Thoracic Cardiovascular Surg, 1989, 97, 90–94.
 5. Hutchings IM, ed. Biotribology – A Personal View, Friction, Lubrication and Wear of Artificial Joints, Professional Engineering Publishing Ltd.: Bury St. Edmunds, UK, 2003.
 6. Buckley DH, Jones WR Jr, Sliney HE, Zaretsky EV, Townsend DP, and Loewenthal SH. Tribology: The Story of Lubrication and Wear, NASA Technical Memorandum 101430, 1985.
 7. Bhushan B, ed. Modern Tribology Handbook, CRC Press: Boca Raton, FL, 2001.
 8. Black J. Biological performance of materials: Fundamentals of biocompatibility, Marcel Dekker: New York, NY, 1992.
 9. Standard terminology relating to wear and erosion, standard G-40-01, American Society for Testing and Materials, 2001.
10. Hutchings IM, The Challenge of Wear, in Stachowiak GW, ed. Wear-Materials, Mechanisms and Practice: Chichester, England, John Wiley & Sons Ltd.: Hoboken, NJ, UK, 2005, Chapter 1, pp. 1–7.

11. Bayer RG, Mechanical wear: fundamentals and testing, Marcel Dekker Inc.: New York, NY, 2004.
12. McKellop HA. The lexicon of polyethylene wear in artificial joints. Biomaterials, 2007, 28: 5049–5057.
13. Stachowiak GW and Batchelor AW. Engineering Tribology, Elsevier Butterworth-Heinemann: Amsterdam, 2005.
14. Buckley DH and Miyoshi K. Friction and wear of ceramics. Wear, 1984, 100: 333–353.
15. Archard JF. Contact and Rubbing of Flat Surfaces. J Appl Phys, 1953, 24: 981–988.
16. Burwell JT Jr. Survey of possible wear mechanisms. Wear, 1957, 1: 119–141.
17. Burwell JT and Strang CD. On the Empirical Law of Adhesive Wear, J Appl Phys, 1952, 23, 18–28.
18. Rabinowicz E. Adhesive wear. Friction and Wear of Materials. John Wiley and Sons: New York, NY, 1965.
19. Bhushan B and Gupta B. Handbook of Tribology. Section 3.3. McGraw-Hill: New York, NY, 1991.
20. Santavirta S, Konttinen YT, Lappalainen R, Anttila A, Goodman SB, Lind M, Smith L, Takagi M, Gdmez-Barrena E, Nordsletten L, and Xu J-W. Materials in total joint replacement, Current Orthopaedics, 1998, 12, 51–57.
21. Kurtz SM, Muratoglu OK, Evans M, and Edidin AA. Advances in the processing, sterilization, and crosslinking of ultra-high molecular weight polyethylene for total joint arthroplasty. Biomaterials, 1999, 20: 1659–1688.
22. Li S and Burstein AH. Ultra high molecular weight polyethylene. The material and its use in total hip joint implants. J Bone Joint Surg Am, 1994, 76: 1080–1090.
23. Klapperich C, Komvopoulos K, and Pruitt L. Tribological properties and microstructure evolution of ultra-high molecular weight polyethylene. Trans ASME, 1999, 121: 394–402.
24. McKellop H, Clarke IC, Markolf KL, and Amstutz HC. Wear characteristics of UHMW polyethylene: a method for accurately measuring extremely low wear rates, J Biomed Mater Res, 1978, 12: 895–927.
25. Chiesa R, Tanzi MC, Alfonsi S, Paracchini L, Moscatelli M, and Cigada A. Enhanced wear performance of highly crosslinked UHMWPE for artificial joints. J Biomed Mater Res A, 2000, 50: 381–387.
26. Buford A and Goswami T. Review of wear mechanisms in hip implants: Paper I – General. Mater Design, 2004, 25: 385–393.
27. Buford A and Goswami T. Review of wear mechanisms in hip implants: Paper II – ceramics IG004712. Mater Design, 2004, 25: 385–393.
28. Daly BM and Yin J. Subsurface oxidation of polyethylene. J Biomed Mater Res, 1998, 42: 523–529.
29. Goldman M, Lee M, Gronsky R, and Pruitt L. Oxidation of ultrahigh molecular weight polyethylene characterized by Fourier Transform Infrared Spectrometry. J Biomed Mater Res, 1997, 37: 43–50.
30. Lee CS, Yoo SH, and Jho JY, Mechanical properties of ultra-high molecular weight polyethylene irradiated with gamma rays. Macromol Res, 2004, 12: 112–118.
31. Rose RM, Goldfarb EV, Ellis E, and Crugnola AN. Radiation sterilization and the wear rate of polyethylene. J Orthop Res, 1984, 2: 393–400.
32. Goldman M and Pruitt L. Comparison of the effects of gamma radiation and low temperature hydrogen peroxide gas plasma sterilization on the molecular structure, fatigue resistance, and wear behavior of UHMWPE. J Biomed Mater Res, 1998, 40: 378–384.
33. McKellop H, Shen F-W, Lu B, Campbell P, and Salovey R. Effect of sterilization method and other modifications on the wear resistance of acetabular cups made of ultra-high molecular weight polyethylene. J Bone Joint Surg, 2000, 82: 1708–1725.
34. Maher SA, Furman BD, Babalola OM, Cottrell JM, and Wright TM. Effect of crosslinking, remelting, and aging on UHMWPE damage in a linear experimental wear model. J Orthop Res, 2007, 25: 849–857.

35. Bracco P, Brunella V, Luda MP, Zanetti M, and Costa L, Radiation-induced crosslinking of UHMWPE in the presence of co-agents: chemical and mechanical characterisation, Polymer, 2005, 46: 10648–10657.

36. Wright TM and Bartel DL. The problem of surface damage in polyethylene total knee components. Clin Orthop, 1986, 205: 67–74.

37. Wright TM, Astion DJ, Bansal M, Rimnac CM, Green T, Insall JN, and Robinson RP. Failure of carbon fiber-reinforced polyethylene total knee-replacement components. A report of two cases. J Bone Joint Surg A, 1988, 70: 926–932.

38. Wright TM, Rimnac CM, Faris PM, and Bansal M. Analysis of surface damage in retrieved carbon fiber-reinforced and plain polyethylene tibial components from posterior stabilized total knee replacements. J Bone Joint Surg A, 1988, 70: 1312–1319.

39. Heisel C, Silva M, Dela Rosa MA, and Schmalzried TP. Short-term in vivo wear of cross-linked polyethylene. J Bone Joint Surg Am, 2004, 86: 748–751.

40. Dorr LD, Wan Z, Shahrdar C, Sirianni L, Boutary M, and Yun A. Clinical performance of a durasul highly cross-linked polyethylene acetabular liner for total hip arthroplasty at five years. J Bone Joint Surg Am, 2005, 87: 1816–1821.

41. Martell JM, Verner JJ, and Incavo SJ. Clinical performance of a highly cross-linked polyethylene at two years in total hip arthroplasty: a randomized prospective trial. J Arthroplasty, 2003, 18: 55–59.

42. Steinberg DR and Steinberg ME. The early history of arthroplasty in the United States. Clin. Orthop Relat Res, 2000, 374: 55–89.

43. Rieker CB, Köttig P, Schön R, Windler M, and Wyss UP. Clinical wear performance of metal-on-metal hip arthroplasties, in Jacobs JJ and Craig TL, ed. Alternative Bearing Surfaces in Total Joint Replacement, ASTM International: West Conshohocken, PA, 1998.

44. Rahaman MN, Yao A, Bal BS, Garino JP, and Ries MD. Ceramics for prosthetic hip and knee joint replacement. J Am Ceram Soc, 2007, 90: 1965–1988.

45. Boutin P, Christel P, Dorlot JM, Meunier A, de Roquancourt A, Blanquaert D, Herman S, Sedel L, and Witvoet J. The use of dense alumina-alumina ceramic combination in total hip replacement, J Biomed Mater Res, 1988, 22: 1203–1232.

46. Hutchings IM, ed. Friction, Lubrication and Wear of Artificial Joints, Professional Engineering Publishing Ltd.: Bury St. Edmunds, UK, 2003.

47. Sato T, Ohtaki S, Endo T, and Shimada M. Science and technology of zirconia. Advances in Ceramics. American Ceramic Society: Westerville, OH, 1988.

48. Affatato S, Frigo M, and Toni A. An in vitro investigation of diamond-like carbon as a femoral head coating. J Biomed Mater Res (Appl Biomater), 2000, 53: 221–226.

49. Dearnaley PA. A review of metallic, ceramic and surface treated metals used for bearing surfaces in human joint replacements. Proc Inst Mech Eng, 1999, 213-H: 107–135.

50. Enke K, Dimigen H, and Hubsch H. Frictional properties of diamond-like carbon layers, Appl Phys Lett, 1980, 36: 291–292.

51. Pharr GM, Callahan DL, McAdams SD, Tsui TY, Anders S, Anders A, Ager JW, Brown IG, Bhatia CS, Silva SR P, and Robertson J. Hardness, elastic modulus, and structure of very hard carbon films produced by cathodic-arc deposition with substrate pulse-biasing. Appl Phys Lett, 1996, 68: 779–781.

52. Ronkainen H, Varjus S, Koskinen J, and Holmberg K. Differentiating the tribological performance of hydrogenated and hydrogen-free DLC coatings. Wear, 2001, 249: 260–266.

53. Holmberg K and Mathews A. Coatings tribology: A concept, critical aspects, and future directions, Thin Solid Films, 1994, 253: 173–178.

54. Collins CB, Davanloo F, Lee TJ, Park H, and You JH. Noncrystalline films with the chemistry, bonding, and properties of diamond. J Vac Sci Technol B, 1993, 11: 1936–1941.

55. Schneider D, Schwarz T, Scheibe HJ, and Panzner M. Non-destructive evaluation of diamond and diamond-like carbon films by laser induced surface acoustic waves. Thin Solid Films, 1997, 295: 107–116.

56. Erdemir A, Bindal C, Pagan J, and Wilbur P. Characterization of transfer layers on surfaces sliding against diamond-like hydrocarbon films in dry nitrogen. Surf Coatings Technol, 1995, 76–77, 559–563.
57. Liu Y, Erdemir A, and Meletis EI. An investigation of the relationship between graphitization and frictional behavior of DLC coatings, Surf Coatings Technol, 1996, 86: 564–568.
58. Lappalainen R, Heinonen H, Anttila A, and Santavirta S. Some relevant issues related to the use of amorphous diamond coatings for medical applications. Diamond Rel Mater, 1998, 7: 482–485.
59. Dearnaley G and Mccabe A. Bioapplications of diamond-like carbon coatings. 4th World Biomater Cong, Berlin, 1992.
60. Sheeja D, Tay BK, Lau SP, and Nung LN. Tribological characterization of diamond-like carbon coatings on Co-Cr-Mo alloy for orthopaedic applications. Sur Coatings Technol, 2001, 146: 410–416.
61. Morshed MM, McNamara BP, Cameron DC, and Hashmi MSJ, Effect of surface treatment on the adhesion of DLC film on 316L stainless steel. Surf Coatings Technol, 2003, 163: 541–545.
62. Schwan J, Ulrich S, Theel T, Roth H, Ehrhardt H, Becker P, and Silva SRP. Stress-induced formation of high-density amorphous carbon thin films. J Appl Phys, 1997, 82: 6024–6030.
63. Morrison ML, Buchanan RA, Liaw PK, Berry CJ, Brigmon R, Riester L, Jin C, and Narayan RJ. Electrochemical and antimicrobial properties of diamond-like carbon-metal composite films. Diamond Related Mater, 2006, 15: 138–146.
64. Bell BF, Scholvin D, Jin C, and Narayan RJ. Pulsed laser deposition of hydroxyapatite-diamond-like carbon multilayer films and their adhesion aspects. J Adhesion Sci Technol, 2006, 18: 221–232.

Chapter 7
Inflammation, Carcinogenicity and Hypersensitivity

Patrick Doherty

7.1 Introduction

Biomaterials, be they metal or polymer, ceramic or glass or some combination of all of these, when placed inside the human body are not normally rejected. The term rejection in this context generally refers to an adverse immunological response analogous to that seen when patients reject an organ transplant. For the most part, non-biological materials when placed in contact with human tissue, either upon or beneath the skin are well tolerated and may become "accepted" by the body. In these circumstances, if the implant is then able to effectively carry out its function for the required period of time, it may be described as expressing biocompatibility.

However, every implant material, no matter how "biocompatible," stimulates a response by the tissues of the body. This response may be such that the recipient (the patient) is unaware of any problem and the implant effectively carries out its function. Or the response may be so severe that the patient's life is compromised. When placed subcutaneously within the body a biomaterial will induce an inflammatory response. This is the opening gambit in the body's defense mechanism and the commencement of the wound healing response.

How our body responds to a wound, and the surgical implantation of a medical device such as vascular prosthesis or a hip joint replacement clearly produces a wound, is well understood. How the presence of a biomaterial modifies that response is somewhat less well defined.

7.2 Granulation Tissue

Granulation tissue defines inflammation during wound healing. Once healing has initiated following injury, granulation tissue will be initiated at the site within a few days. It is characterized by specific cell and tissue activities which can be highlighted using microscopic histological evaluation. Histologists and pathologists have over many years developed staining techniques which can be applied to very thin sections of tissue and permit an accurate examination of the tissue. Four significant components have been identified. (1) Active macrophage cells: these cells are employed in phagocytosis. Phagocytic cells act as mini-vacuum cleaners and

clear up tissue debris. (2) Neutrophil activity: these cells are also employed in a phagocytic role, specifically in the removal of bacterial and other foreign proteins. (3) Angiogenesis: this describes the formation of new blood capillaries infiltrating the site of injury. A new blood supply will provide and effective access route for the wound healing cells and nutrition for the regenerating tissue. (4) Fibroblast cell activity: these cells are employed in the production and deposition of new connective tissue [1, 2].

Wound healing in the absence of an implant, following a clean surgical incision and effective suturing where the wound edges are brought into apposition, will progress with minimal loss of tissue and minimal scar tissue formation. Wound healing in the presence of an implanted biomaterial will lead to a modification of this progression and the extent of this modification will be determined by factors related to the implant and the site of implantation. These factors will include the size and shape of the biomaterial device, the chemical nature of the material and the physical form of the implant. In particular the nature of the biomaterial surface will influence the inflammatory response and the new tissue development. Factors such as porosity and surface topography have been demonstrated to influence significantly this response. A material may be selected for implantation because of its stable, non-reactive nature; nevertheless its presence will fundamentally disrupt the formation of connective tissue, alter the network of new blood capillaries (angiogenesis) and may stimulate an enhanced phagocytic cell activity leading to continued inflammation.

7.3 Foreign Body Response

"*The foreign body reaction is neither a single event nor a simple process, but a broad concept and multi-factorial phenomenon*" [3]. This definition, true of almost every event associated with the inflammatory response, identifies the foreign body response as a part of the body's initial defense mechanism against any invading agent from bacteria to biomaterial. This may be a relatively short term response or part of a long-lasting or chronic reaction. It may stimulate a specific (antibody related) or a non-specific immune response. In simple terms the foreign body response may be considered as that part of the inflammatory process which is activated by the presence of a foreign agent, e.g., bacterial, biomaterial, and will continue to be active until the foreign agent is removed. A foreign body reaction will involve the components of granulation tissue, phagocytic cells such as neutrophils and macrophages and will further see the introduction of another cell type. That is the formation of giant cells or multi-nucleated cells. The foreign body giant cell is by definition a large cells, up to $100\,\mu m$ in diameter (Fig. 7.1). It is formed when a foreign agent stimulates a phagocytic activity which cannot be effectively completed by the cell. This may be due to the large size of the agent (a very realistic factor when considering phagocytic cell activity in contact with biomaterials) or to some chemical effect of the foreign agent which inhibits the phagocytic process.
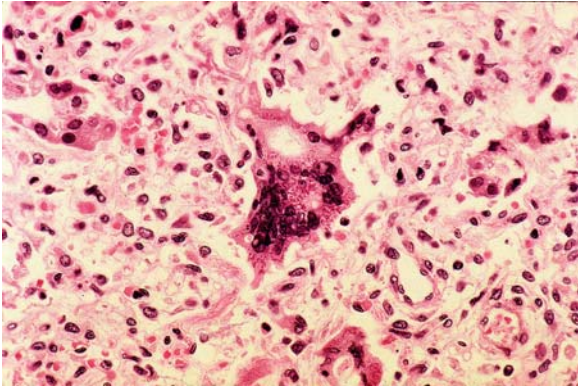
Fig. 7.1 Foreign body giant cell. An implanted material will always stimulate a foreign body response. A sustained macrophage activity in the presence of a degrading or corroding implant material may result in the formation of foreign body giant or multinucleated giant cells. The image above details a giant cell formed in response to a bacterial toxin. (Image from Public Health Image Library)

The outcome is the formation of large, multi-nucleated cells. The nature of the agent which stimulates the response will effectively determine the nature of the response, the activity of the cells at the site and for what length of time the response will remain active. For example, a simple wound with no infection or implant present may heal with very little evidence of a foreign body response. The attendance of invading bacteria and therefore foreign protein, will stimulate an increase in granulation and phagocytic activity. The presence of an implanted biomaterial will further modify this response. This may lead to a process termed "frustrated phagocytosis" where the phagocytic cell is unable to encapsulate the stimulating material. It has been postulated that this is the trigger for the development of foreign body giant cells in contact with implanted biomaterials [4].

7.4 Repair

Repair and remodeling of new tissue is the final stage in the wound healing response and, if the injury is slight, will lead to the regeneration of normal tissue and a return to the "status quo." However if the injury is more severe, and this is the more common scenario, the process will lead to the generation of normal tissue and also scar tissue. Scar tissue can be defined as tissue that is the result of wound healing. It is composed of fibers of collagen and will often lead to a restricted elasticity of the tissue concerned. Most importantly, scar tissue generally is less able to carry out the function of the tissue it is replacing. The extent of scar tissue formation is highly dependent on the severity of the initial injury. In the presence of an implanted biomaterial it may take the form of fibrosis or fibrous encapsulation where the active fibroblast cells lay down protein (collagen) which is physically wrapped around the

implant device, thus sealing it from the rest of the tissue. This encapsulating tissue will be primarily composed of fibroblast cells and collagen. In an ideal scenario the implanted material will then be ignored by the body, subsequent wound healing will continue uninterrupted and the implanted device will function effectively. However, local and systemic factors may be involved in the progression of the tissue response in these circumstances [2]. Local factors will include the physical and chemical nature of the implant material, the nature of the local blood supply and the effects of any invading micro-organisms. Systemic factors will include the health of the patient and the effect of the release of any leachable or degrading or corroding products from the implant itself.

7.5 Acute and Chronic Inflammation

An acute inflammation is usually a short, highly active period which persists for a limited time and is resolved within a few days. It is characterized by the action of the neutrophils, the first cells to be recruited to the damaged site, and other white blood cells. The rapid recruitment of neutrophils to the site of injury will result in effective phagocytosis of micro-organisms and any other foreign materials. The acute inflammation is further characterized by the accumulation of plasma proteins at the injured site, leading to swelling (edema). These plasma proteins are in fluid phase and this leads to a build up and retention of fluid at the site. The subsequent swelling can result in perceived pain for the patient. Recruitment of neutrophils is triggered by chemical signals (chemotactic factors) which stimulate only those cells which have the means to recognize these factors to move towards the site. This triggering process relies on the action of cell surface proteins or receptors. Interactions between chemical signals (generally protein) and cell surface receptors (protein) are highly specific and a cell will only be stimulated or activated if it carries the receptor to a specific signal. In this way neutrophils are activated to pass through the blood vessel wall following appropriate stimulation and migrate towards the source of the chemotactic agent. There are a very large number of proteins which have been identified as having chemotactic properties at inflammatory sites and we are only beginning to understand the mode of action and relevance of a small proportion of these.

    The phagocytic potential of the recruited neutrophil is then enhanced by a process know as opsonization. This is a process by which the body identifies the materials/agents to be phagocytozed by coating or labeling them with specific proteins (opsonins). Opsonins occur naturally in blood plasma and have receptors which are recognized by the neutrophils. The most significant opsonins (i.e., those factors which have been most researched and are best understood) are immunoglobulin G (IgG) and a complement related protein fragment termed activated C3 (C3b) [5]. These proteins have been shown to attach to foreign material at the site of injury (usually they attach to bacterial proteins but they have been shown to attach to implanted materials) and enhance the subsequent attachment and activation of

neutrophils and macrophages leading to phagocytosis. In normal circumstances this would lead to effective removal of the foreign agent from the site. However, when an implant is present, it is highly likely that the phagocytic cell will be unable to engulf and remove the phagocytic stimulus. As stated previously, this may lead to "frustrated phagocytosis" during which the enzyme contents of the phagocytic cell are released, further phagocytic cell recruitment is stimulated and foreign body giant cells may arise.

A biomaterial which is effectively sealed within a fibrous capsule may permit the acute inflammatory phase to progress to completion as would be expected in a wound healing environment. However a biomaterial which is degrading or corroding or in some other way disturbing the surrounding tissue may stimulate a chronic inflammatory response. The presence of the macrophage is probably the most significant factor in any chronic inflammatory response associated with implanted biomaterials. The macrophage is a large phagocytic white blood cell, derived from the monocyte. The monocyte migrates from the blood to the site of stimulus (injury) where it is changed or transformed. The macrophage is present in most tissues and has a scavaging role in the body. When activated at the site of an implanted biomaterial it plays a highly significant role in the inflammatory response (Fig. 7.2). The chronic response is characterized by the presence of large numbers of different cells types at the site of injury or implantation. These cells can include all or any of the following; monocytes, macrophages, lymphocytes, fibroblasts, endothelial cells, foreign body giant cells. There is also likely to be a highly vascularized connective tissue mass. A chronic inflammatory response to implanted materials is generally highly localized around the tissue-material interface and is usually of fairly short duration (2–3 weeks). However, persistent chronic inflammation has been demonstrated with a number of implant materials and devices. It is likely that an inflam-
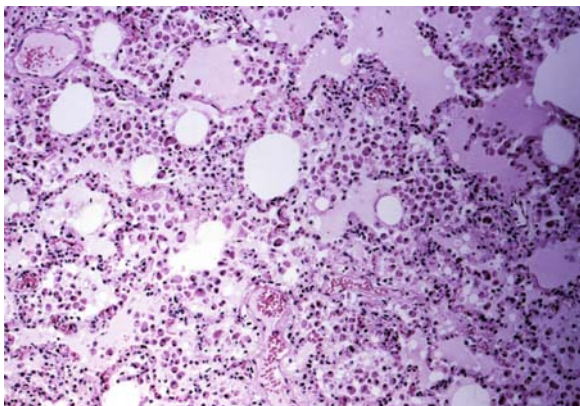
Fig. 7.2 Macrophage activity. This response is indicative of an aggressive tissue reaction to an adverse stimulus. There are numerous macrophages present. The image details the histopathologic changes detected in a lung biopsy tissue. Necrotic changes are indicated by the breakdown of the alveolar walls resulting in pulmonary edema. (Image from Public Health Image Library)
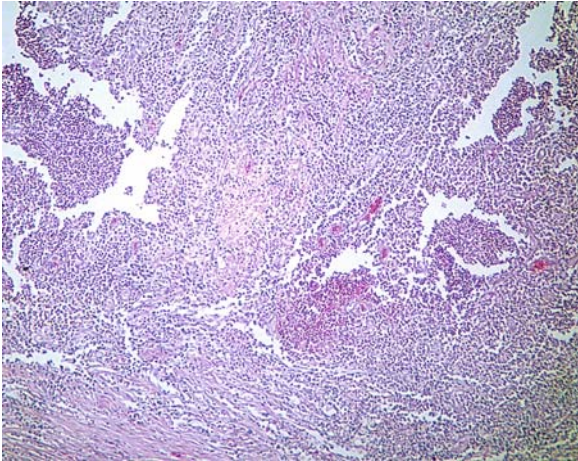
Fig. 7.3 Chronic inflammation. The image details a fibrotic granulation tissue which is indicative of a chronic inflammatory response. The area details a mass of inflammatory cells stimulated as a result of a pulmonary infection. (Image from Public Health Image Library)

matory response which persists for long time periods (months to years) may also be associated with an infection which is implant related [2].

Characterizing a chronic inflammation to an implanted material is notoriously difficult. There will be a mixed cell population and possibly evidence of cell necrosis (cell death). Necrosis may be stimulated by a over-activated inflammation or may be due to the presence of a cytotoxic biomaterial (Fig. 7.3).

7.6 Infection

Infection occurs in up to 10% of patients who have received an implanted medical device [6]. Nosocomial infection rates, i.e., hospital acquired infection, can be very much higher and pose a serious risk to patients who have received implants. Such infections are extremely resistant to antibiotic therapies. Infections which become problematic shortly after implant surgery are often acquired during the surgical procedure or within the hospital environment. These infections will massively exacerbate the inflammatory process and may contribute to early failure and removal of the medical device. This is not only disadvantageous for the patient but very costly for the health care provider and a significant contributing factor to the huge problems associated with nosocomial infection. However, it is also true that the very presence of the implant may be a contributing factor to the virulence of the micro-organism, i.e., the ability of the organism to exert a toxic effect on the body. The implant may provide a convenient means of transport for the invading organism, carrying it to tissues normally inaccessible, where it is more likely to become pathogenic. The presence of the device may inhibit the natural defense mechanisms of the body and

compromise the functioning of the immune system. Increasingly there is concern that the physical presence of the implant may provide a substrate upon which a resistant, pathogenic biofilm may flourish.

7.7 Local and Systemic Responses

It is not uncommon to read in the medical device or materials literature of references to "inert" biomaterials. These references are generally alluding to the older generation of implant materials, i.e., stainless steel, polyethylene, which were selected because it was considered that they would have no interaction with or effect upon the tissues with which they came into contact. Notwithstanding the success these and other biomaterials have had as surgical implants, we are now aware that these materials are not inert. Indeed they not only stimulate a response that may well have adverse consequences for the interfacing tissues, but with time the biological environment will have a deleterious effect on the materials themselves. It is now accepted that no material placed within the body is inert; indeed such an effect would be contrary to the purpose of the majority of implanted medical devices. Most materials used in implant surgery today are designed to have an active role when in contact with tissue and will in due time stimulate a tissue response. In most cases this response will be a simple modification of the normal wound healing process the body will undergo following any traumatic injury. As indicated previously, the nature of the modification will be determined by factors which are related not only to the chemical nature of the implant (e.g., metal, polymer, ceramic), but also to the size and shape of the implant, the surface topography, the porosity, the wettability of the material surface and the propensity for the material to degrade or corrode in the biological environment.

   The nature of any systemic response associated with an implant is more difficult to determine than the local response. In most cases the local response is immediate and by definition located at the site of the implantation. A systemic response may arise months or years following implantation and may be manifest at a site distant from the original implant. Many of the concerns surrounding systemic toxicity are linked with the potential for carcinogenic or mutagenic changes or the ability for the implant material to stimulate a hypersensitive response.

7.8 Soft and Hard Tissue Responses

The presence of an implant may provide an on-going stimulus to inflammation. This may be as a result of interaction of the implant with the soft tissues or through the production of leachable or degradation products from the implant. We may assume that an implant material which is stable and does not produce leachable components will not have a significant effect on the wound healing process other than to provide a physical barrier and a substrate for the deposition of protein and the development of a collagenous capsule (Fig. 7.4). In this situation the implant material is
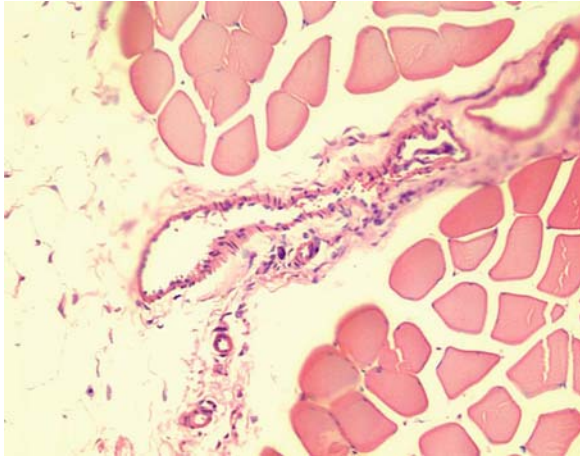
Fig. 7.4 Fibrous capsule. The image details a fibrous tissue which has formed in response to an implanted biomaterial. The biomaterial, in particulate form, was placed within a muscle pocket in an experimental animal (rat). The tissue section shown was prepared after an implantation period of 4 weeks. In many studies using such implantation techniques, the thickness of the fibrous capsule has been used as a "measure" of the biocompatibility of the materials

effectively sealed off from the body and plays no other active role. If the implant can maintain its function then it may remain successfully within the body indefinitely. If the implant material continues to stimulate a response, i.e., by the production of leachable components or corrosion/degradation products, then the fibrous encapsulation may be enhanced. A greater level of material interaction will stimulate a greater cellular activity and consequently a progressively thicker fibrous capsule. In early animal studies the thickness of this capsule was often equated to the "biocompatibility" of the implant material. The thicker the capsule the more irritating and therefore less "biocompatible" was the biomaterial considered. However, an aggressive implant material may never allow the tissue response to stabilize. Continued stimulus from the implant will lead to a persistent and chronic inflammation. This inflammatory environment will mean a more active and aggressive cellular response thus further stimulating greater and greater degradation.

Implantation of a biomaterial into bone tissue will stimulate a wound healing response analogous to that encountered following fracture or other hard tissue injury. This follows a well recognized developmental pattern, involving an initial wound healing inflammation, bone remodeling and bone maturation. The success of a biomaterial implanted into bone may be determined by its ability to permit this process to occur and itself to become integrated into the healing process, leading to a stable fixation of the implant at the hard tissue site. The search for new and better endosseous implants focuses primarily on the development of materials which demonstrate osteogenesis, i.e., materials which actively participate in and/or stimulate the production of new bone tissue during the healing phase and are actively incorporated within this new tissue. There is a huge number of factors

which may influence this process. However, in simplistic terms, biomaterials have been attributed with the properties of contact osteogenesis (bone formation occurs directly in contact with the implant surface and "grows" out towards the surrounding bone) and distance osteogenesis (the surrounding bone is the focus of osteogenesis which may be stimulated by the presence of the implant). However contact and distance osteogenesis are generally identified by microscopic "snapshot" evaluation of recovered tissue/material interfaces, and have become convenient descriptive terms and may not be accurate descriptions of what is occurring during a complex healing process. Roberts [7] has postulated a hard tissue healing timetable based on a rat dental implant study. In agreement with most studies on hard tissue wound healing, the process is categorized into four stages. It is worth noting that these are not distinctly different processes and the wound healing process in both soft and hard tissue healing is a fluid, dynamic process.

The four stages are (1) development of a woven callous, by 6 weeks, (2) lamellar compaction by 18 weeks, (3) interface remodeling by 18 weeks and (4) development of mature compact bone by 54 weeks. The initial inflammatory phase may be considered completed by the 6-week stage and it is during this period that the presence of an implanted material will have the most significant effect on the long-term fate of the bone-biomaterial interface.

7.9 Blood–Material Interactions

When blood contacts a surface that is not the endothelial lining of blood vessels it undergoes a complex series of interdependent reactions which may lead to platelet activation and adhesion, coagulation and the formation of thrombosis. These mechanisms have evolved as part of a sophisticated defense process designed to limit blood loss following injury and facilitate the wound healing process. Unfortunately the introduction of artificial materials into the body via surgical intervention leads to the contact of blood with the biomaterial substrate, with inevitable adverse consequences. These interactions primarily involve the biomaterial substrate, platelets, clotting proteins and the eventual breakdown of the resultant clot or thrombosis by the fibrinolytic process [8]. The body has evolved a number of mechanisms for dealing with thrombus formation once it has been initiated. Therefore, although all biomaterials in contact with blood will stimulate an adverse response, a number of materials can be tolerated and will function effectively. These mechanisms include, the physical/mechanical action of flowing blood to prevent the build up of thrombosis and naturally occurring inhibitors of coagulation to inactivate thrombin. Platelets, coagulation and blood vessel cell lining (endothelial) systems interact to promote local hemostasis, but prevent generalized blood clotting and thrombosis [8]. Determination of the hemocompatibility of biomaterials is complex because these factors are not effective during in vitro evaluation of biomaterials. Yet it is by using in vitro testing methods that researchers and manufacturers garner most of the biological information. And these are the methods used to pass

judgment on the potential for an implant material to function effectively within a blood contacting environment or on its likelihood to lead to adverse or possible fatal biological consequences. The relative merits of cell culture (in vitro) and animal testing (in vivo) is an ongoing argument in biomaterial evaluation. Many researchers will argue that the test environment encountered within the in vitro test is so far removed from the clinical experience as to be meaningless and produce misleading data. Notwithstanding the moral/ethical arguments associated with the use of animals in biomaterial testing, many others will argue that in vivo or animal studies can only be interpreted effectively if we have already garnered the fundamentals of the cell-material interactions using cell culture. The argument is particularly focused in blood-biomaterial interactions where blood cells and blood components introduced to the in vitro environment have already been exposed to adverse conditions before ever meeting the material or device under consideration. But it remains an argument that rages without effective resolution in all areas of the biological evaluation of implant materials and the determination of what is termed biocompatibility.

7.10 Biocompatibility

The term biocompatibility is often associated with cell culture testing of biomaterials. In many cases simple in vitro test methods are employed with a potential biomaterial as the test sample. If a positive result is obtained, i.e., the cells remain viable, then the material is labeled "biocompatible." If the cells die or are significantly adversely affected, the material is labeled "non-biocompatible." This is a serious misuse of the term biocompatibility and a misuse of the in vitro test. Biocompatibility is not the result of a single event or phenomenon. It refers to a collection of processes involving different but independent mechanisms of reaction between a material and its host tissue. Biocompatibility refers to the ability of a material to perform a function. The ability to perform and to continue to perform this function depends not only on its interaction with tissue but also on the intrinsic mechanical and physical properties of the material. The definition of biocompatibility refers to appropriate host response [9]:

> The ability of a material to perform with appropriate host response in a specific application.

It does not stipulate that there should be no response. This may be a minimal response or indeed a more aggressive response. The definition refers to specific application. Biocompatibility should always be described with reference to the situation in which a material or device will be used. No material should be described as "biocompatible" without further qualification. Biocompatibility is therefore not an intrinsic property of any material [10]. The majority of in vitro tests employed in biomaterial evaluation measure cytotoxicity. However lack of toxicity cannot be equated with biocompatibility. Cell culture assays are not biocompatibility assays. The safety evaluation of medical devices is generally based on in vivo testing which evaluates acute and chronic toxicity and inflammation, irritation and sensitivity,

carcinogenicity and mutagenicity, allergic responses, hemocompatibility and systemic toxicity. It is not possible to assess all of these factors using in vitro assays; however, test models have now become established which permit the in vitro evaluation of acute toxicity, genotoxicity [11], hemocompatibility [12], necrotic cell death and apoptosis [13].

The introduction and development of in vitro testing has raised many concerns with respect to reproducibility and proper regulatory control. The establishment of new in vitro assays requires three steps, development, validation and acceptance. Acceptance of in vitro assays involves satisfying the regulatory bodies, the manufacturers and the users. At present there are few, if any, properly validated in vitro tests [14].

International standard ISO 10993 part 5 [15] was developed as the result of a harmonization of various existing standard procedures and the outcome reports of working groups. The standard acknowledges the widespread use of in vitro cytotoxicity testing and defines a scheme for testing which will allow the user to identify the most appropriate test procedures. These tests specify the incubation of cultures of mammalian cells with the medical device or component part of the device (i.e., the direct test method), or with an extract derived from the device (i.e., the indirect test method). Information is provided on appropriate test sample preparation for the direct testing and in some detail the most appropriate means of producing extracts for indirect testing is presented. The standard provides details on a number of recommended established cell lines, although these are not exclusive. The standard further provides details on how cytotoxicity may be most appropriately determined. This evaluation is categorized as follows:

1. Assessment of cell damage by morphological means
2. Measurements of cell damage
3. Measurements of cell growth
4. Measurement of specific aspects of cell metabolism

It is noted that evaluation may be by qualitative or quantitative means. It should also be noted that the standard does not present a single in vitro method for cytotoxicity evaluation, but allows the user to select an appropriate range of tests. Furthermore, at no point does the standard use the term biocompatibility in connection with this type of test. Another important aspect of the standard is the reference made to negative and positive control materials that should be employed in all test systems.

Replacement of animal testing is not the objective of in vitro assays and an appropriate use of both in vitro and in vivo is the most sensible approach to biomaterial evaluation. Simple in vitro assays can provide screening information and play a significant role in the early phase of investigation. Subsequent studies should employ more specific in vitro models alongside animal studies. This level of testing should include the use of human cell lines.

The test procedures identified in ISO 10993 part 5 are designed to evaluate cytotoxicity. The primary goal of such tests is to determine safety. As such these can

be considered screening tests for use in the development and research of novel materials and for the quality control of medical devices (both component parts and final products). Even at this simple level of testing relevant information can be garnered with respect to the final use of the test material. However the user must be careful to appreciate the simplicity of such testing. For example, it is prudent that any cytotoxicity evaluation should employ more than one cell line and where appropriate and possible should make use of human primary cell lines. Even simple screening tests can benefit from some level of quantification where this can be sensibly applied. Although ISO 10993 does make reference to the use of subjective, semi-quantitative grading systems, these will rarely produce any meaningful data and do not easily allow for the comparison of data between different laboratories or indeed different users. Qualitative assessment of cell morphological changes by an appropriately trained individual remains a more meaningful technique. However quantification using measurements of cell metabolic functions is readily employed and can provide accurate and reproducible data easily transferred between researchers. These includes cell-counting procedures, both manual and automatic [16], assessment of DNA levels [17] or the determination of cell activity by MTT assay [18].

7.11 Carcinogenicity

The potential for biomaterials or their degradation products to induce the growth of malignant cells is an area of increasing concern as we strive to develop and manufacture materials which will spend longer and longer times in contact with the tissues of the human body. Sarcomas associated with implant materials have been cited in the literature following procedures with a number of animal models (almost always rats and mice). However, biomaterial induced sarcomas are rare in humans. In experimental animals there has been evidence produced of the carcinogenicity of a wide range of implanted materials (polymers and metals) generally when presented with smooth surfaces. Some metallic powders have also been implicated when placed in contact with connective tissues. In contrast there is little evidence of carcinogenicity with polymeric powders or with the metals primarily used in orthopedic surgery. ISO 10993 part 3 [19] describes carcinogenicity testing as the means to determine the tumorigenicity potential of devices, materials and/or extracts to either a single or multiple exposures over a period of the total lifespan of the test animal. As the animal is generally a rodent and the maximum experimental time is around 12 months, there remains considerable concern as to how relevant such a procedure may be in identifying potential problems with materials which may be implanted within human tissues for periods in excess of 30 or 40 years. The standard further states that although any implant remaining within the body in excess of 30 days should be considered to have long-term or permanent contact, it is only necessary to conduct carcinogenicity testing if there is evidence from other sources which suggest a

problem. Consequently, few implant materials have been subjected to animal testing which is specifically designed to identify carcinogenic potential.

Most carcinogenicity concerns are associated with the physical presence of the biomaterial in direct contact with the tissue. There are a number of factors which may influence the potential for such an implant to trigger an adverse response. Most significant of these appears to be the size, shape and surface finish of the implant material. A highly polished, smooth surface has been most significantly associated with tumors, particularly in the animal model. Indeed materials lose their carcinogenic potential if the smooth nature of the substrate is disrupted. This is often referred to as the "Oppenheimer effect," a reference to the research group who identified this phenomenon in a large animal study. However, there are also concerns that chemicals which leach from implanted materials, either by design or otherwise, may stimulate a toxicity that has an adverse effect on the genetic material of the host cells. We can perhaps take some comfort in the fact that there is now a large population of patients who have been recipients to long term implantations (joint replacements, breast prostheses, etc.) with little evidence that carcinogenicity is a problem. Nevertheless, carcinogenicity concerns are real and it is incumbent upon the biomaterials community to research and develop test methods which will effectively highlight potentially carcinogenic materials.

7.12 Hypersensitivity

Sensitization or hypersensitivity reactions to biomaterials are usually associated with a prolonged or repeated contact with a chemical substance, normally a substance which leaches from the biomaterial and stimulates the body's immune system. The majority of such biomaterial induced reactions have been associated with skin contacting devices; however, increasingly concerns are being expressed over the potential for implanted materials to stimulate both the cell mediated and antibody mediated immune responses. Degradation products from biomaterials which act as antigens within the body are known as allergens and may stimulate an exaggerated immune response which will lead to damaging effects on the host. Metal hypersensitivity is well established in areas outside implantation; however corrosion of metal implants and the formation of metal ion-protein complexes may mediate metal hypersensitivity following implantation. The most significant chemical allergens associated with biomaterials and the degradation products of biomaterials are low molecular weight entities, i.e., metal ions. They are referred to as haptens and only become allergens when complexed with another biological molecule. It is important therefore that appropriate means of assessing the potential allergenicity of an implant material are employed before materials can be considered for long-term implantation. At present there is no effective means of predicting the potential for an implanted material or leachable chemical to generate an allergic effect.

References

1. Pawlowski KJ. Host reactions, in Bowlin GL, ed. Encyclopedia of Biomaterials and Biomedical Engineering, Marcel Dekker: New York, 2004, pp. 770–778.
2. Anderson JM. Inflammation, wound healing and the foreign body response, in Ratner BD, Hoffman AS, Schoen FJ, and Lemons JE, eds. Biomaterials Science: An Introduction to Materials in Medicine, 2nd edn, Elsevier: Amsterdam, 2004, pp. 296–303.
3. Williams DF. The Williams Dictionary of Biomaterials, University of Liverpool Press: Liverpool, 1999.
4. Hunt JA. Foreign body response, in Bowlin GL, ed. Encyclopedia of Biomaterials and Biomedical Engineering, Marcel Dekker: New York, 2004, pp. 641–648.
5. Kopp R, Mottaghy K, and Kirschfink M. Mechanism of complement activation during extracorporeal blood-biomaterial interaction: effects of heparin coated and uncoated surfaces. ASAIO, 2002, 48: 598–605.
6. Costerton B, Cook G, Shirtliff M, Stoodley P, and Pasmore M. Biofilms, biomaterials and device-related infections, in Ratner BD, Hoffman AS, Schoen FJ, and Lemons JE, eds. Biomaterials Science: An Introduction to Materials in Medicine, 2nd edn, Elsevier: Amsterdam, 2004, pp. 345–354.
7. Roberts WE. Bone tissue interface. J Dent Educ, 1988, 52: 804–809.
8. Hanson SR. Blood-material interactions, in Bowlin GL, ed. Encyclopedia of Biomaterials and Biomedical Engineering, Marcel Dekker: New York, 2004, pp. 144–154.
9. Williams DF, ed. Definitions in biomaterials, in Progress in Biomedical Engineering, vol. 4, Elsevier: Amsterdam, 1987, p. 54.
10. Williams DF. Biocompatibility, an overview, in Williams DF, ed. Concise Encyclopedia of Medical and Dental Materials, Pergammon Press: New York, 1990, pp. 51–59.
11. Chauval-Lebret DJ, Auroy P, Tricot-Doleux S, and Bonnaure-Mallet M. Evaluation of the SCGE assay to assess the genotoxicity of biomaterials. Biomaterials, 2001, 22: 1795–1807.
12. ISO 10993. The Biological Evaluation of Medical Devices, Part 4: Selection of Tests for Interactions with Blood, International Standard Organization 2002, pp. 10993–10994.
13. Renvoize C, Biola A, Pallardy M, and Breard J. Apoptosis; identification of dying cells. Cell Biol Toxicol, 1998, 14: 111–120.
14. Jones B and Stacey G. Safety considerations for in vitro toxicology testing. Cell Biol Toxicol, 2001, 17: 247–270.
15. ISO 10993. The Biological Evaluation of Medical Devices, Part 5: Tests for Toxicity; In Vitro Methods, International Standard Organization 2002, pp. 10993–10995.
16. Pizzoferatto A, Ciapetti G, Stea S, Cenni E, Arciola CR, Grachi D, and Savarino L. Cell culture methods for testing biocompatibility. Clin Mater, 1994, 15: 173–190.
17. Kirkpatrick CJ, Bittinger F, Wagner M, Kohler H, van Kooten TG, Klein CL, and Otto M. Current trends in biocompatibility testing. Proc Inst Mech Eng, 1998, 212: 75–84.
18. Doherty PJ. Cell culture assays, in Bowlin GL, ed. Encyclopedia of Biomaterials and Biomedical Engineering, Marcel Dekker: New York, 2004, pp. 292–298.
19. BS EN ISO 10993-3. Biological evaluation of medical devices, Part 3: Tests for genotoxicity, carcinogenicity and reproductive toxicity, 2003.

Chapter 8
Protein Interactions at Material Surfaces

Janice L. McKenzie and Thomas J. Webster

8.1 Introduction

It is important to note that while this chapter will emphasize protein interactions essential for mediating cell functions on implants, the same fundamental principles apply for any application in which proteins interact with surfaces. Applications in which a firm understanding of protein–surface interactions are needed include biomaterials, tissue engineering materials, implants, bioseparations, biosensors, biochips, food science, and aquatics (to only name a few).

Specifically for biomaterials, the first event that occurs when an implant is placed in the body is that water molecules adhere to the surface (in different ways for different surfaces) followed by the formation of a layer of proteins from bodily fluids. Next, cells of the body explore the adsorbed proteins and may subsequently attach to specific binding sites exposed in these proteins (Fig. 8.1). The types of cells that recognize the adsorbed proteins and attach to the implant surface both initially and chronically determine the fate of the implant. Will the implant successfully integrate into the juxtaposed tissues or will the body wall it off to exclude it from interacting as much as possible?

Properties of the implant surface control these initial water–surface interactions which subsequently regulate which proteins will adsorb, how many proteins will adsorb, and what their conformations will be once they adsorb. For these reasons, it is important to design biomaterials with surfaces that optimize select protein adsorption for controlling desirable cell responses pertinent to specific implant functionality.

8.2 Protein Properties

Proteins are an amazing component of biological systems. Though not "alive" like cells, they are macromolecules that perform many actions essential to life such as catalysis and the controlled transport of charged or neutral macromolecules across
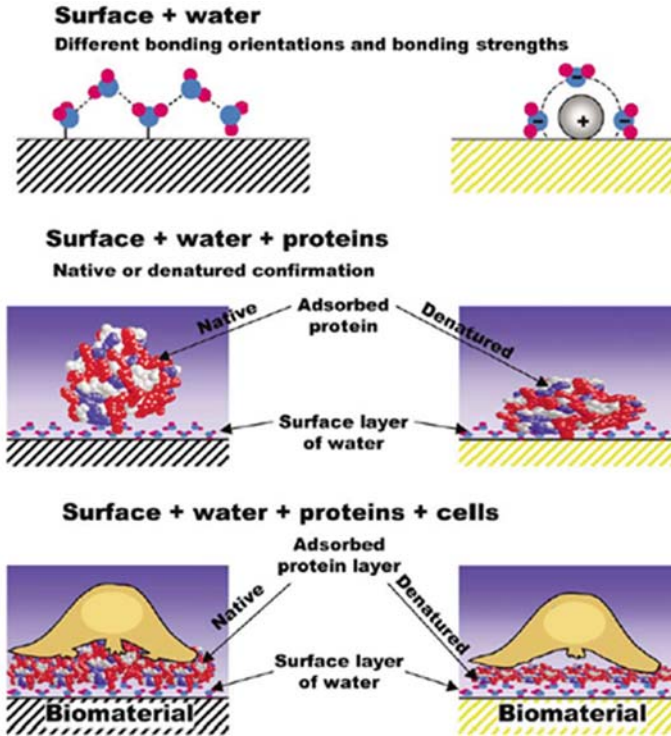
Fig. 8.1 Schematic of the events following implantation of a biomaterial. First, water molecules adhere to the surface followed by protein interactions which influence final cell adhesion [1]

cell membranes. In addition to enzymatic activity and cell membrane transport, proteins such as hemoglobin can transport iron through the blood stream and regulate metabolic activity. Actin and myosin are examples of proteins that are contractile and function in muscular movement. Proteins such as collagen provide structure to our organs and casein functions to store proteins for later use. In addition to controlling activities directly associated with cell and organism homeostasis (or regulations of health), proteins are the integral components that determine the fate of biomaterials deigned for diverse applications such as drug delivery, neural stimulation, bone ingrowth, etc.

8.2.1 Structure

When considering protein interactions with surfaces, of first importance are their constituent amino acids (or their primary structure). Amino acids are the building blocks of proteins. The sequence of these amino acids and the subsequent folding of the chain of amino acids dictate the nature of the final three-dimensional structure that proteins will assume. The chemical nature of the amino acids that compose

a protein and their associated quantities contribute to the ultimate hydrophobicity or hydrophilicity of the protein. (Hydrophilicity can be thought of as water-loving and hydrophobicity as water-hating.) Proteins have both hydrophobicity and hydrophilicity sections but the percentage of each varies from protein to protein. The following sections elaborate on protein structure due to its importance in mediating cell responses on implants.

8.2.1.1 Primary Structure

Proteins can be thought of as polymers, only most polymers are composed of very few monomers and proteins can be composed of 20 or more monomers. The nature of the protein molecule is determined ultimately by the order of the amino acids along the protein backbone once they have been condensed into a polypeptide. The structure of the amino acids can be viewed in Fig. 8.2. The primary sequence of amino acids will ultimately dictate hydrogen bonding and the final folding structure of the protein.

Each amino acid is composed of a carboxyl group (—COO⁻) attached to a hydrogenated carbon atom (CH). This carbon atom is also bonded to a side group (—R) and an amino group (—NH_3^+). The R group can be nonpolar (hydrophobic) or polar (uncharged, acidic, or basic). R groups that only contain carbon and hydrogen atoms can be referred to as aliphatic unless they have a special ring structure which then classifies them as aromatic. The amino and carboxyl groups of two amino acids form a covalent peptide bond which links the amino acids in the primary sequence of the protein together.

DNA sequencing, Edman degradation and mass spectroscopy (MS) techniques allow for the decoding of the primary sequence of a protein. All of these techniques involve cleaving proteins and comparing the fragments, often in iterations to determine the order of the amino acids. Now computer databases have been created that have polypeptide sequences already defined which makes amino acid sequencing easier when amino acids are unsequenced. Table 8.1 displays the frequency at which each amino acid is present in different proteins which ultimately influences their molecular weights. The diversity in amino acid properties and the primary structure of proteins plays a large part in their interactions to surfaces.

8.2.1.2 Secondary Structure

The basic amino acid arrangement in proteins determines the secondary structure of the protein through electrostatic hydrogen bonding to other sections of the protein. The carbonyl and amide groups of each amino acid form hydrogen bonds with water molecules in the solution or with amide or carbonyl groups of other amino acids in the protein backbone. This hydrogen bonding usually leads to the formation of characteristic α-helix or β-sheet structures. The α-helix contains 3.6 amino acids per turn and, at 0.15 nm per residue, is 0.54 nm per turn. The hydrogen bonds in the α-helical are oriented parallel to the helical axis. Bond properties such as length establish the amount, rotation, and flexibility of the amino acids to form secondary structures.

The β-sheet structure of proteins was first theorized in the early 1950s by Linus Pauling who won a Nobel Prize for the discovery of the α-helix structure. In the β-sheet form of the secondary protein structure, each backbone zigzags up and down like a pleated sheet of paper and forms hydrogen bonds with an adjacent string of amino acids that is in a matching zigzag or pleated form. Each amino acid is in one plane, so it takes two amino acids to form the up and down "V-shaped" structure of each pleat that then repeats to make the β-sheet pleats. The β-sheets
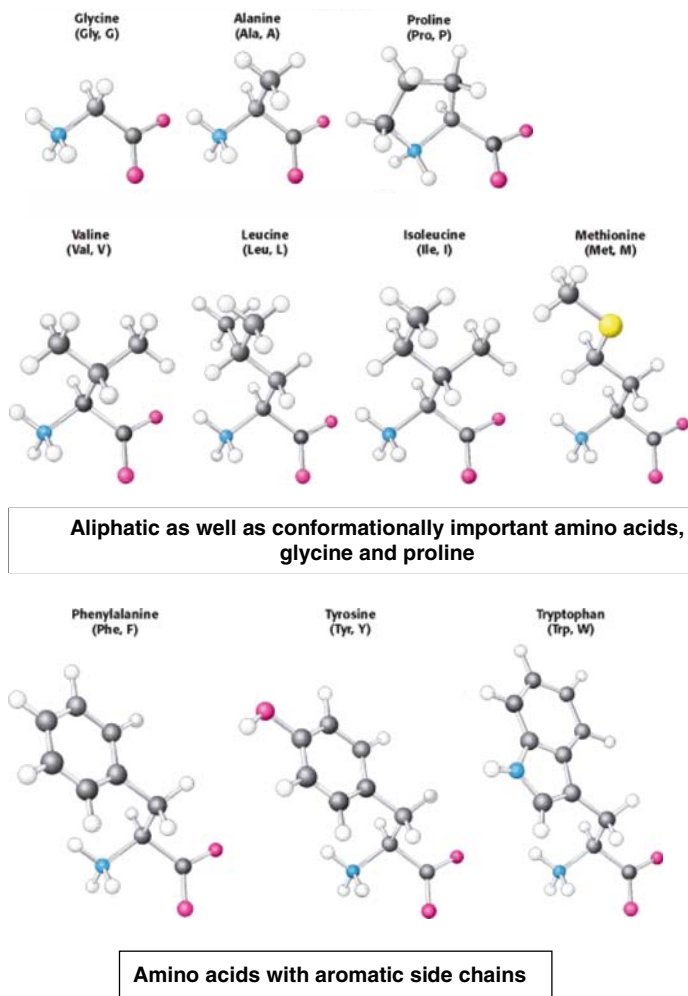
Fig. 8.2 Ball and stick representations of atoms and bonds that show the 20 amino acid structures. *White balls* represent hydrogen; *gray*, carbon; *lighter gray*, oxygen; *light, light gray*, nitrogen; and *lightest gray*, sulfur atoms [2]
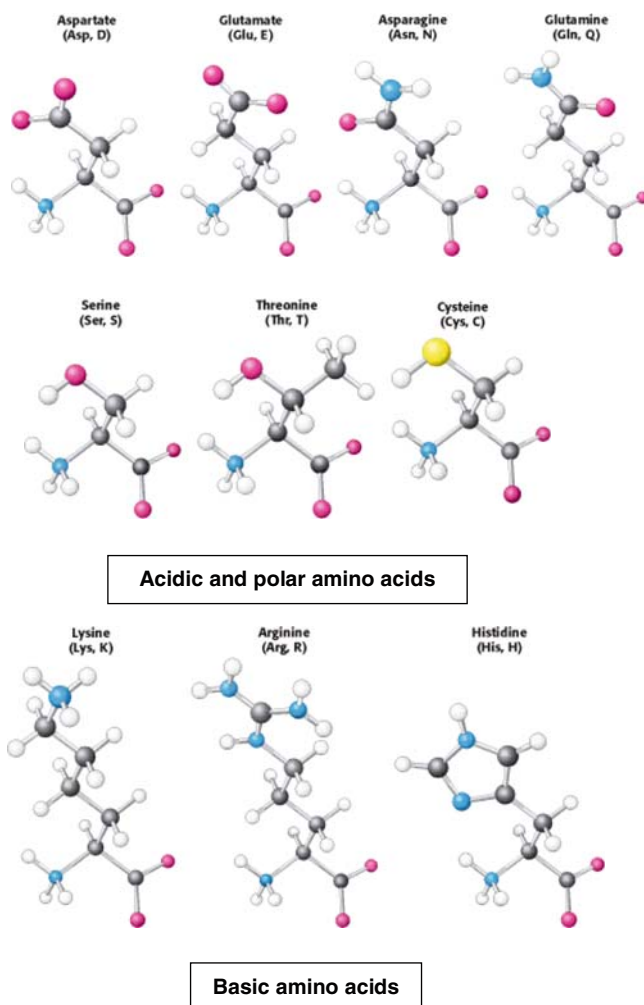
Fig. 8.2 (continued)

can be arranged in either parallel or antiparallel chains. In antiparallel and parallel β-sheets, the distance between chains is 0.325 and 0.347 nm, respectively [3]. Glycine and proline are structurally important amino acids as glycine is the most flexible and is often found at sharp turns in the protein backbone. Proline, on the other hand, is the least flexible amino acid and therefore restricts bending through its rigidity.

X-ray crystallographic analysis can be used to determine levels of protein structure above primary. This method uses X-rays to expose electron densities in a dry protein crystal, thereby revealing structural information. Nuclear magnetic resonance (NMR) measures distances between the nuclei of atoms by magnetically

Table 8.1 Amino acid presence in all proteins

Amino acid	Molecular weight (Da)	Occurrence in proteins (%)
Alanine	71.1	9.0
Arginine	156.2	4.7
Asparagine	114.1	4.4
Aspartate	115.1	5.5
Cysteine	103.1	2.8
Glutamate	128.1	3.9
Glutamine	129.1	6.2
Glycine	57.1	7.5
Histidine	137.2	2.1
Isoleucine	113.2	4.6
Leucine	113.2	7.5
Lysine	128.2	7.0
Methionine	131.2	1.7
Phenylalanine	147.2	3.5
Proline	97.1	4.6
Serine	87.1	7.1
Threonine	101.1	6.0
Tryptophan	186.2	1.1
Tyrosine	163.2	3.5
Valine	99.1	6.9

Source: [3].

stimulating isotopes. Proteins can be in solution for NMR, but both techniques require primary sequence knowledge. These techniques entail mathematical techniques to determine structure. Circular dichroism (CD) is a technique which measures polarized light differences which show structural asymmetry. Infrared (IR) and Raman spectroscopy are vibrational techniques to determine chemical bond energies which aid in analyzing protein secondary structure.

8.2.1.3 Tertiary Structure

The folding of secondary protein structures results in tertiary structure. Noncovalent interactions (Fig. 8.3) largely characterize tertiary protein folding. Hydrophobic interactions exclude water from the interior of globular proteins and are entropically driven. Van der Waals interactions between higher and lower electron densities in hydrophobic residues stabilize the structure. Just as the primary structure dictates the secondary structure, it also impacts tertiary structure. When proteins form α-helices and β-sheets, the type of side chains influence the tertiary structure. α-helices often have hydrophilic residues on one side of the helix and hydrophobic residues on the other side. These aliphatic helices are then well-suited to a globular protein formation with the hydrophilic amino acids on the outside of the protein (facing the aqueous medium) and the hydrophobic groups on the interior of the protein (facing each other). It is important, though, to note that this is a generalization as a mixture of hydrophobic/hydrophilic sections exist on the exterior of proteins. Antiparallel
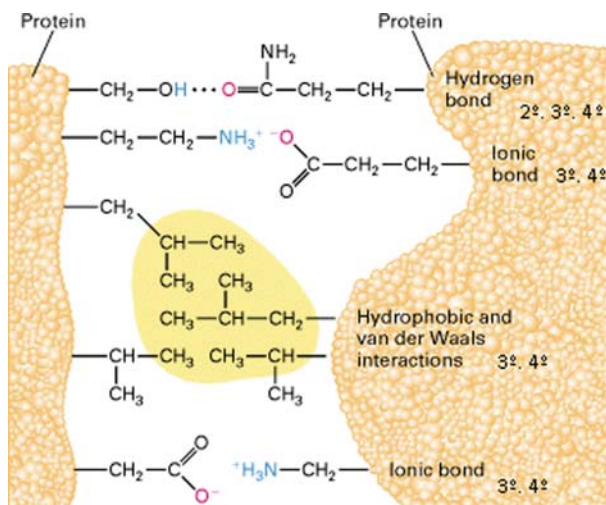
Fig. 8.3 Schematic of potential hydrogen bonding, ionic bonding, hydrophobic and van der Waals interactions between two protein surfaces for secondary (2º), tertiary (3º), and quaternary (4º) structures [4]

β-sheets are usually aliphatic, while parallel β-sheets have hydrophobic chains on both sides and are usually buried on the inside of a globular protein. Hydrophobic interactions involve the nonpolar and aromatic residues: leucine, proline, alanine, valine, methionine, phenylalanine, tryptophan, and isoleucine.

In addition to hydrophobic interactions, other bonds contribute to tertiary protein structures. Two cysteine residues can form a covalent disulfide bond through their sulfur atoms. Hydrogen bonds between polar amino acid residues (such as the alcohols: serine, threonine, and tyrosine and the amides: asparagine and glutamine) participate in tertiary folding. Likewise, ionic bonds can form between charged residues (such as the acids: aspartic acid and glutamic acid or bases: lysine, histidine, and arginine). Protein folding into tertiary shapes is a highly ordered process and can happen spontaneously, but is very difficult to model due to the many folding pathways possible [5].

8.2.1.4 Quaternary Structure

Lastly, the association of monomer subunits into one oligomer structure constitutes quaternary structure. Figure 8.4 displays the four levels of protein structure. There are a few biological predispositions to quaternary protein structure. Shorter polypeptide strands allow for the likelihood of less error than longer strands, so the formation of single polypeptides into monomer units which are then associated with other monomers into oligomers is more logical than the formation of an oligomer from just one polypeptide strand. Additionally, the amount of DNA needed to form one

(a) Primary structure

–Ala–Glu–Val–Thr–Asp–Pro–Gly–

(b) Secondary structure

α belix

β sheet

(c) Tertiary structure

(d) Quaternary structure
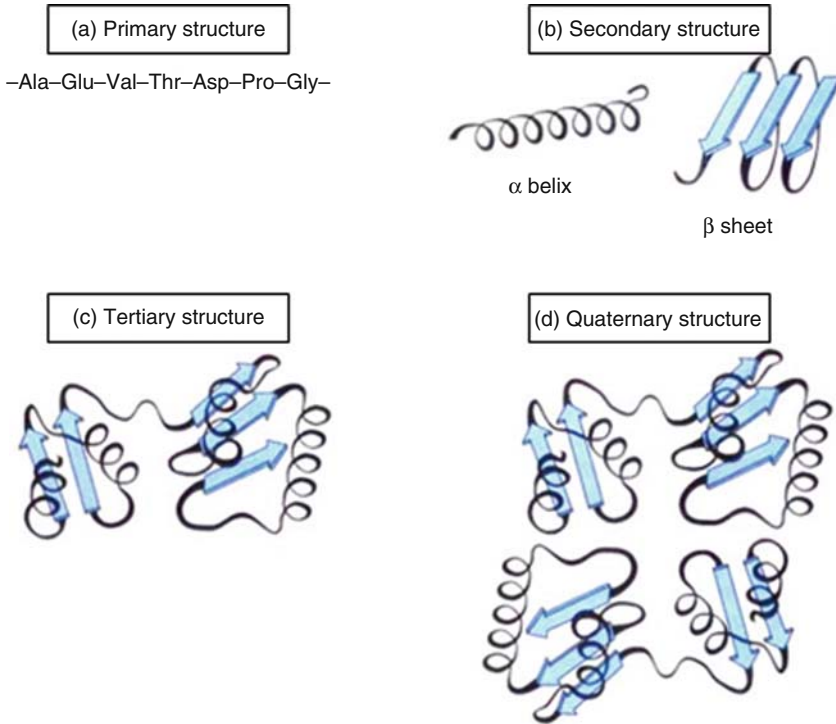
Fig. 8.4a–d Schematics of (**a**) primary, (**b**) secondary, (**c**) tertiary, and (**d**) quaternary protein structures [6]

monomer that will be repeated is less than if the whole oligomer had to be sequenced at one time.

Sometimes the active sites of enzymes are located on adjacent monomers, thus, requiring the interaction of two monomer units to perform catalytic activity on a single substrate. Not only can catalytic activity involve more than one monomer, but this is true for ligand binding as well. When a protein binds to a ligand, its affinity for binding to other ligands, even with its other monomers, may be reduced.

Finally, the surface area to volume ratio is reduced when monomers combine into an oligomer which is energetically favorable. However, the formation of the oligomer results in an entropy loss as the degrees of freedom decrease. This energy loss must be countered with favorable bonding. All of the bond types for tertiary structure are also utilized in quaternary structure. Disulfide bonds are particularly important in quaternary structure to link some monomers together. For example, the extracellular matrix protein, fibronectin, is a dimer with the monomers linked together by a disulfide bond. Insulin has six monomers and hemoglobin has four as originally determined by X-ray diffraction.

8.2.2 Isoelectric Point and Solubility

Due to their aforementioned structures, most proteins are either in globular or fibrous form. Generally, globular proteins are soluble in aqueous solutions and fibrous proteins are not. Globular proteins have a hydrophobic core and present hydrophilic amino acids on their outer surfaces which are in contact with the surrounding aqueous media. Fibrous proteins are usually components of tissues that need structural support such as bones and teeth. They also provide mechanical support at the cellular level and are the structural components of the extracellular matrix that quiescent cells interact with and use for stabilization [7].

Moreover, the large degree of structures proteins may assume provide for unique properties, such as isoelectric points. Proteins have isoelectric points due to charges associated with their amino acid residues. The isoelectric point of a protein is the point where it is uncharged in solution. Since protein molecules would cease to repel each other at this point, the isoelectric point is the least soluble pH value for a particular protein. Uncharged solutes owe their solubility to polar interactions or hydrogen bonding. The solubility of proteins in solution is the point at which equilibrium arises between crystals of the protein and dissolved proteins in solution. The solubility of a protein takes a few days to measure as aliquots are taken until the protein concentration in solution stabilizes [8].

8.2.3 Hydrophobic Composition

Due to the importance of amino acids in determining protein properties, several terms have been developed to describe further the properties of amino acids. For example, hydropathy is the hydrophilic or hydrophobic nature present in each amino acid. In Table 8.2, the hydropathies of various amino acids are listed. The more positive the value in Table 8.2, the more hydrophobic an amino acid is while the more negative the hydropathy, the more hydrophilic an amino acid is. The hydropathy is represented as the free energy of the transfer of an amino acid to water thermodynamically from a lipid bilayer [6].

8.3 Material Surface Properties

Of course, as mentioned, one of the most important aspects of protein structure for numerous biological device applications is how it changes upon interactions with surfaces. The properties of a material at its surface, especially chemistry, roughness, and energy are what elicits water molecule layers to adhere and subsequently mediate the adsorption of different types of proteins. Surfaces in an aqueous environment elicit the adsorption of ordered water layers at the interface with bulk water [9–11]. These water layers are capable of excluding solutes and can be up to $100\,\mu m$ thick with some hydrophilic surfaces [12]. Clearly, the nature of the surface influences

Table 8.2 Hydropathy of amino acid residues

Amino acid	Hydropathy (KJ mol^{-1})
Highly hydrophobic	
Isoleucine	3.1
Phenylalanine	2.5
Valine	2.3
Leucine	2.2
Methionine	1.1
Less hydrophobic	
Tryptophan	1.5
Alanine	1.0
Glycine	0.7
Cysteine	0.2
Tyrosine	0.1
Proline	−0.3
Threonine	−0.8
Serine	−1.1
Highly hydrophilic	
Histidine	−1.7
Glutamate	−2.6
Asparagine	−2.7
Glutamine	−2.9
Aspartate	−3.0
Lysine	−4.6
Arginine	−7.5

Source: [6].

the water interface and subsequent protein adhesion. Differences in the monolayer of proteins adsorbed to material surfaces are based on such properties as surface roughness, energy, and chemistry as will be expanded in the next sections. The atomic structure and bonding interactions can vary considerably even within a material that appears to have the same chemical composition. An example of this is when considering grain size variations which are related to the overall surface roughness. Increased surface roughness can also serve to promote surface energy, thus many surface properties are intertwined and changing one will change others. Lastly, the energy at the surface of a material controls its wettability which many consider to be the most important aspect to promote cell adhesion.

8.3.1 Surface Topography

Until recently, the size of the features on the surface of a biomaterial was not considered to be an important design feature to control protein interactions. However, one important example of how topography can influence protein interactions involves nanomaterials. Materials with micron sized features and up (such as those commonly used in current implants) have significantly different material properties than
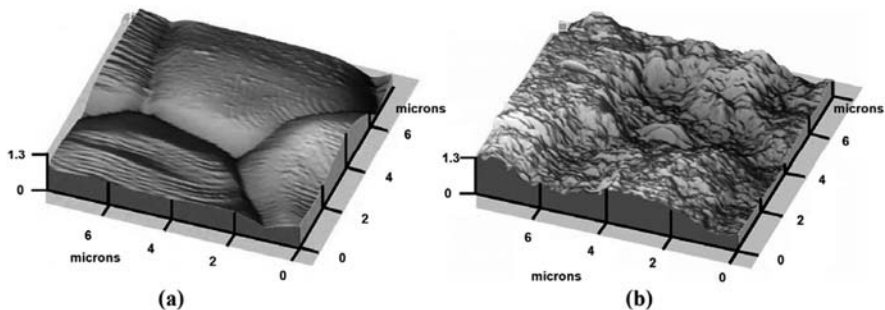
Fig. 8.5 AFM micrographs comparing micronscale surface features with nanoscale surface on titania implants

those with nanoscale features (Fig. 8.5). In fact, the National Science Foundation's definition for nanomaterials states that they not only have smaller constituent sizes (such as grains, particles, fibers, etc.), but also have significantly altered properties compared to conventional materials. Conventional materials have more atoms in bulk and less at the surface, while the reverse is true for nanophase materials. Nanomaterials also have increased defect sites at their surfaces (the highly reactive regions of materials) and therefore greater opportunities for binding to proteins such as enzymes. This gives nanophase materials increased catalytic activity [13] which can be very economical for numerous enzymatic processes.

Other enhanced properties of nanomaterials include superior optical, mechanical, and electrical properties [14]. An example of mechanical property changes when transforming from conventional to nanometer grain sizes is that the smaller grain sizes present in nanophase ceramics slide over one another more easily than in a conventional material; this makes nanophase ceramics less brittle and more ductile when compared to conventional ceramics [15]. But it is now also clear that the unique, biologically-inspired surface roughness of nanomaterials promotes protein interactions important to increase the regeneration of bone, cartilage, vascular, bladder, and nervous system tissues. Nanomaterials are biologically-inspired materials since tissues in our bodies have nanoscale constituent dimensions due to the presence of proteins and hydroxyapatite (as in the case for bone).

Specifically, interactions of two proteins (fibronectin and vitronectin) are significantly enhanced on nanophase compared to conventional materials. Studies have shown that the unique surface energetics present due to nanoscale surface features optimally adsorbs more vitronectin and fibronectin, and once these proteins adsorb, they expose their amino acid sequences (like arginine, glycine, and aspartate; RGD) to a greater extent to promote cell adhesion. Moreover, since proteins are nanoscale entities, investigators are currently matching surface features of nanoscale biomaterials to protruding regions of proteins important for cell functions. In this light, nanometer surface features may have a bright future in controlling protein interactions for a number of tissue engineering applications.

Methods such as gas adsorption, atomic force microscopy (AFM), and scanning electron microscopy (SEM) can be used to assess the surface area, roughness, and porosity of implants. Gas adsorption uses an adsorbent gas such as nitrogen to create a monolayer and then the surface area is calculated since it is dependent on the volume of the adsorbent gas and the cross-sectional area of its molecules. AFM uses a nanoscale tip to scan the morphology of a surface. The tip can also be used to measure force interactions with the surface. In SEM, a surface is interrogated with an electron beam and the resultant secondary or back-scattered electrons provide information about surface topography and composition.

8.3.2 Surface Energy

Another property that influences protein interactions is surface energy. For example, some studies have shown that neutral, and hydrophilic surfaces, tend to manipulate fibronectin conformation to have more RGD binding sites available to cells than charged and hydrophobic surfaces. This trend also holds true for subsequent cell binding interactions. The hydrophilic and neutral surfaces obtain the best cell binding interactions followed by charged and then hydrophobic surfaces [16, 17].

Contact angle analysis can be used to calculate the surface energy of a sample. Since proteins are charged molecules, a surface that is also charged will influence their interactions. By this method, a drop of liquid (often water) is placed on a surface and the angle (θ) made at the liquid, solid, and vapor (air) interface can be measured. As in Fig. 8.6 on the left, if the angle is small, then the surface is wettable, or more hydrophilic. On the other hand, as shown on the right of Fig. 8.6, the angle can be quite large indicating a more hydrophobic surface. In general, if the angle is greater than 90°, then the surface is considered hydrophobic. The surface energy (γ_{sv}) can be calculated as the interfacial tension between the solid and the water drop (γ_{sl}) added to the liquid and vapor surface tension (γ_{lv}) multiplied by the cosine of θ (or the angle created between the liquid and solid interface):

$$\gamma_{sv} = \gamma_{sl} + \gamma_{lv} \cos \theta \qquad (8.1)$$
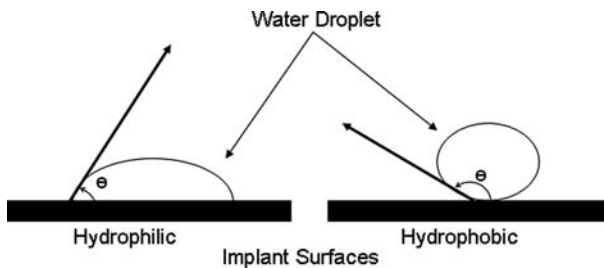
Fig. 8.6 A schematic of contact angle differences on a hydrophilic and hydrophobic surface. θ is the contact angle

Zeta potential is a measure of the charge that forms between the surface and the liquid medium. This potential occurs at the shear plane between the liquid and the surface. Zeta potential (ζ) can be approximated by using the streaming potential (U_s) method in which an electrolyte solution flows past the surface in a channel (L is channel length; A is channel area; R is channel resistance) creating a potential (Fig. 8.7). In (8.2), ΔP is the pressure difference across the channel, μ is the fluid viscosity, ε is the permittivity of the solution, and ϵ_0 is the permittivity of free space:

$$\zeta = \frac{U_s}{\Delta P} * \frac{\mu}{\varepsilon \varepsilon_0} * \frac{L}{A} * \frac{1}{R} \tag{8.2}$$

An understanding of the surface energy available will aid in determining resultant protein interactions.

8.3.3 Surface Chemistry

Surface chemistry is very important for physiological interactions with biomaterials and is closely related to surface energy. It encompasses the type of molecules that reside on the surface of the material as well as their bonding and potentials.
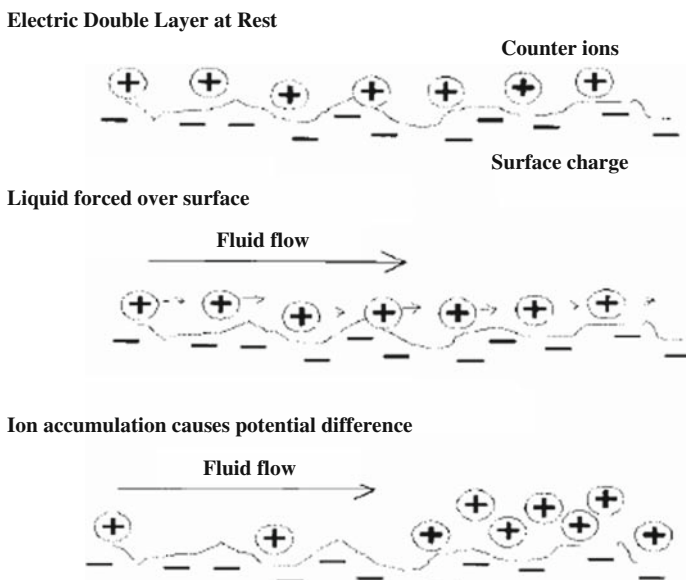
Fig. 8.7 A schematic of the streaming potential used in zeta potential analysis. A surface is shown with its charge and counter ions. When liquid is flowed over the surface, the counter ions accumulate at the end of the channel in the direction of flow creating a potential difference across the channel [18]

As mentioned, cells respond differently to materials of various chemistries which is a direct result of the various protein interactions that occur on materials of altered chemistries. For example, cell migration rates vary on materials of different surface chemistries [19]. Surface chemistry not only influences the amount and types of proteins that adsorb from body fluids, but also affects the conformation of the adsorbed proteins [20–22]. Protein adsorption (fibronectin, fibrinogen, albumin, and others) is much different on materials of altered chemistry and surface energy. All of these factors play vital roles in resultant cell interactions and functions. For example, self-assembled monolayers and microcontact printing are commonly used methods to control surface chemistry for subsequent cellular interactions.

While it is often desirable to promote select protein adsorption for controlling cell adhesion, resisting protein adsorption is important for some applications (such as biosensors) as protein adsorption may interfere with biomolecule exchange necessary for device function. Some materials (such as thiolated dextran) have surface chemistries that are quite protein resistant when tested with proteins such as bovine serum albumin (BSA) [23].

Electron spectroscopy for chemical analysis (ESCA, also referred to as X-ray photoelectron spectroscopy, XPS), auger electron spectroscopy (AES), and MS are some of the methods that can be used to ascertain surface chemistry. XPS detects excited electrons by X-rays that were photoemitted from atoms. AES is similar to XPS, but the sample is excited with an electron beam and the energies determined from secondary electrons, or Auger electrons. Secondary ion MS (SIMS) uses ions to bombard a surface and the surface ions are then emitted and detected to reveal surface chemistry information.

8.4 Protein Adsorption on Surfaces

No matter what specific properties a surface has, most implantable surfaces will adsorb proteins very quickly (in milliseconds) in a monolayer. This layer of proteins that adsorbs to the surface is considered to be generally irreversible. As mentioned, proteins can have different conformations when adsorbing to a surface, such that cell binding sites on the protein can be presented or obscured. Proteins can bind to surfaces with the same types of noncovalent bonding that characterize higher protein structure. Hydrophobic, van der Waals, polar and ionic interactions are all possible [1, 24].

Not only does the type of surface dictate how many proteins will adsorb, and in what conformation they will adsorb, but also what type of protein will adsorb from the numerous possibilities in our blood plasma. As previously mentioned, proteins have various sequences that cells use to adhere. The type of proteins that adsorb does influence the extent of cell adhesion [25–27]. Enzyme linked immunosorbance assays are useful to determine relative amounts of specific proteins that adsorb on different materials. For example, several studies have suggested that a material with

nanoscale features will adsorb greater amounts of vitronectin (a protein important for cell adhesion) than a conventional material.

8.4.1 Kinetics and Thermodynamics

The reason why understanding protein adsorption is so important is that a protein layer can usually be detected within minutes after exposure of a biomaterial surface to a protein solution, and the rate of protein adsorption often levels off after around 1 h [28]. The Langmuir adsorption isotherm for a gas can be used to approximate protein adsorption to a surface. This isotherm models a monolayer as is observed with protein adsorption and includes the assumptions of reversible adsorption with no conformational change or molecular interactions [24]. The Langmuir isotherm is simplified and doesn't account for many of the interactions that occur between a surface and proteins. Certainly, there are more accurate, complex mathematical models that have been developed to understand protein adsorption to surfaces [29]. But in all cases, protein isotherm curves can be recorded by quantifying protein adsorption vs time as in Fig. 8.8.

Surface plasmon resonance (SPR), ellipsometry, and labeling with radioisotopes are methods used to measure protein adsorption kinetics [1]. SPR is an optical technique that measures changes in the refractive index due to protein adsorption. Ellipsometry measures the polarization of the incident and reflected beams from the surface. Labeling proteins with radioisotopes and then quantifying the radioactivity over time is another method, but the labeling itself may alter the protein adsorption kinetics. AFM, CD, IR and Raman methods can also be used to determine protein adsorption kinetics.

The adsorption of proteins to hydrophobic surfaces is driven by entropy whereas the adsorption to hydrophilic surfaces is driven by enthalpy. The former is irreversible to the extent of allowing protein denaturation, but the latter is reversible [24].
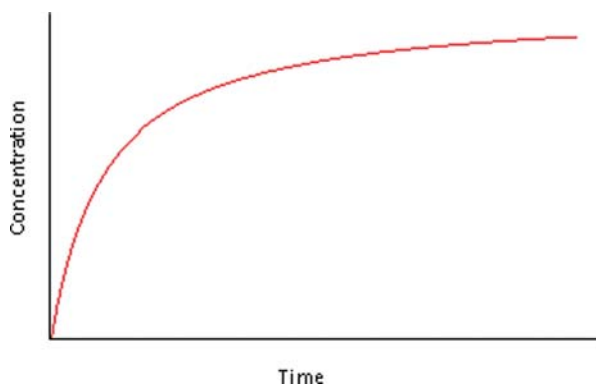
Fig. 8.8 Example of a protein isotherm graph

8.4.2 Density

When assessing the cytocompatibility of surfaces for cell interactions, it is useful to quantify the amount of protein adsorption. While there are a number of methods to determine specific adsorbed protein types and conformations, only a few exist to examine the total quantity of proteins which adsorb to surfaces. Differences in total protein adsorption on various surfaces may explain cell adhesion numbers and cell morphology. To accomplish this, proteins can be desorbed from surfaces by a surfactant, sodium dodecyl sulfate (SDS), which works by disrupting hydrophobic interactions between proteins and surfaces. This enables the determination of total protein quantity in resulting supernatant solutions with the bicinchoninic acid (BCA) protein assay. This assay works on the premise that when proteins are contained in an alkaline medium (OH-), they reduce copper ions ($Cu^{+2} \rightarrow Cu^{+1}$). When BCA is introduced to Cu^{+1}, the solution turns purple. By this simple method, quantities of proteins can be detected by assessing the absorbance of the protein-containing solution at a wavelength of 562 nm on a spectrophotometer. Protein concentrations can be determined by running protein standards (like albumin) at different concentrations during the same assay time that experimental samples are run. Again, such information can prove valuable to determine protein adsorption density and subsequent relationships to cell functions.

8.4.3 Conformation

Another manner in which proteins interact with surfaces that will manipulate cell adhesion is conformation. Proteins undergo a conformational change once adsorbed to the surface of a biomaterial. This conformational change may expose cellular binding sites (Figs. 8.1 and 8.9.). Fibronectin and vitronectin have RGD binding sites which facilitate cellular binding to these proteins via cell membrane integrins. Other proteins, such as laminin which contains binding sequences of tyrosine, isoleucine, glycine, serine, and arginine (YIGSR) and isoleucine, lysine, valine, alanine, and valine (IKVAV) are also important mediators of cell adhesion.
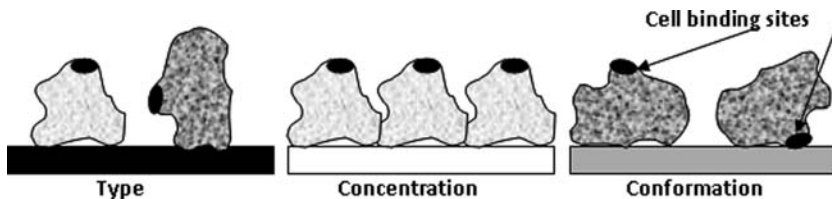
Fig. 8.9 A schematic of protein molecules adhering to a biomaterial surface. Notice that adsorbed proteins can vary in type, concentration, and conformation. Cell adhesion sites are represented by *black dots* and are more available to the cell for adhesion when facing away from the surface

It is difficult to assess the conformation of proteins adsorbed to surfaces. AFM techniques can help to elucidate the conformation of proteins bound to a surface. For example, AFM was useful to determine the orientation of β-amyloids (a peptide implicated in pathogenic plaques of Alzheimer's disease) which correlated to the crystallographic direction of graphite. This interaction was assumed to be driven by hydrophobic forces [30]. Additionally, proteins such as fibronectin can be fluorescently labeled with fluorescent molecules to aid in determining its conformation. One molecule can then excite the other molecule with an emission wavelength and frequency resonance energy transfer (FRET) can then provide information concerning how close the two fluorescent molecules are and therefore reveal protein conformation information.

8.4.4 Extracellular Matrix Proteins

While soluble proteins mediate cell adhesion once adsorbed to material surfaces, extracellular matrix (ECM) proteins play additional roles. These include architecture, mechanical stability, compartmentalization, selective filtration, insulation, defense, nutrition, heat regulation, water equilibrium, and communication. Like soluble proteins, ECMs also influence cell adhesion, proliferation, migration, differentiation, gene expression, morphogenesis, and apoptosis. The size of most ECM proteins are in the nanoscale range as they are all less than 100 nm in size [31, 32].

Collagens are the main component of the ECM. There are about 20 collagens of which all are composed of three helical polypeptides. Other important glycoproteins include fibronectin, laminin, and tenascin which all provide vital guidance functions in embryonic tissue development. Proteoglycans are also glycoproteins, but are much larger complexes that consist of a protein backbone associated with glycosaminolycans (GAG). Some common GAGs include hyaluronic acid, dermatan sulfate, chondroitin 6-sulfate, keratin sulfate, heparin sulfate, and heparin. GAGs are usually sulfated and are the most negatively charged molecules synthesized in the body [33]. These negative ions attract cations such as sodium which causes the ECM to be highly hydrated. Glycoproteins consist of functional domains which can bind with cells, other glycoproteins, or other molecules such as growth factors.

ECM proteins are found in the connective tissue and the basal lamina. These matrices are found in bones and around muscular, fat, vascular and neural tissue [33, 34]. Laminin is a key component of these matrices and without this glycoprotein, nascent ECM formation isn't possible [35]. Figure 8.10 shows the initial laminin matrix assembly. ECM proteins are composed and secreted by cells, however, soluble proteins (such as fibronectin) can be converted by cells from its soluble disc shape to a linear, antiparallel ECM fiber upon adsorption to a biomaterial surface [36]. Vitronectin is responsible for most of the cell adhesion events associated with tissue culture serum [36]. It is for these reasons that, when
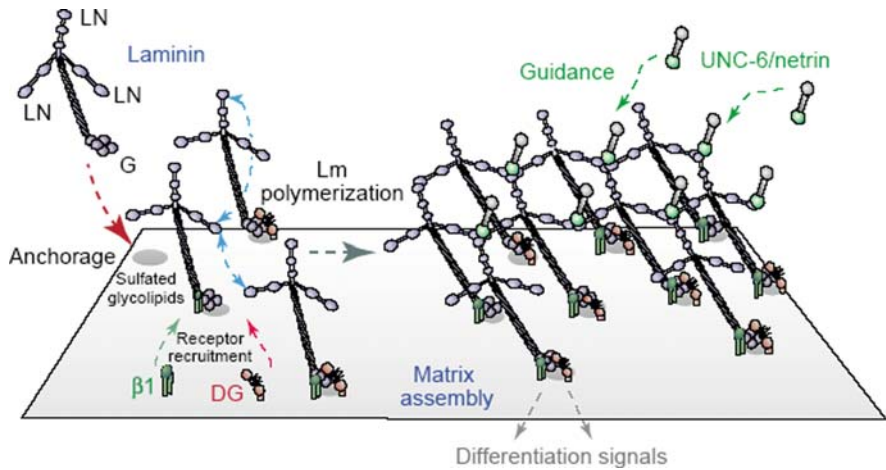
Fig. 8.10 Schematic representations of laminin molecules forming a matrix [35]

discussing protein interactions with surfaces, both soluble and insoluble versions are important.

8.4.5 Cell Adhesive Amino Acid Sequences

Since proteins adsorb to surfaces much faster than cells can adhere, the type and number of proteins that adsorb to the implant surface can control the type of cells that attach to it and how strongly they adhere. Additionally, the conformation of the adsorbed proteins also controls the type and number of cell attachment sites due to the accessibility of binding sites. These binding sites are amino acid sequences that can be recognized by receptors on the cell membranes known as integrins. Integrins span the cell membrane and usually consist of both an α and a β subunit (Fig. 8.11). Examples of common proteins that can adsorb to biomaterials and control cell adhesion are vitronectin, fibronectin, laminin, and collagen.

Cells can have several different types of adhesive interactions with implants mediated by proteins. These include: focal adhesions, close contacts, and extracellular matrix contacts (fibrillar adhesions). The strongest of these is focal adhesion. Focal adhesions have close contact with the surface at a distance of only 10–20 nm. Close contacts often surround focal adhesions, but have a larger gap distance on the order of 30–50 nm. Extracellular matrix contacts are long range adhesions with distances of greater than 100 nm between the cell and surface.

Focal adhesions involve, for example, RGD binding sites and heparin binding domains in proteins such as fibronectin. Heparan sulfate proteoglycans link to both the cell and fibronectin and the cell also attaches to the RGD in fibronectin directly with an integrin [38]. Biomaterial surface chemistry and energy play a large role in focal adhesions on surfaces coated with fibronectin. Previous studies have shown
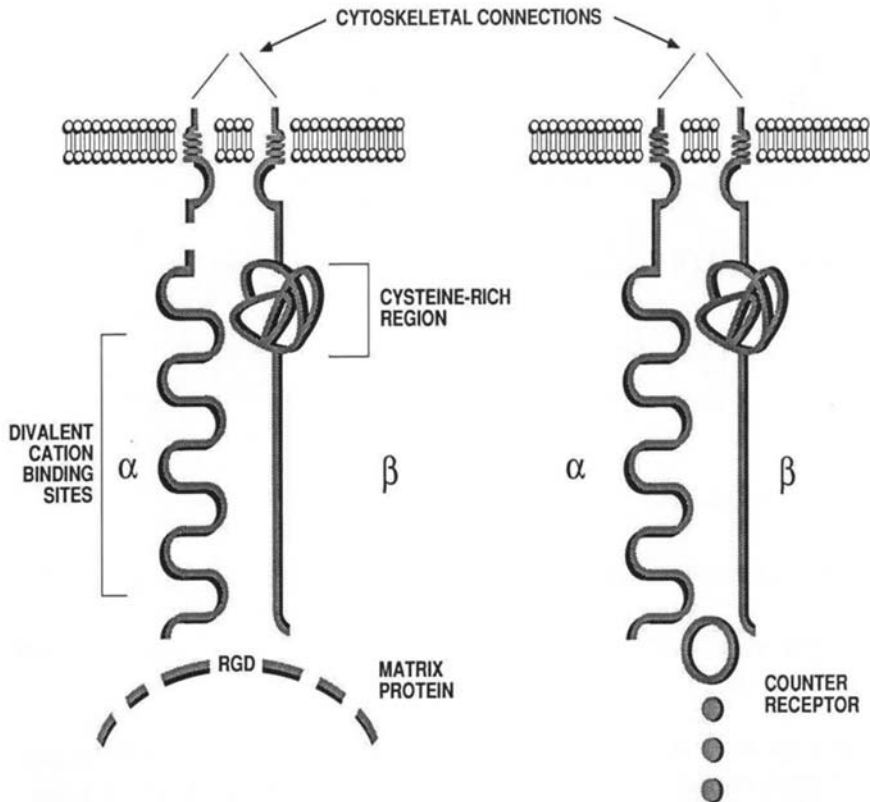
Fig. 8.11 Schematic showing integrin connections to a matrix protein binding sequence or other receptor [37]

that neutral and polar surfaces produce more favorable focal adhesion results than hydrophobic or charged surfaces when all were coated with the same concentration of fibronectin [17].

When a spherical cell adheres to a surface, it begins to spread out on the surface. Sessile cells are dependent on their attachment to a substratum for growth, proliferation, and survival. The attachments that cells have to the extracellular proteins act as mechanosensors, so the cells can mechanically communicate with their environment [39]. Cells organize the proteins they interact with through mechanical interactions [33]. An indication of a cell interacting well with an implanted material is how quickly and completely it spreads on the surface from the spherical form it assumed in aqueous media. The morphology of the spread cell can also provide evidence of how compatible it is with the material. For example, in the case of an osteoblast (bone forming cell), the more spread the cell is into its typical cobblestone morphology, the more likely it is to begin producing new bone.

8.5 Nanoscale Biomaterials

When considering materials for implantation, the reaction of the body to the implant requires careful evaluation. The mechanical properties must also be appropriate to match those of the surrounding material and withstand the normal forces to which the tissue is subjected. One must also consider how the material will withstand chronic implantation. The physiological reaction to the material over time must be considered as well as the way the material itself may degrade. Clearly, though, due to the aforementioned concepts on protein-mediated cell adhesion, one of the most important design considerations for successful implants is the material surface.

Since proteins have nanoscale dimensions, and cells of the body are accustomed to interacting with this size range, nanomaterials are a logical choice for biomaterial design. For example, in Fig. 8.12 a scanning electron micrograph of the basement membrane from the cornea of a monkey is presented [40]. The average fiber and pore diameters in this membrane are approximately 70 nm. Naturally, cells at this location are accustomed to interacting with ECM proteins that are nanoscale in dimension. Pore sizes and surface features affect protein adsorption [41] and therefore cell responses to a material. Studies have shown that nanoscale materials have enhanced properties for cellular interactions such as increased grain boundaries (or
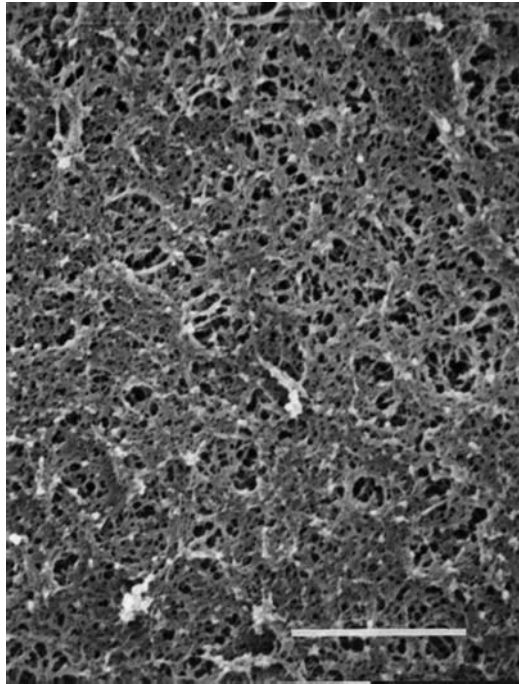
Fig. 8.12 Scanning electron micrograph of the basement membrane from the cornea of a Macaque monkey (*Bar* is 1 μm) [40]

defect sites at the surface) to facilitate optimal protein binding for cell interactions. This is in stark contrast to what is being implanted today, which are materials smooth at the nanoscale.

Quantifying adherent cells, proliferated cells and subsequent cellular functions (such as specific protein production indicative of ECM synthesis) are ways to assess the cytocompatibility of a new biomaterial formulation in the lab. Microscopy and spectroscopy are often used to assess the morphology, viability and functions of cells in contact with novel biomaterials. In vivo (animal) evaluation of new materials is also imperative.

ECMs complete with molecules (such as GAGs) form biological composites. Based on this physiological model, using composite materials that combine mechanical and chemical properties of two or more materials in order to mimic the natural tissue properties as closely as possible, are particularly attractive to design medical implants that will interface well with physiological tissues. For example, carbon fibers represent a very promising reinforcement material to polymer composites while exhibiting promising biological responses [42, 43]. Carbon nanofiber and polyurethane composites contribute promising novel materials for areas as diverse as neural and orthopedic implant applications [43–45]. Most importantly, when these fibers are made to mimic the dimensions of collagen, scar tissue formation is reduced and tissue regeneration is enhanced.

Clearly, looking at the interactions between biomaterial surfaces and proteins that elucidate appropriate cell responses gives us unprecedented control over biological responses. In this way, adverse tissue reactions can be reduced (such as foreign body encapsulation and biofilm formation by bacteria) while implant functionality, longevity, and tissue response can be enhanced.

8.6 Conclusions

Proteins are amazing components of biological systems with diverse and important functionalities. They are stable and resilient, yet have intricate and sometimes fragile architectures at the nanoscale. The interactions between cells and their surrounding ECM is complex and yet essential for understanding how to design biomaterials that will form optimal interfaces with tissues. Since proteins adsorb to a surface upon exposure to biological fluids within milliseconds, insights into the interactions between specific proteins and specific surfaces is a current research focus. Surfaces can be designed to interact optimally with cell adhesive proteins. Using surfaces with nanofeatures not only mimics the physiological size of the natural ECM, but can also enhance properties such as energy without altering chemistry. Clearly, through this understanding, we are on the threshold of formulating biomaterials that will perform better than those developed only 5 years ago. Undoubtedly, the next generation of biomaterials will be designed by considering the biological interactions down to the nanoscale.

References

1. Kasemo B. Biological surface science. Surf Sci, 2002, 500: 656–677.
2. Berg JM, Tymoczko JL, and Stryer L. Biochemistry, W. H. Freeman and Company: New York, 2002.
3. Garrett RH and Grisham CM. Biochemistry, Saunders College Publishing: Fort Worth, 1999.
4. Lodish H, Berk A, Zipursky LS, Matsudaira P, Baltimore D, and Darnell J. Molecular Cell Biology, W. H. Freeman and Company: New York, 2000.
5. Taylor WR and Aszodi A. Protein Geometry, Classification, Topology and Symmetry, Institute of Physics Publishing: Bristol, 2005.
6. Horton HR, Moran LA, Ochs RS, Rawn JD, and Scrimgeour KG. Principles of Biochemistry, Prentice Hall: Upper Saddle River, 1996.
7. Goh C-S, Lan N, Douglas SM, Wu B, Echols N, Smith A, Milburn D, Montelione GT, Zhao H, and Gerstein M. Mining the structural genomics pipeline: identification of protein properties that affect high-throughput experimental analysis. J Mol Biol, 2004, 336: 115–130.
8. Asherie N. Protein crystallization and phase diagrams. Methods, 2004, 34: 266–272.
9. Pollack GH. Cells, Gels and the Engines of Life, Ebner and Sons Publishers: Seattle, Washington, 2001.
10. Ling GN. The physical state of water in living cell and model systems, Ann NY Acad Sci, 1965, 125: 401–417.
11. Ling GN. In Search of the Physical Basis of Life, Plenum Publ. Co.: New York, 1984.
12. Zheng JM and Pollack GH. Long range forces extending from polymer surfaces. Phys Rev E, 2003, 68: 1–7.
13. Takhistov P. Electrochemical synthesis and impedance characterization of nano-patterned biosensor substrate. Biosens Bioelectron, 2004, 19: 1445–1456.
14. Zhang J, Chia Q, Albrecht T, Kuznetsov AM, Grubb M, Hansen LG, Wackerbarth H, Welinder AC, and Ulstrup J. Electrochemistry and bioelectrochemistry towards the single-molecule level: theoretical notions and systems. Electrochim Acta, 2005, 50: 3143–3159.
15. Webster TJ. Advances in Chemical Engineering, Academic Press: New York, 2001, pp. 125–166.
16. Keselowsky BG, Collard DM, and Garcia AJ. Surface chemistry modulates fibronectin conformation and directs integrin binding and specificity to control cell adhesion. J Biomed Mater, 2003, 66A: 247–259.
17. Keselowsky BG, Collard DM, and Garcia AJ. Surface chemistry modulates focal adhesion composition and signaling through changes in integrin binding. Biomaterials, 2004, 25: 5947–5954.
18. Elimelech M and Childress AE. Reverse Osmosis Membranes: Implications for Membrane Performance, U.S. Department of the Interior Bureau of Reclamation: Denver, 1996.
19. Saltzman W. Cell Interactions with Polymers, in Lanza R, Langer R, and Vacanti J, eds. Principles of Tissue Engineering, Academic Press: San Diego, CA, 2000.
20. Michael KE, Vernekar VN, Keselowsky BG, Meredith JC, Latour RA, and García AJ. Adsorption-induced conformational changes in fibronectin due to interactions with well-defined surface chemistries. Langmuir, 2003, 19: 8033–8040.
21. Kowalczynska HM, Nowak-Wyrzykowska M, Dobkowski J, Kolos R, Kaminski J, Makowska-Cynka A, and Marciniak E. Adsorption characteristics of human plasma fibronectin in relationship to cell adhesion. J Biomed Mater Res, 2002, 61: 260–269.
22. Shen M and Horbett TA. The effects of surface chemistry and adsorbed proteins on monocyte/macrophage adhesion to chemically modified polystyrene surfaces. Mater Res, 2001, 57: 336–345.
23. Frazier R, Davies M, Matthijs G, Roberts C, Schacht E, Tasker S, and Tendler S. The self-assembly and inhibition of protein adsorption by thiolated dextran monolayers at hydrophobic metal surfaces, in Ratner BD and Castner D, eds. Surface Modification of Polymeric Biomaterials, Plenum Press: New York, 1996, pp. 117–127.

24. Chinn JA and Slack SM. Biomaterials: Protein-surface interactions, in Palsson B, Hubbell JA, Plonsey R, and Bronzino JD, eds. Tissue Engineering, CRC Press: Boca Raton, FL, 2003, pp. 10-11–10-13.

25. Webster TJ, Schadler LS, Siegel RW, and Bizios R. Mechanisms of enhanced osteoblast adhesion on nanophase alumina involve vitronectin. Biomaterials, 2001, 22: 291–301.

26. Webster TJ, Ergun C, Doremus RH, Siegel RW, and Bizios R. Specific proteins mediate enhanced osteoblast adhesion on nanophase ceramics. J Biomed Mater Res, 2000, 51: 475–483.

27. Price RL, Gutwein LG, Kaledin L, Tepper F, and Webster TJ. Osteoblast function on nanophase alumina materials: Influence of chemistry, phase, and topography. J Biomed Mater Res, 2003, 67A: 1284–1293.

28. Slack S and Horbett T. Changes in the state of fibrinogen adsorbed to solid surfaces: An explanation of the influence of surface chemistry on the Vroman effect. J Colloid Interface Sci, 1989, 133: 148–165.

29. Fang F and Szleifer I. Kinetics and thermodynamics of protein adsorption: A generalized molecular theoretical approach. Biophys J, 2001, 80: 2568–2589.

30. Legleiter J and Kowalewski T. Atomic force microscopy of B-amyloid, in Ricci PCBAD, ed. Atomic Force Microscopy Biomedical Methods and Applications, Humana Press Inc.: Totowa, NJ, 2004, pp. 349–364.

31. Ayad S, Boot-Hanford R, Kadler K, and Shuttleworth A. The Extracellular Matrix Facts-Book, Academic Press: San Diego, CA, 1994.

32. Olsen B. Matrix molecules and their ligands, in Vacanti J, ed. Principles of Tissue Engineering, Academic Press: San Diego, CA, 2000.

33. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, and Walter P. Molecular biology of the cell. Garland Sci, 2002.

34. Bellail AC, Hunter SB, Brat DJ, Tan C, and Meir EGV. Microregional extracellular matrix heterogeneity in brain modulates glioma cell invasion. Int J Biochem Cell B, 2004, 36: 1046–1069.

35. Yurchenco PD and Wadsworth WG. Assembly and tissue functions of early embryonic laminins and netrins. Curr Opin Cell Biol, 2004, 16: 572–579.

36. Yamada KM. Fibronectin and other cell interactive proteins, in Hay ED, ed. Cell Biology of Extracellular Matrix, Plenum Press: New York, 1991.

37. Ruoslahti E. Integrins as receptors for extracellular matrix, in Hay ED, ed. Cell Biology of the Extracellular Matrix, Plenum Press: New York, 1991.

38. Toole BP. Proteoglycans and hyaluronan in morphogenesis and differentiation, in Hay ED, ed. Cell Biology of Extracellular Matrix, Plenum Press: New York, 1991, pp. 305–341.

39. Geiger B, Bershadsky A, Pankov R, and Yamada KM. Transmembrane extracellular matrix-cytoskeleton crosstalk. Nat Rev, 2001, 2: 793–805.

40. Flemming RG, Murphy CJ, Abrams GA, Goodman SL, and Nealey PF. Effects of synthetic micro- and nano-structured surfaces on cell behavior. Biomaterials, 1999, 20: 573–588.

41. Suh CW, Kim MY, Choo JB, Kim JK, Kim HK, and Lee EK. Analysis of protein adsorption characteristics to term-pore silica particles by using confocal laser scanning microscopy. J Biotechnol, 2004, 112: 267–277.

42. Alexander H, Brunski JB, Cooper SL, Hench LL, Hergenrother RW, Hoffman AS, Kohn J, Langer R, Peppas NA, Ratner BD, Shalaby SW, Visser SA, and Yannas IV. Classes of materials used in medicine, in Buddy Ratner AH, Schoen F, and Lemons J, eds. Biomaterials Science, Academic Press, Inc.: San Diego, CA, 1996, pp. 37–130.

43. McKenzie J, Shi R, and Webster T. Material Design for Neural Applications Using Carbon Nanofibers, American Society for Metals International Conference Proceedings: St. Paul, MN, 2004.

44. McKenzie J, Waid M, Shi R, and Webster T. Cytocompatibility of astrocytes on carbon nanofibers. Biomaterials, 2004, 55: 1309–1317.

45. Webster T, Waid M, McKenzie J, Price R, and Ejiofor J. Nano-biotechnology: Carbon nanofibres as improved neural and orthopaedic implants. Nanotechnology, 2004, 15: 48–54.

Chapter 9
Sterility and Infection

Showan N. Nazhat, Anne M. Young, and Jonathan Pratten

9.1 Sterilization

Biomaterials implanted into the body need to be sterile in order to avoid infections which can lead to illness and even mortality. Sterility can be defined as the absence of all living organisms, particularly microorganisms. Indeed, the presence of one bacterium would render a medical device non-sterile. Nearly one-third of implant related infections can be prevented by strictly following established infection control guidelines, but in a period of healthcare cost reductions this is increasingly difficult to sustain.

The sterility assurance level (SAL) minimum, that the probability of a given implant will remain non-sterile after exposure to a given sterilization process, is one in a million implants. Therefore, in order to ensure sterility, a number of physical and chemical sterilization methods are employed. These are discussed below and summarized in Table 9.1. Additionally, the operating theatre area in which the surgical procedure is to be carried out must also be considered. Under aseptic conditions, 5,000–50,000 skin particles are delivered daily from each physician's flora in intensive care units and pathogenic bacteria such as *Staphylococcus aureus* can be recovered from about 90% of "clean" wounds at the time of closure [37].

9.1.1 Steam Autoclaves

Moist heat, in the form of pressurized steam, is the most dependable medium for the destruction of all forms of microbial life. Steam sterilizers (autoclaves) are instruments that produce superheated steam under high pressure, and are used for both decontamination and sterilization. However, they must be properly used to be effective. The effectiveness of steam autoclaving depends upon various loading factors that influence the temperature to which the material is subjected and the contact time. Particular attention must be given to packaging, including the size of containers and their distribution in the autoclave. Material to be sterilized must come in contact with steam and heat and hence the prevention of entrapment of air is critical to achieving sterility. Containers must have good steam permeability and must

Table 9.1 Sterilization procedures carried out for disinfection of biomaterials and medical devices

Procedure	Use	Parameters	Advantages	Disadvantages
Physical methods				
Steam autoclaves	All critical materials that can withstand temperature and pressure, e.g. surgical implements, dressings, contact lenses	Temperature: 121–132 °C Pressure: 15–19 lbs/in^2 (gauge) Air pressure: 0–15 lbs/in^2 (gauge) Exposure time: 5–45 min	Liquids can be sterilized, good steam penetration, 100% effective	Dampens fabric and corrodes metal. Larger autoclaves expensive Packaging and their distribution in autoclave is important
Dry heat	Heat stable dry powder pharmaceuticals, oils, and products that are heat stable but either sensitive to moisture or not penetrated by moist heat	Temperature: 140–170 °C Exposure time: 60–180 min	Simplicity, penetrating power, and lack of toxic residues, 100% effective	Slow procedure, penetration is poor, high temperatures may cause melting
Radiation	Air and surface disinfectant e.g. sutures, gloves, gowns, face masks, dressing, syringes	Dose: 1.5–3.5 Mrad	Kills bacteria, some viruses and some fungi, leaves no residue	Surface sterilization only, limited effectiveness against microbial spores
Chemical method				
Ethylene oxide	Catheters, tracheostomy tubes, mechanical heart valves, sutures	Temperature: 25–75 °C Pressure: subatmospheric to 25 lbs/in^2 (gauge) Gas concentration: 450–1,000 mg/L Relative humidity: 50–80% Exposure time: 1–12 h	Low processing temperature and the wide range of compatible materials	Cost of gas and engineering/environmental controls required to assure safe low residual products

be arranged in the autoclave in a manner that promotes free steam circulation. For example, tight-fitting containers do not permit steam penetration, and thus are not acceptable for use in autoclaves. Stacking containers above one another and overloading an autoclave can also result in poor performance. A chemical indicator (e.g. autoclave tape) must be used with each load placed in the autoclave. The use of autoclave tape alone however is not an adequate monitor of efficiency. Autoclave sterility monitoring should be conducted regularly using an appropriate biological indicator such as *Bacillus stearothermophilus* spore strips. The spores, which can survive $121\,°C$ for 5 min, but are killed at $121\,°C$ in 13 min, are more resistant to heat than most, thereby providing an adequate safety margin when validating decontamination and sterilization procedures.

9.1.2 Dry Heat

Dry heat is less efficient than wet heat and requires longer times and/or higher temperatures to achieve sterilization. It is suitable for the destruction of viable organisms on impermeable non-organic surfaces such as glass, but it is not reliable in the presence of shallow layers of organic or inorganic materials which may act as insulation. Sterilization of glassware by dry heat can usually be accomplished at $160–170\,°C$ for 2–4 h. Dry heat sterilizers should be monitored on a regular basis using appropriate biological indicators as discussed above.

9.1.3 Radiation

Gamma and electron-beam irradiation are among the most popular and well established processes for sterilizing polymer-based medical devices. It has long been known, however, that these techniques can lead to significant alterations in the materials being treated. High-energy radiation produces ionization and excitation in polymer molecules. These energy-rich species undergo dissociation, abstraction, and addition reactions in a sequence leading to chemical stability. Gamma radiation is a penetrating sterilant. Generally no area of a product, its components, or packaging is left with uncertain sterility after such treatment. Even high-density products, such as pre-filled syringes, can be readily processed and used with confidence. Gamma radiation kills microorganisms by attacking their DNA. Both direct and indirect mechanisms are instrumental in the disruption of DNA bonds, which results in the prevention of cellular division and, consequently, propagation.

9.1.4 Ethylene Oxide

Among the sterilization technologies currently available to the medical device industry, 100% ethylene oxide (EtO) gas remains one of the most popular. Ethylene oxide is a chemical gas with alkylating effects. This kills microorganisms by destruction of proteins and nucleic acids. Important parameters for sterilization with

ethylene oxide are the gas concentration, humidity and temperature. As a consequence of the low temperatures during the process, sterilization with EtO is suitable for heat-labile instruments such as endoscopes with sensitive optics, medical utensils or implants. The major problem associated with EtO sterilization arises from difficulties in measuring and controlling the critical physical parameters for the sterilization process. In addition, ethylene oxide is an explosive and toxic gas, and a long degassing period is required after completion of the sterilization process. International standards recommend control with biological indicators for each sterilization cycle due to these elements of uncertainty.

9.1.5 New Technologies

Several new technologies are emerging that have the potential for sterilizing medical devices. These include gaseous chlorine dioxide, low-temperature gas plasma, gaseous ozone and vapor-phase hydrogen peroxide which are being developed as potential alternatives to ethylene oxide. Additionally, machine-generated X-rays are being investigated as a substitute to gamma radiation as they have having similar penetrating power but with the advantage of a non-isotopic source.

9.2 Biomaterials Associated Infections

All biomaterials are prone to infection [4, 5]. The source, and the type of infecting organism as well as the duration of infection depends on the biomaterial being used as well as the time of application which varies from site to site within the body. These are one of the leading causes of nosocomial infections (infections that are acquired while a patient is in a hospital) and primary septicemia (infections of the bloodstream), which result in an adverse affect on the quality of life of the patient, culminating in extended stays in hospitals, and therefore incurring major consumption of health care resources [9].

Once in the body, biomaterials can be in contact with a number of fluids such as blood, urine saliva, urine fluid, tear fluid and tissue fluid, depending on the site of usage. These fluids act as sources of proteins such as fibronectin and fibrinogen, which make up a conditioning film coating the surface of biomaterials and act as receptors to bacterial attachment. The physico-chemical properties of biomaterials' surfaces and their interaction with the bacterial cell play a major role in the adhesion process [2]. Bacterial organisms have the ability to adhere to all types of materials that are made of polymers, glasses, ceramics and metals. Table 9.2 gives the types of implants particularly prone to infection.

9.2.1 Biofilms

Intact host defense systems usually eliminate transient bacterial contamination or colonization, unless the number of organisms exceed threshold levels, host defenses

Table 9.2 Annual frequency of implantable devices usage in USA and rate of infection [4, 5, 19, 12, 16]

Biomaterials/implant/ medical device	Estimated number used annually in USA	Rate of infection
Bladder catheters	30,000,000	10–30%
Central venous catheters	5,000,000	3–10%
Fracture fixation devices	2,000,000	5–10%
Dental implants	1,000,000	5–10%
Prosthetic hip/knee joint	600,000	1–3%
Vascular grafts	450,000	1%
Cardiac pacemakers	300,000	1–7%
Mammary implants	130,000	1–2%
Prosthetic heart valves	85,000	1–2%
Penile implants	15,000	1–3%
Heart assisted devices	700	25–50%

are impaired, tissue surfaces are traumatized, or a foreign body is present [37]. Bacterial adhesion is the first and most important step in implant infection. At 50 nm, reversible, low specificity (e.g. van der Waals') forces are important before short range forces (e.g. hydrogen bonds) create a stronger bond. Additionally, hydrophobic molecules as well as pili and fimbria predispose to bacterial adhesion. Binding of bacteria is governed by factors such as: atomic structure, electronic state, dipole-to-dipole forces, oxidation state and the biomaterial's individual properties such as topography, trace chemicals, and ionic and glycoprotein sequences.

Once adhered, the organisms will usually form a biofilm. Biofilms are matrix-encased communities which are specialized for surface persistence. This community of microorganisms becomes irreversibly attached to a surface and exhibits distinctive phenotypic properties [18]. The presence of biofilms was first demonstrated in the airways of patients with cystic fibrosis, helping to explain the chronic lung infections in this population. There has since been a growing realization that similar biofilms are a factor in almost every aspect of health care. Indeed, biofilms account for over 65% of human bacterial infections [33] and are a major contributing factor to the difficulty of treating infections associated with implanted devices. Examples include infections of the oral soft tissues, teeth and dental implants, gastrointestinal tract, lung tissue, urinary tract prostheses, indwelling catheters, cardiac implants such as pacemakers, prosthetic heart valves, sutures and tracheal and ventilator tubing.

Contact with a solid surface triggers the expression of a panel of bacterial enzymes, which catalyze the formation of sticky polysaccharides that promote bacterial colonization and protection. Bacteria express new, and sometimes more virulent phenotypes when growing within a biofilm. The microorganisms tend to be far more resistant to antimicrobial agents and it becomes particularly difficult for the host immune system to render an appropriate response [42]. With the advent of non-destructive methods for studying organisms in their living, hydrated state (e.g. confocal laser scanning microscopy) it has been shown that biofilms develop not as a uniform layer of cells, as electron microscopy of (dehydrated) biofilms may show,

Fig. 9.1 Three dimensional confocal laser scanning micrograph reconstruction of a mixed-species biofilm growing on hydroxyapatite based on a series of *xy* projections. Within the structure, stacks and water channels are visible. These can function as a primitive circulation system, a characteristic of the structure of organisms growing in this phenotype. Image is 1 mm × 1 mm
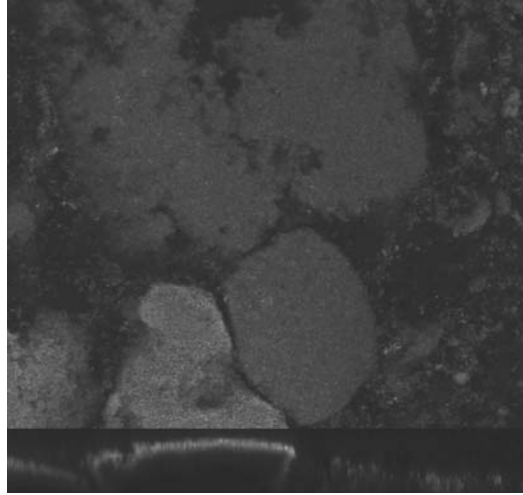
but as discrete stacks of cells separated by channels (Fig. 9.1). Molecular techniques have now identified genes which are up- and down-regulated both between biofilm grown organisms and those grown in suspension, and within different regions of the biofilm. This has important implications in assessing the behavior of organisms in terms of virulence, susceptibility and response to host defense mechanisms. Additionally, the realization that population-dependent gene regulation (quorum sensing) occurs within biofilms has revised the concept that bacteria function autonomously.

The formation of biofilms on medical devices presents three major problems. First, bacterial communities on these surfaces represent a reservoir of bacteria that can be shed into the body, leading to a chronic infection. Second, biofilm bacteria are highly resistant to treatment with antibiotics; therefore, once these bacterial communities form, they are extremely difficult to eliminate with conventional antimicrobial therapies. Finally, because host responses and antimicrobial therapies are often unable to eliminate bacteria growing in a biofilm, a chronic inflammatory response at the site of the biofilm may be produced. The host responses are continually primed by planktonic organisms shed by the biofilm, but insufficient to eradicate the infection. This host response produces a chronic inflammation at the site of the biofilm and may account for tissue damage, separate from that produced by bacteria directly. A non-shedding surface on a medical device encourages the establishment of biofilms.

9.2.2 Types of Medical Related Biofilms

Although a wide range of bacteria and fungi have been shown to be responsible for infections, most infections are actually caused by a small number of species. Staphylococci (predominantly coagulase-negative strains) are considered to be the

Fig. 9.2 Scanning electron micrograph of a single species, *Staphylococcus epidermidis,* biofilm growing on the surface of a titanium implant. Bar $= 20\,\mu$m
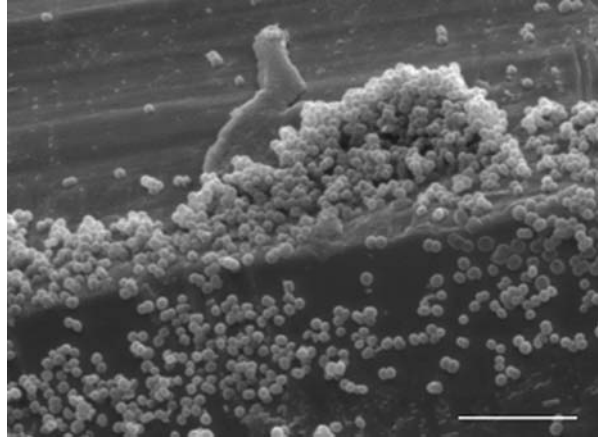
most frequent cause, being responsible for more than half of all infections [22], while streptococci and a collection of Gram-negative organisms account (in approximately equal proportions) for most of the remaining infections. Of the Gram-negative organisms *Escherichia coli*, *Klebsiella pneumoniae*, *Proteus mirabilis* and *Pseudomonas aeruginosa* are the most commonly isolated [17]. These organisms may originate from the skin of patients or health-care workers, tap water to which entry points are exposed, or other sources in the environment. Biofilms may be composed of single- (Fig. 9.2) or multi-species depending on the device and its duration of use in the patient. However there are cases, for example, in urinary catheters, where biofilms are initially composed of single-species but longer exposure leads to a multi-species infection. The microbiological profile of biomaterials associated infections, however, is determined by the type and location of use.

9.2.3 Infections Associated with Implantable Devices

9.2.3.1 Central Venus Catheters

One of the most frequently used biomaterial devices is the central venous catheter (CVC). This is used in situations as varied as the administration of blood products, gastroenterology (to deliver fluids and nutrients) and oncology (to administer cytotoxic drugs). Infection can occur at any time during the use of the catheter, which can last for several months. Microorganisms, most of which are staphylococci, *Klebsiella* spp. and *Candida albicans* [4, 5] are able to enter both from the inside and outside of the catheter.

9.2.3.2 Urinary Catheters

Urinary catheters are tubular latex or silicone devices, which when inserted may readily acquire biofilms on the inner or outer surfaces. The organisms commonly contaminating these devices and developing biofilms are *S. epidermidis, Enterococcus faecalis, E. coli, Proteus mirabilis, P. aeruginosa, K. pneumoniae,* and other Gram-negative organisms. The longer the urinary catheter remains in place, the greater the tendency of these organisms to develop biofilms and result in urinary tract infections. For example, 10–50% of patients undergoing short-term urinary catheterization become infected (7 days) whereas virtually all patients undergoing long-term catheterization (>28 days) become infected. Organism/bacteria adhesion to catheter materials is dependent on the hydrophobicity of both the organisms and the surfaces [10]. Catheters display both hydrophobic and hydrophilic regions and thus allow colonization of the widest variety of organisms. Divalent cations (calcium and magnesium) and increase in urinary pH and ionic strength all result in an increase in bacterial attachment. Certain component organisms of these biofilms produce urease, which hydrolyzes the urea in the patient′s urine to ammonium hydroxide. The elevated pH results in precipitation of minerals such as struvite and hydroxyapatite. These mineral-containing biofilms form encrustations that may completely block the inner lumen of the catheter.

Since the 1960s attempts have been made to incorporate antibiotics into urinary catheters. However, the mixed populations of resistant nosocomial organisms that inhabit the catheterized urinary tract present formidable challenges to any antibiotic impregnated biomaterial. The release of sub inhibitory concentrations of the drugs into the fluids surrounding the catheters has been highlighted to potentially having a negative effect in terms of inducing resistance to resident microbes. Therefore, although this strategy may be initially successful, in the long term it could be counterproductive.

9.2.3.3 Prosthetic Heart Valves

Prosthetic valve endocarditis is a condition caused by the attachment and subsequent biofilm formation on components of mechanical heart valves and the surrounding tissues of the heart. Implantation of the medical heart valve causes tissue damage, and circulating platelets and fibrin tend to accumulate where the valve has been attached. Once again, many of the organisms responsible originate form the skin, however, some can be attributed to other indwelling devices and dental work. The identity of the causative organism is therefore related to its source, whether at time of surgery (*Staphylococcus epidermidis*), from dental work (*Streptococcus* spp.) or from an indwelling material (a variety of organisms).

9.2.3.4 Orthopedic Prosthetic Infections

Over the past five decades, hip or knee replacement surgery has become one of the most frequent prosthetic surgeries in restoring function to disabled arthritic individuals. Currently, more than a million joint replacements are performed

world-wide each year. Second to aseptic loosening, infection is the most frequent complication leading to long periods of hospitalization, morbidity, severe functional impairment and sometimes increased mortality [13]. In the 1970s, prosthesis associated infection was devastating with 16% mortality after treatment of a prosthesis associated infection and only 13% of the survivors left with a functioning prosthesis [25]. Since then, clean air enclosures have been introduced in the operating theatres to prevent the direct contamination of the wound and the biomaterial resulting in significant reductions of infections to 1–4%. However, infections remain a serious problem requiring multiple operations and sometimes leading to amputations and mortality [41]. Most infections are caused by staphylococci which gain access to the device during insertion.

The treatment of infection following total joint arthroplasty involves the combination of surgery and antimicrobial therapy. Surgical methods include debridement and prosthesis retention; re-implantation with a one or two stage exchange arthroplasty; arthrodesis (knee); and excision arthroplasty (shoulder, hip). Antimicrobial therapy should always be combined with surgery. When used alone, as with chronic suppressive treatment of infection, antimicrobial therapy is rarely successful. The duration of infection is an important factor in determining optimal treatment. With an infection of greater than 1 month, the biofilm related bacterial disease would have progressed to such a degree that cure with prosthetic retention is less achievable than with resection. The two stage prosthetic exchange separated by 6 weeks of antibiotic therapy is most common. The prolonged course of antibiotics aimed at the causative bacteria can be administered systemically or locally. Local delivery can be achieved by implanting antibiotic bead chains or antibiotic loaded bone cements. The concept of using bone cement as a carrier for antibiotics is logical as it allows for their delivery at the site of infection. The first time this was carried out was in 1970 [11]. Gentamicin sulfate was the first antibiotic of choice due to its broad-spectrum antimicrobial activity, good water solubility, thermal stability and low allergenicity [14]. Novel and challenging therapeutic approaches have been attempted, particularly in hip prosthetic infections, based on altering the host immune system the type and route of infection, the surgical procedure, the bacteria cultured and the antibiotics employed [7, 8].

The knee, shoulder and elbow revision uses the two stage exchange which involve the removal of the prosthesis and resection of all infected tissue, with the use of antibiotic loaded cement. However there are many variables associated with this technique. These include the use of antibiotic loaded cement in the form of beads or a temporary spacer between the first and second stage; duration of post operative antibiotic therapy; timing of re-implantation, use of allograft bone and selection of a cemented or a cementless implant [35]. The ideal pause between surgeries is not well established, but this interval is normally accepted to be a minimum of 6 weeks, during which antibiotic therapy is prescribed [20]. Once this antibiotic period is completed, and if the results of diagnostic studies indicate the eradication of infections, a new prosthesis is implanted. Although this two stage approach provides a high success rate (about 95%) for eradication of the infection, it frequently results in major morbidity due to prolonged immobilization of the patient.

The one stage revision is mainly applied in hip replacement which involves the excision of all prosthetic components and infected tissues and implantation of new components during the same operation. Although the one stage exchange is attractive as allows earlier mobility, it exposes the patient to the risk that remaining bacteria will lead to re-infection. However, 80% success rates are reported.

9.3 The Use of Antibiotics in the Treatment of Biomaterials Associated Infections

Both systemically administered and locally applied antibiotics or antimicrobial agents have been used with the objective of preventing biomaterials associated infections. However, it is becoming increasingly acknowledged that the treatment of device-associated infections with antibiotics is often destined to fail. The reasons why biofilms are so refractory to antimicrobial agents have not been fully determined but the following have been suggested as contributing factors: protection by the matrix of the biofilm, altered growth rate of the organism, altered phenotype of the organism, entrapment of antibiotic-degrading enzymes, inactivation of the antibiotic by the environment within the biofilm [27, 21]. In some cases, however, such as cerebrospinal fluid shunts, there are additional pharmacological problems with getting the drugs to the site of infection. Despite these problems, precautionary use of antimicrobials is often undertaken in order to help reduce the incidence of infections. For example, in endocarditis antimicrobials are usually administered during valve replacement and whenever the patient has dental work. The aim is to prevent initial attachment by killing any microorganisms introduced into the blood stream. The consequence of the biofilm problem, however, is that treatment failures and relapses are common and in some cases, such as prosthetic hip replacements, surgical removal of the infected device is required as discussed in Sect. 2.2.4.

9.3.1 Systemic Antibiotic Prophylaxis

The ideal antibiotic would be one that has broad-spectrum antimicrobial and is active against the Gram positive and Gram negative species. Antibiotics such as penicillins, gentamicin, and vancomycin are commonly used. An advantage of systemic antibiotic prophylaxis is that it may disseminate via bloodstream to provide activity against organisms remote from the surgical site, as in the case of a patient with an intravascular catheter with a heart valve replacement. Systemic applications of antibiotics may also protect against superficial wound infections [16]. However, not all antimicrobial strategies have been proven to protect significantly against infections with surgical implants due to differences in the microbiology and the duration of time for indwelling implants to become covered with endothelial tissue (in the case of intravascular implants) or fibrotic tissues (in the case of extravascular

implants). The clinical efficacy of a particular antimicrobial approach may also differ among various implants.

Testing the efficacy of systemic antibiotic prophylaxis involves clinical trials which compare a group of patients receiving systemic antimicrobial drugs to a control group receiving placebo. Only a few prospective, randomized clinical trials have demonstrated that systemic antibiotic prophylaxis significantly reduces the rate of infection of implants.

9.3.2 Local Delivery of Antibiotics and Antimicrobial Agents

Since most cases of biomaterials related infection that clinically manifest within 1 year of surgery are thought to result from preoperative inoculation of pathogens, the major purpose of local antibiotic prophylaxis is to prevent the organisms from colonizing the implant and/or contaminating the tissue adjacent to the implant. Table 9.3 compares the four types of local antibiotic prophylaxis.

9.3.2.1 Antimicrobial Irrigation of a Surgical Field

The efficiency of antimicrobial irrigation of a surgical field is affected by a number of variables. These include volume and pressure of irrigation, nature of irrigation flow (i.e. whether it is pulsatile or constant), antimicrobial concentration in the irrigation solution, amount of locally available antimicrobial drugs that either bind to the surgical implant or accumulate in the immediate vicinity, duration (usually less than 1 day) and the choice of antimicrobial agents.

9.3.2.2 Dipping of Biomaterials in Antimicrobial Solutions

The dipping of biomaterials in antimicrobial solutions currently has no established method of application and so results in an undetermined amount of locally available antimicrobial drugs. Further problems with the dipping approach include the incorporation of relatively small amounts of antimicrobial drugs onto the surface of the implant thus giving low systemic detection and a local activity of only a few hours.

9.3.2.3 The Antimicrobial Coating of Biomaterials

Antimicrobial coating of biomaterials possesses the advantages associated with using an established method, knowing the amount of locally available antimicrobial drugs, having a low likelihood of detectable systemic antimicrobial levels, having a persistent local antimicrobial activity lasting from weeks to months, and using a predetermined selection of non therapeutic drug or drugs of choice.

9.3.2.4 Placement of an Antimicrobial Carrier

Antimicrobial therapies can be incorporated in a carrier either preoperatively or during surgery. Impregnation of biodegradable or nondegradable polymers with

Table 9.3 Comparison of locally administered antimicrobial therapies used to prevent biomaterials associated infection [16]

	Antimicrobial irrigation of a surgical field	Dipping of biomaterials in antimicrobial solutions	Antimicrobial coating of biomaterials	Placement of an antimicrobial carrier
Established method of antimicrobial application	No	No	Yes	Yes/no[a]
Known amount of locally available antimicrobial[b]	No	No	Yes	No
Detectable antimicrobial levels in serum	Variable[c]	Unlikely	Unlikely	Unknown
Duration of local antimicrobial activity	<1 day	Hours	Weeks to months	Weeks to months
Utilizes drugs of choice	Surgeon dependent	Surgeon dependent	No	Yes
Demonstrated clinical efficacy[d]	No	Yes[e]	No	No

[a]Yes, for preoperatively prepared antimicrobial carriers; no, for intraoperatively assembled antimicrobial carriers

[b]Locally available antimicrobials bind to the surgical implant or accumulate in the vicinity of the implant

[c]The likelihood of detecting antimicrobial levels in serum varies on implant location (i.e. intravascular or extravascular) and the total amount of applied antimicrobials

[d]As demonstrated by significant reduction in the rate of implant associated infection in a prospective, randomized clinical trail

[e]Dipping of prosthetic heart valves has been evaluated in a single prospective, randomized clinical trial [34]

antimicrobials is often performed intraoperatively in patients with infected orthopedic prosthesis. The local antimicrobial activity usually persists for weeks to months and is dependent on the rate of degradation of biodegradable carriers or on the timing of surgical removal of nondegradable carriers. However, the clinical efficacy of this approach has not been demonstrated in a prospective randomized fashion.

9.4 Developing Infection-Preventing Biomaterials

Despite ongoing advances in the treatment of biomaterials associated infections the results are moderate at best and achieved at considerable cost to the patient [24] and prevention still remains the best form of treatment. Therefore, there is increasing

demand for infection-preventing biomaterials that have antimicrobial activity and/or resist biofilm colonization. The activity should endure for the lifetime of the device, not be reduced by contact with body fluids and not encourage drug resistant organisms. Although many antimicrobial biomaterials have shown promising activity in vitro, few anti-infective prosthetic devices manufactured from these materials have yet achieved any degree of success in clinical trials.

This chapter has already covered drug-releasing materials such as antimicrobial carriers where the agent is added to the material during manufacture or even introduced into the material post manufacture. In both cases, the drug is dispersed throughout the material. Other coating or surface modifications are also, however, increasingly being considered. The physico-chemical design and morphology of a biomaterial can significantly influence the risk of colonization. One approach in reducing the attractive forces between bacteria and implant is to modify the surface properties of the biomaterials. Most organisms are negatively charged; therefore a negatively charged surface should exert a repulsive electrostatic force on the organism. Furthermore bacterial adhesion is expected to be reduced on hydrophobic surfaces. Protein coating is another potential method to prevent adhesion. Albumin and heparin for example are macromolecules of biological origin used as coatings to decrease the adhesiveness of biomaterials.

Various other antimicrobial coatings have been developed to address the problem of biomaterials associated infections. In general there are two types of approach: passive or active. Passive antimicrobial coatings bind the antimicrobial agents permanently onto the device. The drugs are not released but the microorganisms are killed on surface contact. Active antimicrobial coatings use the principles of controlled drug release. These incorporate chemical compounds into the coating matrix for release into the vicinity of devices to perform bactericidal functions [40].

Metal ions such as gold, silver and copper all have broad-spectrum antimicrobial activity. Silver in particular has high antimicrobial activity with low human toxicity and has been used as an antiseptic compound since the fourteenth century. Silver cations are highly reactive chemical structures that strongly bind to electron donor groups containing sulfur, oxygen and nitrogen, which are generally present in biological molecules in the form of thio, amino, imidazole, caboxylate, and phosphate groups. Inorganic silver salts and complexes such as argyrols, silver nitrate, carbonate, silver halides, and the metal itself are used in various medicines [39]. The biocidal activity of silver is related to the biologically active silver ions by binding to the bacterial DNA, which may inhibit a number of important transport processes (such as phosphate and succinate uptake) and can interact with cellular oxidation processes [29]. Several methods have been proposed for surface coating with elemental silver including plasma deposition or spluttering of silver coatings and ion beam assisted deposition /ion implantation techniques using a vacuum process. Silver has activity against a wide range of nosocomial organisms, including multiresistant *staphylococci* and *entercocci*, Gram-negative *enteric bacilli*, *Pseusdomonas aeruginosa* and *Candida albicans*.

9.5 Case Study: Oral Infections and Biomaterials

Two of the most common diseases of man include dental caries and periodontitis. These cause destruction of tooth enamel and dentine or periodontal tissues such as the alveolar jaw bone and periodontal ligament (Fig. 9.3). Estimates suggest 10–15% of the worldwide population is affected by periodontitis [23]. Other frequent infections in the mouth affecting the pulp are termed Periapical or may be associated with implants (Peri-implantitis). Micro-organisms can reach the pulp and travel to the apex of the tooth through damaged enamel and dentine tubules.

Bacteria in the mouth generally form plaque biofilms in which they are protected from both host defense mechanisms and antibacterial agents. Environmental changes can lead to ecological shifts and subsequent population changes which may predispose to a more "pathogenic" microbial community. Hence, different diseases are associated with a dominance of certain organisms at the infected site (Table 9.4). Although with some infections complete bacterial irradiation is required, in the mouth where it is impossible to achieve this goal, it may be preferable to change the bacterial composition to one which is considered less pathogenic.

9.5.1 Dental Caries and Periapical Disease

Cariogenic bacteria produce organic acids that cause localized destruction of the tooth enamel. Regular brushing, mouth rinsing and dental flossing with additional professional treatments such as scaling or application of slow antibacterial agent release varnishes can reduce plaque build up. Toothpastes contain fluoride that reacts with hydroxyapatite in the tooth forming more acid resistant fluorapatites.
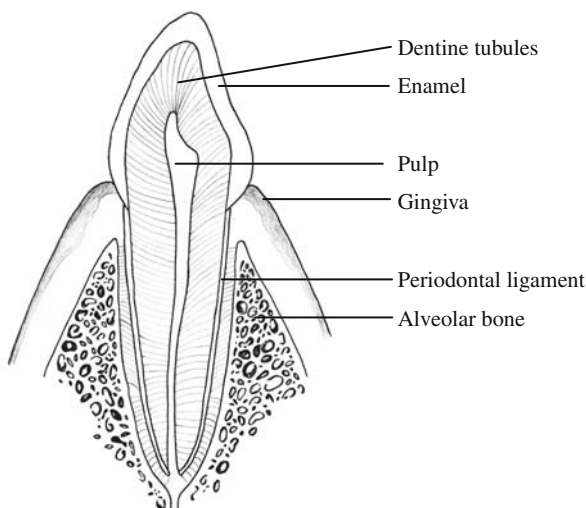
Fig. 9.3 Cross section through a tooth and supporting tissues

Table 9.4 Bacteria associated with various oral diseases

Disease	Examples of causative organisms
Caries	*Actinomyces* spp., *Lactobacillus* spp., *Streptococcus mutans*
Periodontitis	*Actinobacillus actinomycetemcomitans, Capnocytophaga* spp., *Eubacterium* spp., *Fusobacterium* spp., *Peptostreptococcus* spp., *Porphyromonas gingivalis, Prevotella* spp., *Tannerella forsythensis*
Pulpal and periapical	*Fusobacterium nucleatum, Peptostreptococus* spp., *Porphyromonas gingivalis, Prevotella intermedia*
Osteomyelitis	*Staphylococcus aureus, Staphylococcus epidermidis*

In its ionic form, fluoride can also be toxic to bacteria, but relatively high concentrations are required. Mouthwashes and slow release devices contain various antibacterial agents including chlorhexidine, cetylpyridinium chloride and thymol (Fig. 9.4). Once bacteria erode the enamel, however, diseased tissue must be surgically removed and replaced with a restorative material.
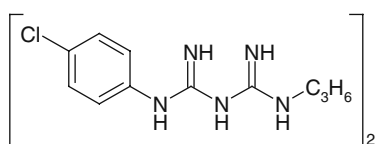
Chlorhexidine
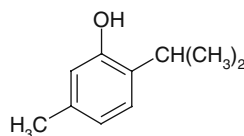
Thymol
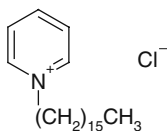
Cetyl pyridinium chloride
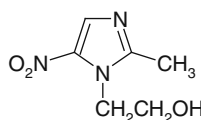
Metronidazole
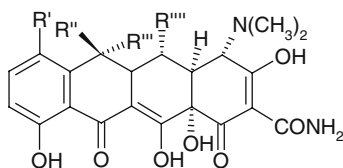
Tetracycline, Doxycycline, Minocycline
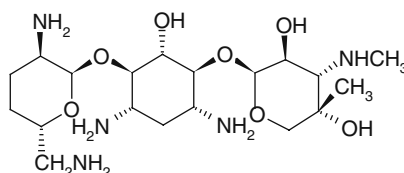
Gentamicin

Fig. 9.4 Chemical structures of selected antimicrobials

Fig. 9.5 Electron
micrograph showing the gap
between dentine and an
amalgam restoration.
Bar $= 100\,\mu$m
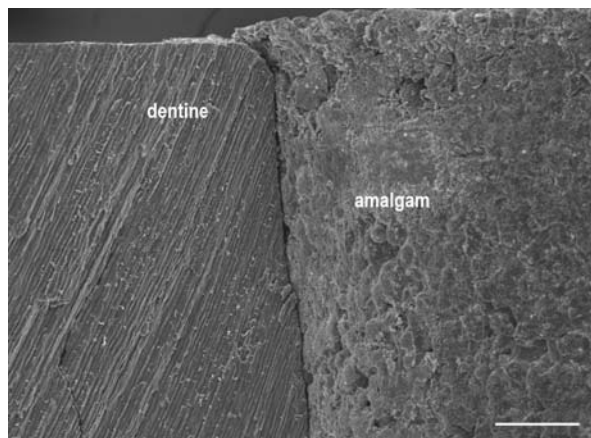
A major problem with all dental restorations is microleakage. This is the passage
of bacteria and other species through the micron-sized gaps between the restoration and dentine (Fig. 9.5). Such gaps can be enhanced by restorative material
dimensional change during set or thermal stresses and mechanical loading [15].
Microleakage can cause sensitivity, marginal discoloration, filling degradation, secondary caries, pulpal inflammation and subsequent apical periodontitis.

The main choice as a dental restorative material has traditionally been amalgam.
Due largely to patient demand, this is increasingly being replaced by, more esthetic,
non-mercury-containing materials. A summary of dental restorative materials typically used is given in Table 9.5. Two basic direct (i.e. set in the mouth) esthetic
restorative materials are composites and glass-ionomer cements (GICs). Composites consist of glass particles embedded in hydrophobic methacrylate monomers
that set forming a crosslinked polymer network on exposure to a blue dental light.
These are bonded to the tooth using more hydrophilic and fluid methacrylate based
adhesives. The polymerization setting reaction, however, leads to shrinkage, damaging the adhesive bond and increasing marginal gaps [31]. This can be partially
overcome through build up of the material in thin layers but higher levels and a

Table 9.5 Compositions of dental restorative materials

Dental restorative	Typical chemical composition (approximate wt%)
Amalgam	Mercury (50), Silver (25), Tin (15) and Copper (10)
Composite	Inorganic glass particles (80) and dimethacrylate monomers with light cure initiators (20)
Glass ionomer cement	Fluoroaluminosilicate glass particles (76), water (17) and polyacrylic acid (17)
Resin–modified glass ionomer cement	Silane–treated fluoroaluminosilicate glass particles (75), polyacrylic acid (10), methacrylate monomers with light and chemical cure initiators (10), water (5)

greater diversity of bacteria have still been detected beneath composites in extracted teeth than below amalgam [38].

GICs are set by reaction of an aqueous polyacid solution (often polyacrylic) with a fluoro-alumino silicate glass. Water sorption by GICs restricts shrinkage and promotes fluoride release. Unfortunately, these cements are too weak to be used in load-bearing applications [30]. To improve the mechanical properties of GICs they have been modified with methacrylate resins. Recent resin–modified glass ionomer cements (RMGICs) have mechanical properties between GICs and composites and fluoride release levels comparable with GICs [43].

Based on clinical investigations, levels of GIC fluoride release can only partially arrest caries progression. Various alternative antibacterial agents have therefore been added to dental materials to try to reduce bacterial microleakage, but much of this work has only been in vitro and has yet to reach clinical trials. Chlorhexidine for example has been added to GICs [32] and RMGICs [36]. In addition, silver sol gel particles [44] and antibacterial monomers [26] have been included within composites. With all these modifications, one problem is the limited release of the antibacterial agent. Any antibacterial exposed surface that does not release its active ingredients is likely to become rapidly ineffective in the mouth due to coverage with a polymeric pellicle layer. In the case of chlorhexidine, the drug is trapped in the GIC by interaction with polyacrylic acid. Chlorhexidine has been released from experimental composite adhesives but the release rate is substantially reduced with increased hydrophobicity and polymer crosslinking [46].

Although some antibacterial agent modified dental restorative materials have shown improved antibacterial behavior in vitro, this may not translate to any significant effect in vivo. Many studies have investigated the antibacterial properties of dental materials by using relatively simple microbiological tests. Such methods, however, are often a poor assessment of efficacy of materials in the mouth where organisms are protected within biofilms as discussed previously.

Biofilms can be grown in vitro using apparatus to model the oral cavity, one such model being the constant depth film fermentor (CDFF). Using this technique, biofilms have been observed to develop marginally less on exposed GIC or RMGIC surfaces than on a composite with no fluoride release. Bacterial microleakage into a GIC or RMGIC bovine dentine restoration has, however, been demonstrated, using a CDFF, to be very much reduced [45] compared with an amalgam filling despite the latter's high level of antibacterial metal ions [28]. This may be due more to the better adhesion and bonding of the cements to dentine rather than any antibacterial properties. Indeed, polymerization shrinkage allows higher numbers of bacteria to penetrate a composite compared with amalgam restoration in the model (Fig. 9.6).

In order to overcome bacterial microleakage with esthetic restorations in load-bearing situations, GICs and RMGICs can be used beneath a composite using a sandwich or layering technique [3]. These are, however, difficult to place and there are additional problems with potentially poor bonding between the various layers. There is therefore still a great need and a huge market for a new dental material with high mechanical properties that can also prevent bacterial microleakage.

Fig. 9.6 A composite restoration surface. After 10 weeks within the CDFF, bacteria can form an extensive gap-filling biofilm between a composite and dentine. This consists of a wide range of bacteria embedded within polymers. $Bar = 5\,\mu m$
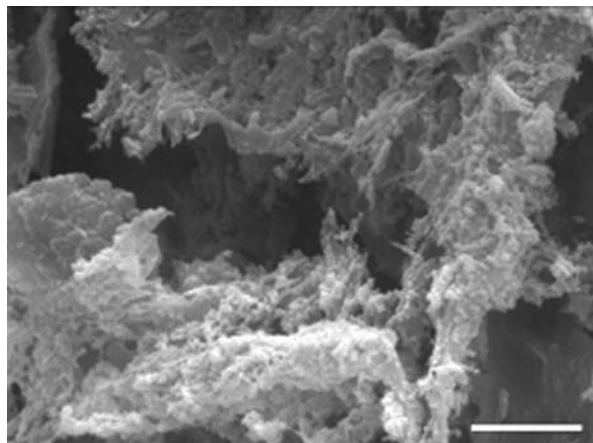
9.5.2 Periodontal Disease

Periodontal disease is the usual sequel to untreated bacteria/plaque induced swelling of the gingiva (gingivitis). The disease causes formation of an increased periodontal pocket that bleeds with probing. Within this pocket, the bacterial population shifts towards mainly gram negative anaerobic bacteria which cause bone loss. Treatments include mechanical scaling, root planning and in some cases, surgery to encourage bone regeneration and epithelium re-attachment. Although, in combination with improved oral hygiene, these treatments are generally sufficient to stop periodontal destruction, in severe, aggressive or some non-responding sites other antimicrobial options are advocated [6]. Systemically applied antimicrobials have side effects such as hypersensitivity, gastrointestinal intolerance, development of bacterial resistance and low local tissue concentration at the required site [1]. Various sustained local delivery formulations have, therefore been developed for the periodontal pocket (see Table 9.6). Some of these formulations have also been used for treatment of infections around oral implants (perio-implantitis). Chemical structures of the antimicrobial agents used in these devices are provided in Fig. 9.4.

Carriers for periodontal antimicrobials have been in fiber, microsphere, liquid, gel or chip form and produced from both degradable and non-degradable polymers. The former have the disadvantage of requiring removal after treatment. Drug release is via a combination of diffusion and carrier degradation mechanisms. Rates of diffusion of drugs from the polymers into a surrounding aqueous phase can be reduced by enhancing polymer crosslinking or hydrophobicity. With degradable polylactides rates of degradation can be enhanced by production of copolymers with glycolide.

The most common antibiotics in current controlled release periodontal devices are tetracycline including tetracycline itself, minocycline and doxycycline. Typical minimum inhibitory concentrations of tetracycline for subgingival microorganisms are $\sim$1–2 μg/mL but some bacterial species are more resistant and others develop

Table 9.6 Commercial antimicrobial agent releasing periodontal formulations

Trade name	Company	Dosage form	Main carrier (*degradable)	Drug	Drug loading
Actiste	Alza	Fiber	Ethylene vinyl acetate	Tetracycline HCl	25%
Arestin	OraPharma	Micro-spheres	Polylactide-co-glycolide*	Minocycline HCl	1 mg
Atridox	Atrix Lab	Liquid	Polylactide in N-methyl-2-pyrrolidone*	Doxycycline hyclate	10%
Dentomycin	Lederle Lab	Gel	Hydroxyethyl cellulose	Minocycline	2%
Elyzol	Dumex-Alpharma	Gel	Glycerol monooleate in sesame oil	Metronidazole	5–10%
Periochip	Perio Products	Chip	Crosslinked gelatine + glycerine*	Chlorhexidine gluconate	33% 2.5 mg

resistance after exposure to sub-inhibitory concentrations. Atridox and Periochip gingival crevicular fluid GCF levels are both approximately 1,000 µg/mL initially but this decreases by a factor of 10 by 7 days after placement. GCF bioactive antibiotic levels from Atrigel and Actiste exceed 250 µg/mL over 7 days and 600 µg/mL over 10 days respectively. This drug level is generally effective against oral bacteria even within biofilms.

Some of the disadvantages of current formulations include difficulty in placement and movement from the site of infection providing opportunities for future developments. Clinical trials using the antibacterial periodontal devices in combination with current treatments have also demonstrated only limited advantage partially because of the high success rates of the conventional mechanical methods alone. Further clinical trials with the antibacterial devices in difficult to treat cases are therefore still required, to provide irrefutable evidence for their beneficial effects.

References

1. Addy M and Martin MV. Systemic antimicrobials in the treatment of chronic periodontal diseases: a dilemma. Oral Dis, 2003, 9: 38–44.
2. An YH and Friedman RJ. Concise review of mechanisms of bacterial adhesion to biomaterial surfaces. J Biomed Mater Res, 1998, 43: 338–348.
3. Andersson-Wenckert IE, van Dijken JW, and Kieri C. Durability of extensive class II open-sandwich restorations with a resin-modified glass ionomer cement after 6 years. Am J Dentistry, 2004, 17(1): 43–50.
4. Bayston R. Medical problems due to biofilms: Clinical impact, aetiology, molecular pathogenesis, treatment and prevention, in Newman HN and Wilson M, eds. Dental Plaques Revisited: Oral Biofilms in Health and Disease, Bioline: Cardiff, UK, 1999.
5. Bayston R. Biofilms in medicine and disease: An overview, in Wimpenny J, Gilbert P, Walker J, Brading M, and Bayston R, eds. Biofilms: The Good, the Bad and the Ugly, Bioline: Cardiff, UK, 1999.
6. Beikler T, Prior K, Ehmke B, and Flemmig TF. Specific antibiotics in the treatment of periodontitis – a proposed strategy. J Periodontol, 2004, 75: 169–175.
7. Berbari EF, Hanssen AD, Duffy MC, Steckelberg JM, Ilstrup DM, Harmsen WS, and Osmon DR. Risk factors for prosthetic joint infection: case control study. Clin Infect Dis, 1998, 27: 1247–1254.
8. Bernard L, Hoffmeyer P, Assal M, Vaudaux P, Schrenzel J, and Lew D. Trends in the treatment of orthopaedic prosthetic infections. J Antimicrob Chemother, 2004, 53: 127–129.
9. Boxma H, Broekhuizen T, Patka P, and Oosting H. Randomised controlled trial of single dose antibiotic prophylaxis in surgical treatment of closed fractures: the Dutch Trauma Trial. Lancet, 1996, 347: 1133–1137.
10. Brisset L, Vernet-Garnier V, Carquin J, Burde A, Flament JB, and Choisy C. In vivo and in vitro analysis of the ability of urinary catheters to microbial colonization. Pathol Biol, 1996, 44: 397–404.
11. Buchholz HW and Englebrecht H. Depot effects of various antibiotics mixed with Palacos resins. Chirurg, 1970, 41: 511–515.
12. Burrows LL and Khoury AE. Infection of medical devices, in Wnek GE and Bowlin GL, eds. Encyclopedia of Biomaterials and Biomedical Engineering, Marcel Dekker Inc: New York, 2004.
13. Charnley J. Postoperative infection after total hip replacements with special reference to air contamination in the operating room. Clin Orthop, 1972, 87: 167–187.

14. Ching DW, Gould IM, Rennie JA, and Gibson PH. Prevention of late haematogenous infection in major prosthetic joints. J Antimicrob Chemother, 1989, 23: 676–680.
15. Combe E. Dental Biomaterials, Kluwer Academic Publishers: Boston/London, 1999.
16. Darouiche RO. Antimicrobial approaches for preventing infections associated with surgical implants. Clin Infect Dis, 2003, 36: 1284–1289.
17. Donlan RM. Biofilms and device-associated infections. Emerging Infect Dis, 2001, 7: 277–281.
18. Donlan RM and Costerton JW. Biofilms: Survival mechanisms of clinically relevant microorganisms. Clin Microbiol Rev, 2002, 15: 167–193.
19. Furno F and Bayston R. Antimicrobial/antibiotic (infective resistance) materials, in Wnek GE and Bowlin GL, eds. Encyclopedia of Biomaterials and Biomedical Engineering, Marcel Dekker Inc: New York, 2004.
20. Garvin KL, Fitzgerald RH Jr, Salvati EA, Brause BD, Nercessian OA, Wallrichs SL, and Ilstrup DM. Reconstruction of the infected total hip or knee arthroplasty with gentamicin-impregnated Palacos bone cement. Instr Course Lect, 1993, 42: 293–302.
21. Gilbert P, Das J, and Foley I. Biofilm susceptibility to antimicrobials. Adv Dent Res, 1997, 11: 160–167.
22. Götz F. Staphylococcus and biofilms. Mol Microbiol, 2002, 43: 1367–1378.
23. Henderson B and Nair SP. Hard labour: bacterial infection of the skeleton. Trends Microbiol, 2003, 11: 570–577.
24. Hendriks JGE, van Horn JR, van der Mei HC, and Busscher HJ. Backgrounds of antibiotic loaded bone cement and prosthesis related infection. Biomaterials, 2004, 25: 545–556.
25. Hunter G and Dandy D. The natural history of the patient with a total hip replacement. J Bone Joint Surg Br, 1977, 59: 293–297.
26. Imazato S, Russel RR, and McCabe JF. Antibacterial activity of MDPB polymer incorporated in dental resin. J Dent, 1995, 23: 177–181.
27. Mah TF and O'Toole GA. Mechanisms of biofilm resistance to antimicrobial agents. Trends Microbiol, 2001, 9: 34–39.
28. Matharu S, Spratt DA, Pratten J, Ng YL, Mordan N, Wilson M, and Gulabivala K. A new in vitro model for the study of microbial microleakage around dental restorations: A preliminary qualitative evaluation. Int Endod J, 2001, 34: 547–553.
29. Modak SM and Fox CL. Binding of silver sulfadiazine to cellular components of *Pseudomonas-aeruginosa*. Biochem Pharmacol, 1973, 22: 2391–2404.
30. Nicholson JW. Chemistry of glass-ionomer cements: a review. Biomaterials, 1998, 19: 485–494.
31. O'Brien WJ. Dental Materials and Their Selection, 2nd edn, Quintessence: Illinois, 1997, pp. 97–113.
32. Palmer G, Jones FH, Billington RW, and Pearson GJ. Chlorhexidine release from an experimental glass ionomer cement. Biomaterials, 2004, 25: 5423–5431.
33. Potera C. Microbiology – forging a link between biofilms and disease. Science, 1999, 283: 1837–1838.
34. Riedl CR, Plas E, Hubner WA, Zimmerl H, Ulrich W, and Pfluger H. Bacterial colonization of ureteral stents. Eur Urol, 1999, 36: 53–59.
35. Salvati EA, Gonzalez Della Valle A, Masri BA, and Duncan CP. The infected total hip arthroplasty. Instr Course Lect, 2003, 42: 223–345.
36. Sanders BJ, Gregory RL, Moore K, and Avery DR. Antibacterial and physical properties of resin modified glass-ionomers combined with chlorhexidine. J Oral Rehabil, 2002, 29: 553–558.
37. Schierholz JM and Beuth J. Implant infections: a haven for opportunistic bacteria. J Hosp Infect, 2001, 49: 87–93.
38. Splieth C, Bernhardt O, Heinrich A, Bernhardt H, and Meyer G. Anaerobic microflora under class I and class II composite and amalgam restorations. Quintessence Int, 2003, 34: 497–503.

39. Stickler DJ. Biomaterials to prevent nosocomial infections: is silver the gold standard? Curr Opin Infect Dis, 2000, 13: 389–393.
40. Su SH, Conroy S, Lin TL, Sheu MS, and Loh IH. Surface coatings, in Wnek GE and Bowlin GL, eds. Encyclopedia of Biomaterials and Biomedical Engineering, Marcel Dekker Inc: New York, 2004.
41. Wang CJ, Huang TW, aWang JW, and aChen HS. The often poor clinical outcome of infected total knee arthroplasty. J Arthroplasty, 2002, 17: 608–614.
42. Wilson M. Bacterial biofilms and human disease. Sci Prog, 2001, 84: 235–254.
43. Xu X and Burgess JO. Compressive strength, fluoride release and recharge of fluoride-releasing materials. Biomaterials, 2003, 24: 2451–2461.
44. Yoshida K, Tanagawa M, and aAtsuta M. Characterization and inhibitory effect of antibacterial dental resin composites incorporating silver-supported materials. J Biomed Mater Res, 1999, 47: 516–522.
45. Young AM, Leung D, Ho S, Rafeeka SA, Pratten J, and Spratt DA, Gulabivala K. Chemical, Mechanical and Biological Studies of New Antibacterial Aesthetic Restorative Materials. Medical Polymers, Rapra Technology Ltd: Shawbury, 2003.
46. Young AM, Leung D, Ho S, Sheung Y, Pratten J, and Spratt DA. New Injectable Antibacterial Polymers for Bone and Tooth Repair. Medical Polymers, Rapra Technology Ltd: Shawbury, 2004.

Chapter 10
Biocompatibility Testing

Kirsten Peters, Ronald E. Unger, and C. James Kirkpatrick

10.1 Introduction

The insertion of a foreign material into the body induces a cascade of events, basically at the interface between the implanted material and the tissue, which results in the recognition of the material as foreign matter. The degree of this physiological response depends on the location of implantation and the composition and mechanical properties of the material. Thus, the body's response to an implanted material is affected by a number of different factors. To evaluate and reduce the risk for unexpected or unwanted side effects, biocompatibility testing is used to examine new biomaterials and biomedical devices destined for implantation. The biological evaluation of the material's safety is a complex task since it requires knowledge in the disciplines of medicine, biology, pathology, engineering and materials science.

The word biocompatibility can be defined as the compatibility between a material and the biological system. Probably the best definition of biocompatibility is that agreed upon at the Consensus Conference on Definitions in Biomaterials in 1986 (European Society for Biomaterials), namely "the ability of a material to perform with an appropriate host response in a specific application" [1]. The corollary of this is that the type of testing method will depend on the intended function of the biomaterial being used. It then becomes obvious that the testing strategy must take into account the biological situation in which the biomaterial will find itself. The biological situation is influenced by the type of affected tissue or organ, physico-chemical features of the material, and the implant duration period.

The biocompatibility concept includes two principal elements. First there is the absence of a cytotoxic effect and second, there is the aspect of functionality. Cytotoxicity deals mainly with the survival of cells and the maintenance of specific cellular functions under the influence of a material and/or its degradation products. Functionality includes the integration of the biomaterial into a biological system (the tissue or organ). Functionality also assumes the absence of impairment of cellular function and requires that the mechanical, chemical and physical features of the material are sufficient for the performance of cell-specific functions. Since biomaterial-induced modifications of cellular functions might also accumulate or act synergistically, cell-material-interactions could lead to deleterious effects with respect to implant acceptance and long-term stability.

The protocols for the evaluation of the safety of biomaterials planned for use as a medical device are guided by specific organizations. In the USA, the Food and Drug Administration (FDA), which is a government body, is responsible for the determination of standards in testing biocompatibility of medical devices. In Europe, the International Organization for Standardization (ISO), a commercial organization, is the authorized body. The definitions for medical devices are similar in Europe and the USA. In Europe the testing of medical devices is guided by an ISO standard (ISO 10993). The US standards (Quality System Regulation/QRS) are mostly comparable to the ISO standard. This chapter refers generally to the ISO standard when accredited test systems are mentioned. Currently, 20 parts of the ISO 10993 standard are either accepted or under preparation. Tests that may be used in the assessment of medical device biocompatibility include procedures for cytotoxicity, sensitization, irritation, acute systemic toxicity, subchronic toxicity, mutagenicity, genotoxicity, hemocompatibility etc. The testing procedures can be performed in vitro as well as in vivo (mainly in animal studies).

The main aspects of biocompatibility testing are discussed in the following sections. It must be pointed out that due to the widely facetted aspects that require to be taken into account one cannot deal with the topic exhaustively. Since the preparation of the samples is an important point in biocompatibility testing we will start this chapter with a short overview of the issue of sample preparation.

10.2 Sample Preparation

The preparation and processing of the test specimens is a fundamental aspect in biocompatibility testing. Both ISO and FDA standards address sample preparation and reference materials (e.g. [2]). The types of test samples, suitable extraction medium and conditions, and appropriate reference materials to be used as controls are covered. The standards recommend testing medical devices in their final product form and composition whenever possible. If finished devices are not available for testing, the evaluation of representative subcomponents of the device is also acceptable. If the device is too large or cannot be tested as a whole, each different material component that has the potential of coming into contact with body tissues should be represented in the same proportion in the test sample as it is in the final product.

Since biomaterials are normally insoluble and the methods for biocompatibility testing are often designed for soluble compounds, the biocompatibility tests can be modified for the evaluation of liquid extracts. The selection of the appropriate extraction medium depends on the test system of choice. For example, bacterial test systems are frequently exposed to physiological osmolarity solutions (0.9% sodium chloride) or either ethanol or dimethyl sulfoxide extracts. Since in vitro mammalian cell culture-based test systems require media that induce cell growth, cell culture medium is often employed as an extraction medium. In vivo test models frequently employ aqueous and nonaqueous extraction fluids that are capable of extracting both water-soluble and lipid-soluble chemicals.

However, one has to keep in mind that the problems often lie in the details since the materials can undergo a wide range of annealing, passivation, electropolishing, and cleaning processes which may lead to extensive changes of their character. Thus, although the underlying material can clearly be identified, the exact surface chemistry, surface corrosion, and durability of the device over a prolonged time period can differ significantly depending on the specific processing and finishing steps.

10.3 Mammalian Cell Culture

A basic requirement in biocompatibility testing is the reduction and replacement of animal studies wherever possible. Biocompatibility testing in mammalian cell cultures in vitro offers an excellent screening method. Due to detailed cell isolation protocols, the availability of different recombinant growth factors and highly defined cell culture media a large number of different mammalian cell types can be more or less easily propagated in vitro (Fig. 10.1). These cell types originate from different tissues and are characterized by unique functions, in accordance with the functions they exhibit within the original organ, some of which are retained during cultivation in vitro. These cell type-specific functions are e.g. pro-inflammatory cytokine production in the case of activation (inflammation in vitro), in vitro capillary formation by endothelial cells after stimulation with blood vessel-inducing growth factors (blood vessel formation in vitro), mineralization after long-term cultivation of osteoblasts (bone formation in vitro) or the production of large amounts of extracellular matrix by fibroblasts in vitro.

There are two large subgroups of cells that can be cultured and these can be grossly divided into the so-called primary cells and permanent cell lines. Primary cells are directly isolated from fresh mammalian tissues. A large number of cell types is able to adapt to cell culture conditions and may be subcultivated for a few passages (cell strains). The primary cells and their resulting cell strains normally have high similarity to the respective cell type in vivo. The preparation of primary cells as isolated single cell type cultures is labor intensive. Primary cells are characterized by an individual heterogeneity due to the genetic heterogeneity of the donor. This heterogeneity may have effects on the reproducibility of results. Primary cells and the resultant cell strains can be maintained in vitro only for a limited period of time. Depending on the cell type, primary cells become senescent after a limited number of cell divisions. This senescence is observed as a reduction in proliferation rate and the loss of cell-specific functions [3].

Permanent cell lines evolve from primary cells after genetic transformation. This transformation can occur in vivo (i.e. malignant transformation in tumor cells) or in vitro. Genetic transformations develop spontaneously or by the influence of chemical or physical mutagens (e.g. Co-60 gamma rays, nickel chloride, 4-nitroquinoline 1-oxide). The development of genetically transformed cells is a multi-step process involving accumulation of multiple genetic derangements over an extended time-period. One step found in nearly all permanent cell lines is the acquisition of an
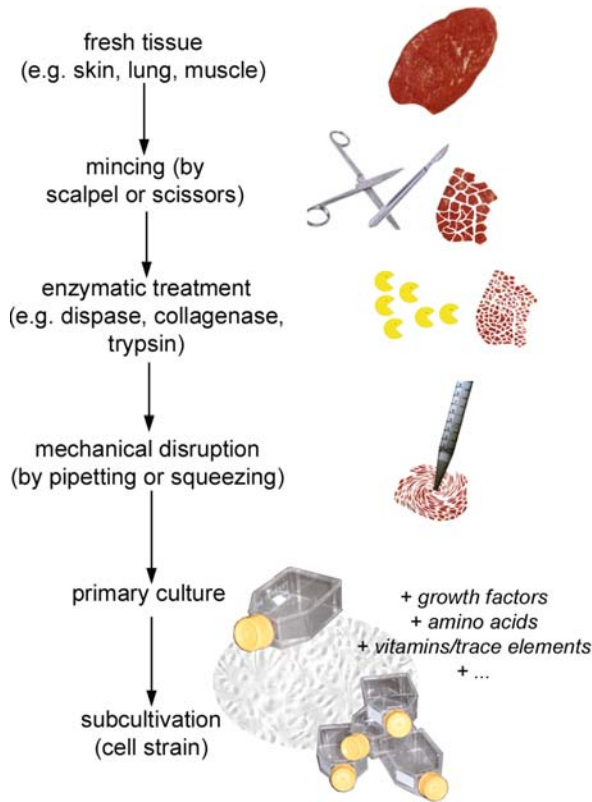
Fig. 10.1 Exemplary isolation protocol for cells from fresh tissues. First, tissues must be mechanically minced into small pieces. This is generally followed by an enzymatic digestion. The type of enzyme utilized (e.g. dispase, collagenase, etc.) depends on various characteristics of the tissue. For example, when the tissue has a significant amount of collagen fibers, the more appropriate enzyme for the disintegration of the tissue might be a collagenase; when the tissue is characterized by large amounts of elastin fibers, the appropriate enzyme is most likely an elastase. Therefore, the isolation methods for different cell types show large variations. The enzymatic digestion is again followed by a mechanical treatment to disconnect the loosened cells from the disintegrated tissues. The isolated cells are placed into cell culture flasks that are, dependent on the cell type, coated with specific adhesion factors (e.g. collagen, fibronectin) to initiate a rapid attachment of the cells. The cell culture media are supplemented with growth factors, amino acids, carbohydrates, vitamins etc. These cell cultures deriving from fresh tissues are referred to as primary cultures. After growing to confluence (i.e. all available surface area is covered by cells) the cells can be subcultivated. This subcultivation leads to cell propagation that is accompanied by cell proliferation. The subcultivation step marks the transition from the primary culture to cell strain

unlimited proliferative potential that requires overcoming the senescence barrier that is typical for primary cells. Permanent cell lines are clonal since they arise from one single cell, although genomic alterations also take place subsequently in vitro to give a certain degree of population heterogeneity.
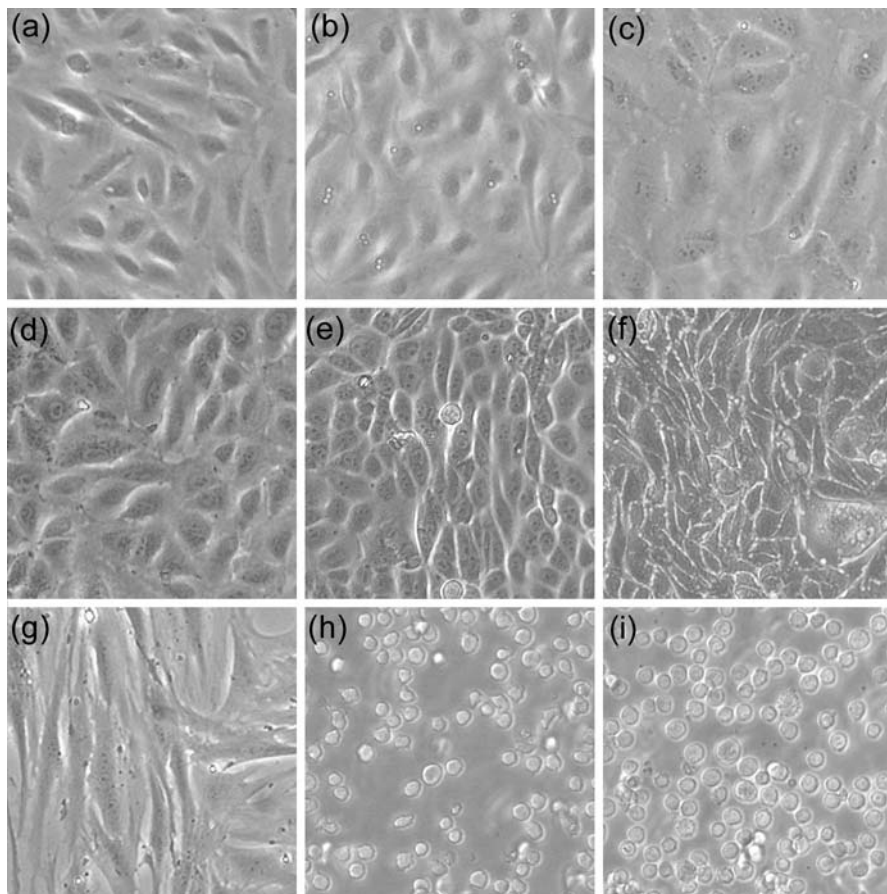
Fig. 10.2 Different primary and permanent cell types in vitro. (**a**) HUVEC (primary culture, passage 0), (**b**) HDMEC (primary culture/cell strain, passage 2), (**c**) HPMEC-ST1 (lung microvascular endothelial cell line), (**d**) A549 (alveolar epithelial cell line), (**e**) 16HBE140- (human bronchial epithelial cell line), (**f**) Calu-3 (airway epithelial cell line), (**g**) human dermal fibroblasts (primary culture/cell strain, passage 15), (**h**) U937 (cell line of a human histiocytic lymphoma, non-stimulated), (**i**) U937 (phorbolester-stimulated) (magnification 10x)

Figure 10.2 shows a number of different human cell types in vitro. These cells differ in their origin (i.e. endothelial, epithelial, fibroblastic, monocytic) and their transformation status (i.e. primary or permanent type). Whereas the different endothelial cells, epithelial cells and fibroblasts require adherence for growth, the monocytic cell line normally grows in suspension. All adherent growing cells are shown in the confluent state. This is the stage reached when the cells have covered the entire cultivation-area available. The same cell type derived from different tissues/organs of the same species shows different phenotypes (Fig. 10.2a: primary human umbilical vein endothelial cells, HUVEC, b: primary human dermal

microvascular endothelial cells, HDMEC, c: a permanent cell line of human pulmonary microvascular endothelial cells, HPMEC-ST1). Most endothelial cells in culture develop a so-called cobblestone morphology. This morphology is variable and is dependent on the origin of the cells (type of tissue) and the transformation status. Such phenotypic differences are also observed in different permanent epithelial cell lines (Fig. 10.2d: A549, e: 16HBE140-, f: Calu-3). Phenotypic differences refer to differences in the expression status of the cells which are induced by varying gene activation.

In comparison to both endothelial and epithelial cells, human fibroblast cultures (Fig. 10.2g) show large differences in phenotype in vitro. The intercellular contacts of fibroblasts are not as pronounced and the growth pattern of fibroblasts is not as regular as for the other two cell types. The difference between the appearance of fibroblasts and endothelial/epithelial cells is due to their arrangement in the tissue: endothelial and epithelial cells are responsible for barrier formation (e.g. between the tissue and the blood or between the air and the tissue) and thus show a polarized growth pattern as monolayers where they have an apical and a basal surface. Fibroblasts on the other hand are three-dimensionally embedded within tissues without an apical/basal polarization.

Most cells derived from solid tissues are dependent on substrate adhesion. However, a few cell lines are able to replicate while floating in the culture medium (suspension culture). Anchorage-independent growth is usually an indication of genetic transformation. This of growth often reflects the origin of the cell type from which they are derived, e.g. from leukemia or lymphomas (e.g. U937 cells derived from a human histiocytic lymphoma, Fig. 10.2h). Interestingly, these cells are also able to attach to the cultivation surface after a specific stimulation (Fig. 10.2i, phorbolester-activated U937, most cells attach to the cultivation surface). This attachment after stimulation is mediated by the expression of different cell surface adhesion molecules [4]. However, the spreading of the cell body onto the surface does not occur. On the other hand, many types of adhesion-dependent cells undergo cell death (i.e. the so-called programmed cell death or apoptosis) if they are prevented from attaching to an appropriate substrate.

Due to the almost limitless expansion potential of permanent cell lines these are often preferred for testing cytotoxicity instead of primary cells. However, one has to keep in mind that permanent cell lines have often lost a significant part of their cell type-specific functions and may be less sensitive to modulating effects of biomaterials than primary cell cultures. Also, the examination of cell cycle-modulating aspects by biomaterials might be questionable in permanent cell lines, since these cell lines have often lost important cell cycle-regulating abilities. In addition to the choice between permanent cell lines or primary cells there is the choice of species. Various comparative studies demonstrate that human cells react in a very different manner compared to cells from other mammals. Since biomaterials are intended for clinical use, in the authors' laboratory only cells of human origin are utilized.

In addition to the fact that in vitro testing leads to the reduction of animal experimentation cell culture methodologies are generally more economical. However, it must not be forgotten that cell culture represents a reduction of the complexity of

the entire organism generally to a single cell type. Thus, the buffering capacity of complex cellular and humoral systems in the intact organism is missing, so that a biomaterial may not perform well in the in vitro test, but be biocompatible in vivo. Nevertheless, this is a risk which is inherent to the system.

The work with mammalian cells in culture requires specific laboratory equipment and technical experience. For example, a cell culture laboratory should be segregated from often utilized rooms and a laminar air flow cabinet, an inverted microscope and a carbon dioxide-incubator are part of the basic requirements (Fig. 10.3). The primary reason for the safety precautions for cell cultivation is to avoid any hazard to individuals and the environment stemming from the cell cultures (this is especially true for human and primate cell cultures). In addition, an undesired contamination of the cell cultures by bacteria, fungi, viruses, and mycoplasma needs to be prevented [5]. Furthermore, experience in cell cultivation is needed since different cell types possess unique demands and can exhibit large variations in phenotype induced by the cultivation methods and handling. Thus, a basic knowledge in cell
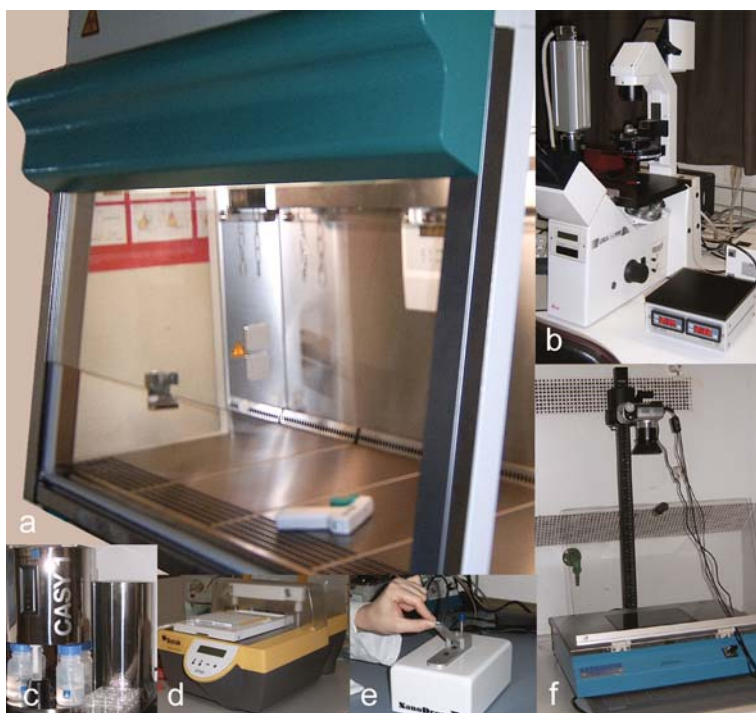
Fig. 10.3 Basic equipment for cell culture-based experiments. (**a**) Laminar air flow cabinet (Heraeus, Germany), (**b**) inverted microscope (Leica Microsystems, Germany), (**c**) cell counter/analyzer (Schärfe System GmbH, Germany), (**d**) multiwell plate-washer (Amersham, Germany), (**e**) spectrophotometer for nucleic acid quantification (NanoDrop Technologies, USA), (**f**) UV-source for nucleic acid analysis (Bachofer, Germany)

biology is considered a prerequisite for working with cell cultures in biocompatibility testing and is essential for ensuring reproducibility of results.

The utilization of human cells is regulated by the specific organizations, depending on the national laws. A number of key guidelines need to be followed: e.g. the donor must be informed and has to sign a consent form whenever a sample is extracted (donors should understand what the sample is used for and how the results might impact scientific problems). Furthermore, the extraction of human tissue/samples must exclude any possibility of harm for the donor. All human tissues/samples extracted for research studies must be approved by an ethics committee. In most cases sample anonymity is adequate. Dependent on the type of extracted tissue or cells there may be formalities which need to be clarified with the local responsible authorities.

10.3.1 Cytotoxicity Testing

As mentioned above, the absence of cytotoxicity (a toxic effect on cellular functions) is a prerequisite for the biocompatibility of a material. For the assessment of risk of a compound to human health the testing of toxicity at the cellular level in vitro has been very useful. The intention is to reduce the complexity of the entire organism generally to a single cell type. Nowadays, the testing of cytotoxic effects of biomaterials or their degradation products in cells in vitro is a major part of the development of new biomaterials.

An approximate procedure for the evaluation of biomaterials in vitro has been developed by ISO (International Organization for Standardization) and is documented in ISO 10993-5 (Biological evaluation of medical devices – Part 5: Tests for cytotoxicity: in vitro methods) [6]. Due to the general applicability of testing cytotoxicity in vitro and the widespread utilization to evaluate large amounts of medical devices and biomaterials this part of ISO 10993 does not consist of a fixed test protocol. ISO 10993-5 guidelines serve as a starting point for the determination of a scheme that requires incremental decisions and leads to the most appropriate cytotoxicity test. Thus, a basic knowledge about the different test systems and of their specific attributes is absolutely necessary.

ISO 10993-5 deals with three testing categories: (1) testing of extracts (conditions for the extraction are specified), (2) testing by direct cell/material contact, (3) testing by indirect cell/material contact. The choice of one or more of these testing categories depends on the character of the tested material, the area of application within the body and the task that it has to fulfill.

The cell type selected for cytotoxicity testing is important. The origin of the cells (human vs. other mammalian cells, tissue type from which the cells are isolated, differentiation status) as well as the transformation status (primary vs. permanent cell lines) is dependent on the application. Whereas the choice of the cell origin (e.g. osteoblasts, fibroblasts, endothelial cells etc.) depends on the future application of the tested biomaterial, the choice of the transformation status of the cells

depends on the availability of fresh tissue (for primary cultures) and suitable and reproducible cell isolation procedures (see above). In ISO 10993-5 a preference is made for permanent cell lines that come from accredited suppliers (e.g. American Type Culture Collection/ATCC, German Collection of Microorganisms and Cell Cultures/DSMZ). However, ISO 10993-5 also recommends the utilization of primary cells/cell strains and organ-typic cultures if a specific sensitivity is needed and if reproducibility and accuracy can be assured. This is an important point since permanent cell lines often suffer from low sensitivity and non-congruent behavior compared to the respective primary cell culture.

An important factor in cytotoxicity testing is choosing between qualitative or quantitative assays (Table 10.1). A qualitative assay evaluates morphological changes microscopically (morphological changes, vacuolization, detachment, etc.). An example for the validation of morphological changes is given in Fig. 10.4. Here,

Table 10.1 Aspects of testing cytotoxicity in vitro

1. Destroyed cellular structures	a) Evaluation of morphological changes	→ Qualitative/descriptive
	b) Measurement of biochemical aspects (e.g. detection of cytoplasmic enzymes in the cell culture supernatant)	Quantifiable (mostly semi-quantitative)
2. Proliferation (e.g. measurement of DNA-synthesis, quantification of specific markers of proliferation)		
3. Metabolic aspects (measurement of specific metabolic pathways, e.g. quantification of mitochondrial dehydrogenase activity by tetrazolium salt assays		
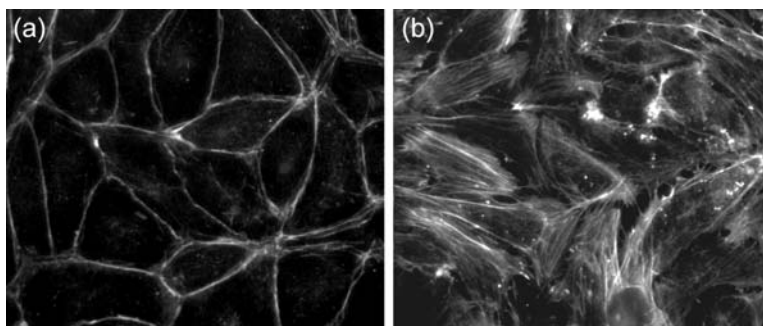
Fig. 10.4 F-actin cytoskeleton in human endothelial cells (**a**) non-treated endothelial cells, (**b**) Co^{2+}-treated endothelial cells; fluorescence microscopy, magnification 40x

the f-actin cytoskeleton of human endothelial cells is stained by the fluorescent-labeled actin-binding molecule phalloidin. Whereas non-treated endothelial cells show a peripheral actin ring (negative control) (Fig. 10.4a), cobalt ion (Co^{2+})-treated endothelial cells exhibit a number of f-actin stress fibers and loosened intercellular contacts (Fig. 10.4b). These cytotoxicity validation methods often provide impressive images and may also indicate a number of possible pathomechanisms [7]. However, the significance of these validation methods depend on the cell type, the interpretation of results requires experienced personnel and the results cannot be exactly or easily quantified.

Quantitative cytotoxicity assays on the other hand are mostly indirect methods for the detection of destroyed cellular structures, impairment of cell proliferation, or impaired metabolic cell functions (Table 10.1). Within these categories, there are different possibilities for measurement and most important is the use of appropriate negative and positive control materials that exhibit the reproducibility of the test system. The negative control should be a material that does not induce a cytotoxic reaction (e.g. cell culture polystyrene). This control is used to show the basal reaction of the cell culture and is necessary since all cells in culture show a variable fraction of dying cells. A positive control must also be used to evaluate the sensitivity of the test system. Therefore, a material or compound should be utilized that shows a reproducible cytotoxic response in the exposed cell culture.

The range of commercially available and accredited cytotoxicity assays is large (Table 10.2). An example for the biochemical quantification of destroyed cellular structures (Table 10.2, item: membrane integrity) is the detection of the cytoplasmic, soluble lactate dehydrogenase (LDH) that is released into the cell culture supernatant when the cell membrane is disrupted. A disruption of the cell membrane generally correlates with an increase in cell death. Therefore, the activity of LDH detected in the cell culture supernatant can be used as a quantitative measurement of cell death in vitro [8]. LDH activity catalyzes the oxidation of lactate to pyruvate in a reversible manner by reducing NAD^+ to $NADH+H^+$.

$$NAD^+ + lactate \xrightarrow{LDH} pyruvate + NADH + H^+ \qquad (10.1)$$
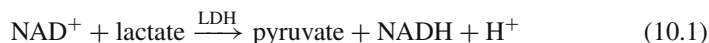
The development of reduction equivalents by LDH is used for the LDH-activity assay: the developed $NADH+H^+$ reduces (by means of the added enzyme diaphorase) the light yellow tetrazolium salt INT (2-[4-iodophenyl]-3-[4-nitrophenyl]-5-phenyl-2H-tetrazolium-chloride), that is added to the test system, to a red formazan salt. This conversion can be measured photometrically at 492 nm.

$$NADH + H^+ + INT\ (yellow) \xrightarrow{diaphorase} NAD^+ + formazan\ (red) \qquad (10.2)$$
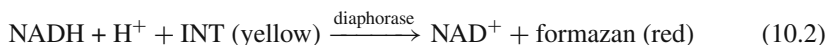
Thus, an increase in the number of dead cells results in an increase of LDH activity in the cell culture supernatant, which is reflected in an increase of absorption at 492 nm after the addition of the tetrazolium salt INT and the enzyme diaphorase.

Table 10.2 Classification of accredited cell viability/cytotoxicity assays (the list is not exhaustive)

Cellular function	Test system
Membrane integrity	– Trypan blue exclusion
	– ^{51}Cr-release
	– Neutral red release
	– LDH-release
	– Calcein-AM-release
Energy metabolism	– Detection of reduction equivalents by tetrazolium salt assays (i.e., MTT, MTS, WST-1, XTT...)
	– ATP-content
Proliferation	– Cell number
	– [^{3}H]-Thymidine uptake
	– BrdU-incorporation
	– Crystal violet staining (DNA-binding dye)
	– Ki67-expression (immunocytochemical detection of a nuclear proliferation marker)
	– Cell cycle analyses (by flow cytometer)
Apoptosis	– Nuclear condensation/fragmentation
	– TUNEL staining (terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling)
	– DNA ladder formation
	– Annexin V-staining
	– Caspase detection
	– Detection of specific pro- and anti-apoptotic molecules (bax/bcl)

Other cytotoxicity assays also utilize the color changes of tetrazolium salts to formazan by reduction. Very prominent representatives of these are the tetrazolium salt assays (e.g. MTT, MTS, WST-1, XTT) that indicate the energy metabolic state of the cells. The tetrazolium salts are membrane-permeable. When the salts are added to the cell culture medium they permeate into the living cells. Here, the light yellow tetrazolium is reduced by the physiologically developing NADH+H$^+$ to (mostly) water soluble, membrane-permeable red or orange formazan products [9, 10] that are released into the cell culture supernatant (since the MTT reagent leads to a water insoluble product a further solubilization step is necessary). As in the LDH release assay this color change can be detected photometrically. In contrast to the conclusion of the LDH assay, an absorbance increase in the tetrazolium salt assays refers to an increase in metabolic activity and thus indirectly in cell viability, a decrease of absorbance is normally an indirect measurement for an impairment of cell viability.

A further important parameter in cytotoxicity testing is the evaluation of the proliferative state, i.e. the capacity of the cells to divide and propagate. In this case a number of different assays is also available (Table 10.2). The most direct way to obtain information about proliferation is to measure the cell number. This can be done very classically using the Neubauer cell counting chamber. More progressive is the method of cell counting using an electronic cell counter that works by the resistance measurement principle. The advantage of the electronic cell counting

systems is that additional information is provided since cell size, volume etc. are also analyzed and the recorded data are saved electronically. However, both methods are relatively time consuming for adherent cells since the cells have initially to be detached from the cultivation surface primarily by an enzymatic digestion.

Crystal violet staining can be used as an indirect method for the quantification of relative cell number. Crystal violet is a DNA-binding dye. The staining of cells by crystal violet can be performed in a microplate format and analyzed spectrophotometrically. These data indicate the relative number so that the absorbance of the relevant samples can be compared to that of the non-treated controls (i.e. negative control) [11]. A further method of proliferation quantification is the Ki67 assay. The Ki67 antigen is only expressed in the nucleus of cells, actively present in the cell cycle. The detection of this antigen can be used to quantify proliferating cells using an enzyme-linked immunoassay based on a peroxidase staining reaction and subsequent dye quantification by a microplate-spectrophotometer [12].

Proliferation can also be quantified by labeling the newly synthesized DNA by [3H]-thymidine or bromodeoxyuridine (BrdU). Both compounds are incorporated into newly synthesized DNA strands of actively proliferating cells during the S-phase of the cell cycle. Whereas the radioactive thymidine-incorporation is measured by scintillation counting of autoradiography, the BrdU-incorporation is detected immunocytochemically. The [3H]-thymidine-incorporation assay has several limitations, primarily in the handling and disposal of radioisotopes and the necessity for specialized equipment.

A comparison of two assays that reflect the proliferative state is shown in Fig. 10.5. Human endothelial cells were exposed to different concentrations of the metal ion salt $CoCl_2$ (Co^{2+} ions are corrosion products of CoCr-alloys). Low concentrations (0.1 mM) induced only slight changes of the relative cell number detected by crystal violet staining, higher concentrations of $CoCl_2$ (0.7 mM) induced a reduction of more than 20%. However, the results of the proliferation-inhibiting effects of $CoCl_2$ are more pronounced, when the proliferative state is examined by the Ki67 assay (0.1 mM: more than 10%; 0.7 mM: nearly 55%). The differences
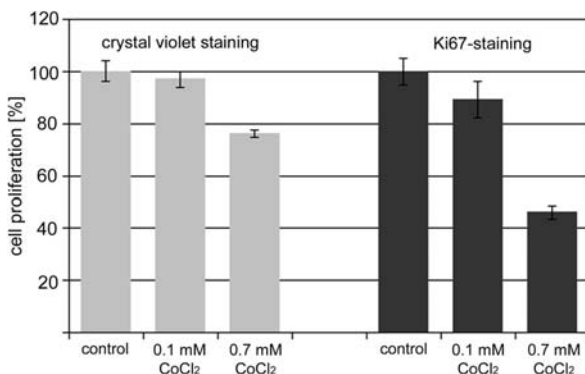
Fig. 10.5 Detection of the proliferative state in human endothelial cells after exposure to metal ions (crystal violet staining vs. Ki67-staining, exposure 24 h, untreated control – negative control – set as 100%)

between these assays occur since different parameters of proliferation are being examined; crystal violet stains the cell nuclei since it binds to DNA and is thus an indirect method for the quantification of the relative cell number, whereas the Ki67 assay detects a protein that is only expressed during the active cell cycle. Thus, the Ki67 assay reflects a parameter that is mechanistically ranged ahead of the cell propagation and might therefore be more sensitive than the crystal violet staining. Since crystal violet staining is much easier to perform than the complex Ki67 assay, the selection of the assay depends on the individual case and has to be carefully considered.

There are different mechanisms of cell death which generally can be divided into necrosis and apoptosis. Whereas necrosis is also called the "accidental" cell death, apoptosis is referred to as "programmed" cell death. This original classification depends on the different morphological, biochemical and molecular hallmarks that are expressed during cellular breakdown.

Necrosis is caused by extreme conditions (e.g. toxins, heat, radiation, trauma, oxygen deficiency etc.). These physical or chemical stimuli can lead to the lethal disruption of cell structure and generally affect a number of cells simultaneously. In necrosis the cells swell, the plasma membrane disintegrates and intracellular compounds are released. The release of intracellular material leads to pro-inflammatory processes surrounding the dying cells.

In contrast to necrosis, apoptotic cells are characterized by condensation of the cytoplasm and the nuclei, so that affected cells shrink. In the course of apoptosis nuclear fragmentation and cleavage of chromosomal DNA into internucleosomal fragments occurs. The degradation of DNA in the nuclei of apoptotic cells occurs by a number of ways following activation of executioner enzymes, the caspases. Characteristic is also the packaging of the deceased cells into apoptotic bodies without plasma membrane breakdown. These apoptotic bodies can be recognized and removed by phagocytic cells. Therefore, apoptosis is also notable for the absence of inflammation around the dying cells. Currently this original classification of necrosis (as the accidental cell death) and apoptosis (as programmed cell death) is extended to a number of alternative cell death styles. There is growing evidence that, besides apoptosis, autophagic and necrotic forms of cell degeneration may be programmed [13].

Apoptosis-typical hallmarks can be analyzed in vitro and a number of different test systems are available (see Table 10.2). The nuclear condensation and fragmentation is easily detected by nuclear staining (e.g. by Hoechst 33342, DAPI). An example of different stages of apoptosis-typical nuclear break-up is shown in Fig. 10.6a. Non-affected nuclei of endothelial cells possess an oval, regular shape (1); during the course of apoptosis the nuclei condense (2) and are fragmented (3); at a late stage of apoptosis the nuclear fragments come apart (4).

A further method in the detection of apoptosis is the analysis of fragments of chromosomal DNA. Characteristic in apoptosis is a specific form of DNA degradation in which the genome is cleaved at internucleosomal sites, generating a "ladder" of discretely sized DNA fragments when analyzed by agarose gel electrophoresis (Fig. 10.6b). The nucleic acid preparations contain RNA in addition to the DNA of
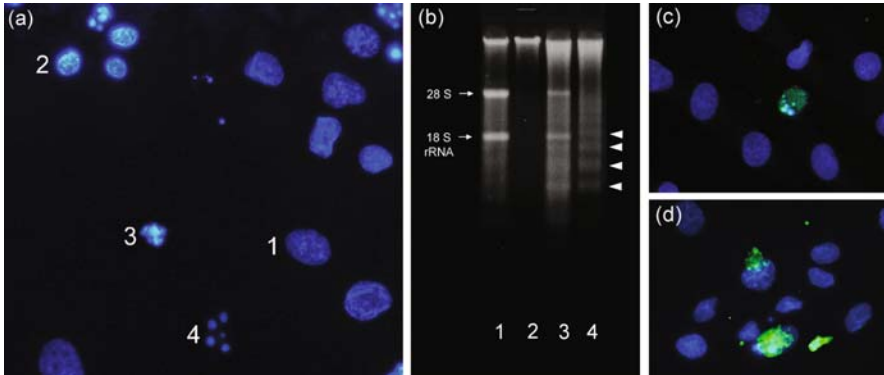
Fig. 10.6 Detection methods of apoptosis. (**a**) Nuclear staining of H_2O_2-treated human endothelial cells (0.7 mM, 24 h, staining with Hoechst 33342; different stages of the nuclear degradation: (1) non-affected nucleus, (2) nuclear condensation, (3) nuclear fragmentation, (4) nuclear breakup); (**b**) DNA ladder formation in human endothelial cells after metal ion-treatment (agarose electrophoretic separated nucleic acid preparations of: lane 1 – non-treated control cells, lane 2 – non-treated control cells after RNase-treatment, lane 3 – Co^{2+}-treated cells, lane 4 – Co^{2+}-treated cells after RNase-treatment; arrowheads mark the distinct "steps of the ladder"); (**c**) TUNEL staining in sporadically occurring apoptosis in human endothelial cells (*green*: TUNEL positive, *blue*: nuclear staining, digital overlay of the TUNEL and the nuclear staining, fluorescence microscopy); (**d**) detection of active caspase-3 (*green*) in cobalt-particle treated endothelial cells (digital overlay of caspase-3 and nuclear staining, fluorescence microscopy)

interest. Since the RNA masks the DNA-ladder (Fig. 10.6b, lane 1 and 3) removal of RNA with a specified enzyme (RNase) is necessary (Fig. 10.6b, lane 2 and 4). After the treatment of endothelial cells with high concentrations of Co^{2+} a DNA-ladder is detectable (Fig. 10.6b, lane 4). This apoptosis-typical fragmentation can also be detected by the so-called TUNEL method (terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling). Therefore, the low molecular weight DNA fragments as well as single strand breaks ("nicks") in high molecular weight DNA are labeled at the apoptosis-typical breaks with modified nucleotides in an enzymatic reaction. Incorporated nucleotides are detected by fluorescence or light microscopy depending on the modifications of the incorporated nucleotides (Fig. 10.6c: fluorescently labeled). The TUNEL-positive endothelial cell nucleus (green) in Fig. 10.6c shows simultaneously fragmented nuclear material.

The core effectors for the cellular breakdown during apoptosis are a family of proteolytic enzymes, the caspases. Caspases exist as latent precursors in most nucleated mammalian cells. The caspase precursors are activated by cleavage or proximity-induced dimerization. One of the executioner caspases is caspase-3, which mediates the apoptotic cascade. The active enzyme can be detected in the apoptotic cells by immunostaining (Fig. 10.6d). Here, the specific binding of an antibody against the active caspase-3 (not the precursor) is detected in cobalt-particle treated endothelial cells by labeling with a fluorescent secondary antibody and a subsequent fluorescence microscopic analysis.

Thus, there are numerous possibilities for testing cytotoxicity in vitro. The significance of these assays often depends on the choice of the assay and the mechanisms that are influenced by the tested compound. Therefore, the results achieved by a special cytotoxicity test have to be evaluated critically. To exclude misinterpretations of the assays the utilization of at least two different tests that are based on different test methods is highly recommended. Thus, if the cytotoxicity assays are utilized correctly these in vitro test systems might be helpful for the pre-selection of compounds and materials and lead to the reduction of animal studies.

10.3.2 Hemocompatibility

Hemocompatibility, the ability of a material to perform in contact with blood, is one of the most complex fields in biocompatibility testing. This is due to the large number of interacting cell types and factors which are addressed during the cascade of activation steps (see below). All materials implanted come into direct contact with tissues and thus into direct contact with blood. ISO 10993 requires an evaluation of hemocompatibility for any medical device that has contact with circulating blood, directly or indirectly, during routine use. In other words, testing of hemocompatibility is relevant for all devices and implants that come into large-area contact with the circulatory system. Thus, in addition to heart valves, vascular prostheses and vascular stents, most forms of extracorporeal devices involve a blood-biomaterial interaction. Guidelines for such evaluations and the respective devices are found in ISO 10993-4 (Biological evaluation of medical devices – Part 4: Selection of tests for interactions with blood) [14].

The human blood consists of cellular components (40–45%) and the plasma (55–60%). Red blood cells (erythrocytes), white blood cells (leukocytes) and the platelets (thrombocytes) constitute the cellular part of the blood (Fig. 10.7a). The blood circulates within the vascular system and fulfils a number of different functions, including immune responses and tissue repair as well as oxygen and nutrient supply and the disposal of metabolites. Because of the broad range and the critical nature of these functions, any source of cytotoxicity to blood cells can cause significant harm. For example, the breakdown of erythrocytes (called hemolysis), which can be induced by a material or a toxin or might result from mechanical damage, directly impairs the ability of the circulatory system to provide the body's tissues with oxygen. Moreover, adverse interactions with leukocytes can impair the body's ability to repel invading pathogens efficiently. Furthermore, blood contains several soluble multi-component protein systems that systematically interact in various ways to perform critical functions. For example, the complement system participates in inflammatory reactions and facilitates removal of invading pathogens. The clotting cascade operates to initiate blood coagulation, thereby preventing excessive loss of fluid and facilitates tissue repair (Fig. 10.7b). The impairment of and other cascade systems (e.g. kinin system, fibrinolytic system) can have significant adverse effects on the body.
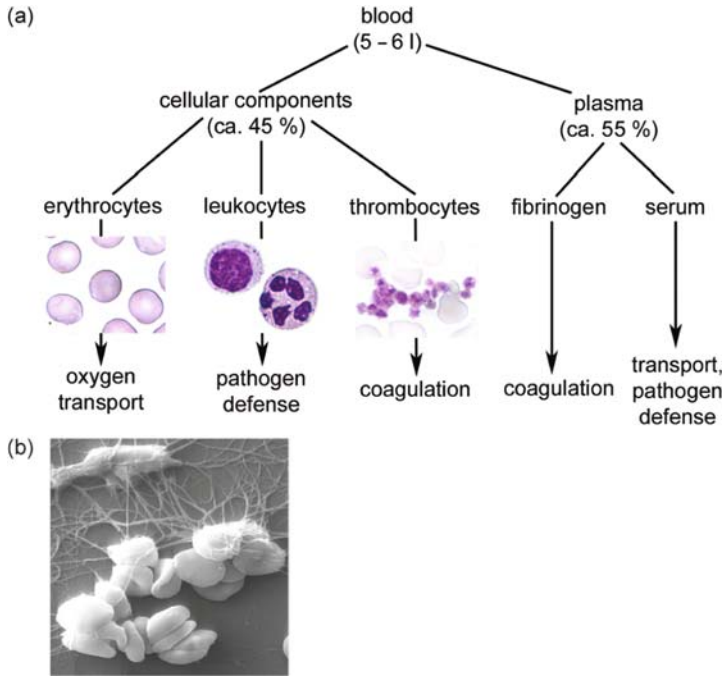
Fig. 10.7 Human blood. (**a**) Composition and function of blood components in humans (staining of blood cells: panoptical staining according to Pappenheim), (**b**) blood coagulation, human blood cells entrapped in a fibrin network (scanning electron microscopy)

The above mentioned clotting cascade consists of two sub-pathways that lead to the formation of a fibrin clot: the intrinsic and the extrinsic system. These systems are initiated by distinct mechanisms but work together for a common result, the blood clot. In Fig. 10.8 a highly simplified depiction of the cascade is shown. Only the major components are shown; a large number of interacting factors is not included. Beside the process of blood clotting, the subsequent dissolution of the clot following repair of the injured tissue is also an important aspect (for more information see: http://www.labtestsonline.org/images/coag_cascade.pdf).

There are three different approaches for testing hemocompatibility: (a) in vivo (in animals), (b) ex vivo (test systems that shunt blood directly from a human subject or test animal into a test chamber that is located outside the body), and (c) in vitro (test systems with fresh, usually anticoagulant-treated blood samples). The types of tests required by ISO 10993-4 depend on the blood contact category of the device or material. In vivo experiments cannot focus on a single parameter of hemocompatibility as the different blood cells/components and their versatile functions act together. These complex interactions can not be significantly mimicked in vitro. On the other hand, the in vitro experiments allow a detailed examination of single factors involved in hemocompatibility.
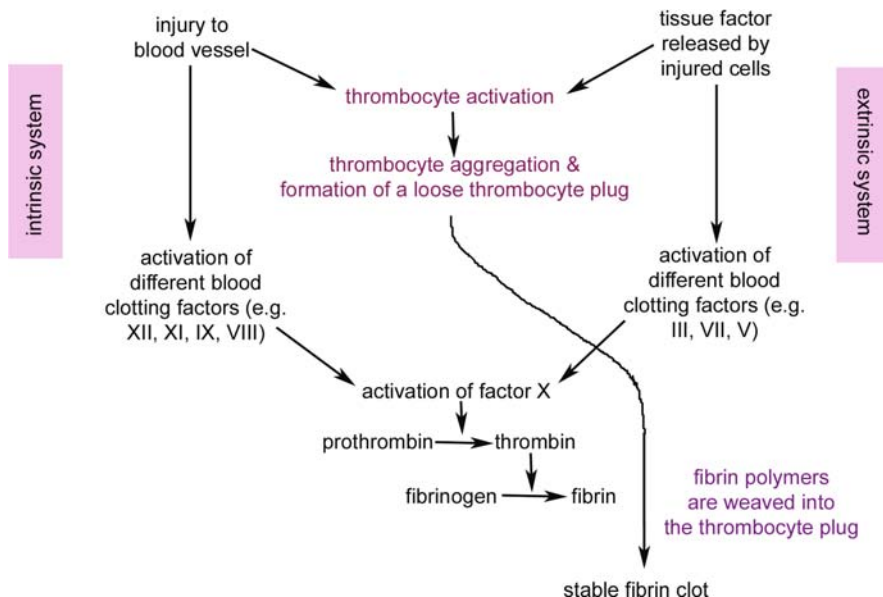
Fig. 10.8 Diagram of the clotting cascade: the intrinsic system is initiated when contact is made between blood and the cell-denuded vessel wall. The extrinsic system is initiated by the release of the so-called tissue factor by injured cells. The two pathways converge at the activation of factor X. The active factor X (called Xa) activates prothrombin to thrombin. Thrombin ultimately converts fibrinogen to fibrin and activates a fibrin-crosslinking factor (factor XIII). The developing fibrin polymers are woven into the thrombocyte aggregations forming a stable fibrin clot

The methods used for testing hemocompatibility are classified into different categories. These categories are (1) thrombosis, (2) coagulation, (3) platelets, (4) hematology and (5) immunology. This classification is based on the different aspects and detection methods for testing hemocompatibility.

(1) Thrombosis: Thrombosis is an in vivo phenomenon resulting in the partial or complete occlusion of the vessel or device following blood coagulation. A detached and thus moving thrombus is called an embolus. When an embolus becomes lodged in a blood vessel it is called embolism. Depending on the affected tissue an embolism might induce a severe life-threatening situation, such as in the lungs or brain.

In most applications non-thrombogenic behavior of the materials is required. The characterization of the thrombogenicity of a material includes the microscopic analysis (light, fluorescence or electron microscopic) of adherent platelets, leukocytes, aggregates, erythrocytes, fibrin, etc. by in vitro studies. The adsorption of proteins to the material surface and the interactions between different proteins and the material surface has a crucial effect on the thrombogenicity of the material. Materials that do not induce adhesion of thrombocytes by protein adsorption of plasma proteins or by interaction of blood compounds on contacting the material are referred to as non-thrombogenic.

(2) Coagulation: A number of different parameters are involved in the induction of coagulation in vivo. A simple method to examine hemocompatibility of a material is the blood coagulation test: Blood is brought into contact with the surface and the coagulation time is determined. The time frame for coagulation determines if a material activates or impairs coagulation. Also, different parameters of the coagulation cascade can be detected in vitro and can be quantified. Among these are the detection of TAT (thrombin-antithrombin III-complex), and some key enzymes of the coagulation cascade such as Xa (activated factor X), XIIa (activated factor XII) utilized for evaluation of the coagulative capacity of a material. The so-called "partial thromboplastin time" (PTT) is used to assess the extent of activation of the intrinsic coagulation pathway. Therefore quickly thawed, pooled, citrated, fresh human plasma is used and the material is brought into contact with this plasma (37°C). The plasma is immediately removed from the material samples and placed on ice. Afterwards the plasma is incubated with $CaCl_2$ and cephalin (phospholipids derived from rabbit brain). The re-calcification together with the addition of cephalin induces the clotting. The plasma samples are monitored until clot formation is seen. The time that is needed for clot formation is recorded and gives the PTT.

(3) Platelets: The activation of platelets can be confirmed by the determination of the number of material-adherent platelets and their degree of spreading. Furthermore, the surface expression or secretion of the platelet activation markers CD62, CD63 and platelet factor 4 (PF4) can be useful for the evaluation of the activation status.

(4) Hematology: Testing hemolysis (i.e. measurement of erythrocyte fragility) in contact with materials or their extraction products is a significant screening method used in hematology. The hemolysis tests are performed using anticoagulated blood that is brought into direct contact with the materials or co-incubated with sample extracts. After incubation at 37°C, the blood cells are removed by centrifugation and the supernatant is analyzed spectroscopically at 540 nm for the release of hemoglobin. These data are compared to the corresponding positive (e.g. deionized water that will cause bursting of all erythrocytes) and negative controls (e.g. phosphate- buffered saline solution). The higher the absorption at 540 nm, the higher the hemolytic properties of a material or its extraction products. A further aspect in the hematology category is the evaluation of leukocyte total number. Leukocyte counting is carried out on samples of whole blood after in vitro contact with the material.

(5) Immunology: The complement system is part of the immune response in the blood. It consists of a series of approximately 25 proteins that work to "complement" the activity of antibodies in destroying bacteria. Therefore, the unintended activation of this system by a material could induce a chronic triggering of complement. Complement proteins circulate in the blood in an inactive form. When the first complement member is triggered, a cascade of activation steps is initiated. Generally the activation of the complement system is possible by two different pathways: (1) the so-called classical pathway that is triggered by bacteria to which antibodies are bound and (2) the alternative pathway, triggered by bacterial components, other microorganisms and material surfaces without the assistance of an antibody-binding (Fig. 10.9).
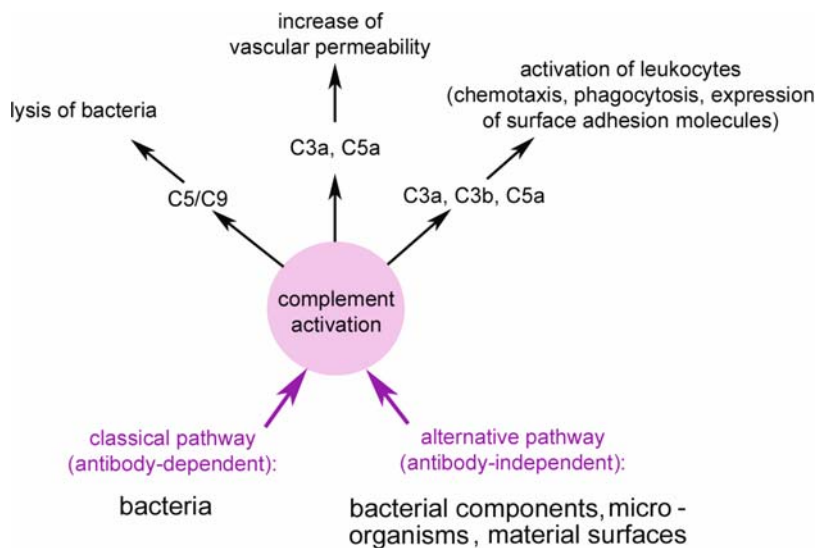
Fig. 10.9 A schematic drawing of the complement system (highly simplified): the classical pathway is induced by bacteria to which antibodies have bound. The alternative pathway is activated by bacterial components, microorganisms and material surfaces without antibody-binding

A panel of complement proteins is available that can be utilized for the detection of complement activation in hemocompatibility testing. These different proteins encompass the various complement activation pathways. Among these are the proteins C3a, C5a, TCC (terminal complement complex), C4d, and SC5b-9. Elevated levels of these complement components indicate the activation of the complement system. Usually these factors are detected and quantified by enzyme immunoassays.

However, although a number of different test systems are available the evaluation of hemocompatibility is unique in that the design of the final device can determine its performance. Thus one needs to consider specific surface properties, stiffness, shape etc. that may cause deleterious effects to the blood elements in addition to the accepted hemocompatibility tests.

10.3.3 Hypersensitivity/Allergic Responses

Allergies are the most common disorder of immunity, affecting more than 25% of the Western World population. The term allergy is used to describe the body's response to a substance which is harmless in itself but results in an immune response and a reaction that causes symptoms and disease in a predisposed person. The symptoms of an allergy may include everything from a runny nose, itchy eyes, and skin rash to life-threatening complications such as asthmatic attacks and anaphylactic shock.

Allergic reactions are usually classified by the type of tissue damage that they cause. Some allergic reactions produce more than one type of tissue damage, and other reactions involve antigen-specific lymphocytes (a subclass of the white blood cells/leukocytes) rather than antibodies. There are four recognized types of allergic reactions based on different subclasses of antibodies (or immunoglobulins, Ig) and cell types that interact with each other (see also Table 10.3): (1) Type I: Anaphylaxis (immediate-type hypersensitivity) – anaphylaxis occurs when a specific allergen combines and cross-links antibodies (i.e. the Ig-subclass IgE) affixed to so-called mast cells; anaphylactic disturbances occur only if the body had previous contact to the allergen. Shortly after re-exposure to the allergen the reaction occurs (within minutes, e.g. hay fever), (2) Type II: Cytotoxic Reactions – These are antigen-antibody reactions usually mediated by IgG and IgM antibodies at cell surfaces that result in the lysis of blood cells (red cells, white cells, and platelets) due to the activation of complement (well-known incidences in this category are blood transfusion reactions), (3) Type III: Immune complex reactions – These are antigen-antibody reactions mediated by IgG and IgM. The reactions produce toxic antigen-immunoglobulin complexes that can circulate in the blood. In these cases, the complexes cause damage by adhering to blood vessel walls and initiating an inflammatory response (such as vasculitis), (4) Type IV: Delayed-type hypersensitivity or cell-mediated immunity reactions – These are reactions between antigens and sensitized antigen-specific T lymphocytes (subclass of leukocytes), but not antibodies. The reaction subsequently releases inflammatory and toxic substances, and lymphokines that attract other leukocytes. The induction of hypersensitivity is called sensitization. Clinical examples include transplant rejection, and contact dermatitis

Table 10.3 Types of allergic response

Type	Name	Initiation time	Mechanism	Clinical examples
I	Anaphylaxis	2–30 min	Antigen induces cross-linking of IgE bound to mast cells	Anaphylactic shock, hay fever, eczema
II	Cytotoxic reactions	5–8 h	Antibody directed against cell surface antigens mediates cell destruction	Blood transfusion reactions, autoimmune haemolytic anaemia
III	Immune complex reactions	2–8 h	Antigen-antibody-complexes induce release of inflammatory factors by various cells	Disseminated rash, arthritis, glomerulonephritis
IV	Delayed type hypersensitivity	1–3 days (3–4 weeks for granuloma)	Antigen-specific T lymphocytes activate leukocytes by release of inflammatory factors	Contact dermatitis, tuberculin reaction, granuloma

in response to metals such as chromium, cobalt and nickel. Type I and Type IV reactions are the most common.

Most allergens are proteins. However, non-protein "allergens" (in quotes since this term is usually reserved for the inducers of IgE reactions, therefore better: sensitizer) also exist. These non-protein sensitizers include nickel, chromium, cobalt, gold, palladium ions, and other metals and metal-containing substances [15]. Due to their small size these metals referred to as "haptens", in that they must bind to a protein in order to become allergenic [16]. Furthermore, acrylic compounds [17], epoxy moieties [18], and Teflon implants, and a few mineral substances are shown to induce hypersensitivity [19]. Most of these non-protein compounds do not cause Ig reactions. Thus, most allergic reactions induced by implanted artificial materials are Type IV reactions, which belong to the cell-mediated immunity reactions and do not involve Ig.

In the early 1970's it was observed that metallic materials used in bone surgery could cause the development or an increase in contact dermatitis [20, 21]. From this it was concluded that contact hypersensitivity to metals might explain some orthopedic complications. This was supported by the fact that in patients developing such complications after implantation of metallic devices, a positive contact dermatitis (detected by the so-called patch test) with metals was more frequent than in other patients [22]. Hypersensitivity reactions usually occur as a result of repeated or prolonged contact with sensitizing compounds (such as with an implant). Biomaterials and devices that cause sensitization reactions do so by means of the extractable compounds.

To exclude a hypersensitivity-inducing potential of materials different test systems have been developed. Methods for performing such tests are described in ISO 10993-10 (Biological evaluation of medical devices – Part 10: Tests for irritation and delayed-type hypersensitivity) [23]. Due to the complexity of reactions in sensitization appropriate in vitro methods are not available. However, there are different methods for testing the hypersensitivity/sensitization-potential of materials in animals. Most of those reactions to biomaterials are Type IV reactions (cell mediated immunity) and are tested on the skin of laboratory animals. Dermal hypersensitivity reactions in laboratory animals are marked by redness and swelling. Biomaterials and other device materials are tested for the presence of sensitizing chemicals using guinea pigs, a species known to be nearly as responsive to dermal sensitizers as humans.

The repeated-patch test method (also called Buehler-test) involves exposing the skin of guinea pigs directly to the test material under occlusive dressings for a minimum of 6 h [24]. This procedure is repeated three times a week for three weeks (induction phase). Following a two-week rest to allow for the development of a delayed response (recovery phase), the animals are exposed to a patch of the biomaterial (final exposure). The repeated-patch model is used primarily for topical devices such as dermal electrodes, since this method of applying test materials to the animals simulates clinical use.

A further test system that is based on guinea pigs is the maximization test method (called guinea pig maximization test/GPMT, described by Magnusson and Kligman)

[25]. In this case extracts of the test material are prepared in saline and vegetable oil and guinea pigs are exposed repeatedly to the extracts. The guinea pigs are first injected with an extract along with an adjuvant intended to enhance an immune response, and then receive a topical application. Following a two-week recovery phase, the animals are covered with a topical patch containing the extract. Generally considered more sensitive than the repeated-patch model, the maximization test is used for device materials that will contact areas other than skin. The use of both a saline extract and an oil extract simulates extraction by body fluids and by intravenous liquids and other pharmaceutical products that first contact the device and then the patient. Both guinea pig sensitization tests require nearly two months and thus take the longest time to complete of all the biocompatibility tests described in the 10993 standards. In both techniques, the area of the challenge patches is examined for reactions (redness and swelling) that are not present in negative-control animals. In addition, known sensitizing chemicals are used periodically to validate the methods.

Human allergens generally sensitize guinea pigs although this animal is known to be less sensitive than the human. However, some substances known to be human sensitizers have failed to sensitize guinea pigs. These include some metals, drugs, latex-based products and macromolecular proteins such as collagen [26]. Thus, these methods are not perfect in their ability to detect weak sensitizers or chemicals, and they do not detect chemicals that act as adjuvant, enhancing an immune response to other chemicals to which a patient might be exposed. To overcome these problems additional test methods have been developed. One of these tests is the local lymph node assay (LLNA) in mice [27]. The LLNA has been validated for pharmaceutical and agricultural chemicals. In contrast to the classical sensitization tests no complete sensitization has to be caused in the animal. In the LLNA, the ears of mice are treated with the extract/chemical and then the surrounding lymph nodes are examined for a lymphocyte proliferative response as demonstrated by the accumulation of radio-labeled DNA constituents (uptake of the labeled constituent if there is a lymphocyte activity induced by the tested compound). This testing method is also mentioned in the Annex of ISO 10993-10 [23]. However, one should keep in mind that no single test method will adequately identify all substances with a potential for sensitizing human skin and which is relevant for all substances. Much effort also has been devoted to developing in vitro alternatives to the guinea pig sensitization models, but thus far no suitable replacements have been identified.

10.3.4 Genotoxicity

A further striking issue of biocompatibility testing is the detection of genotoxicity. Genotoxicity is defined as any toxic modification of the structure or function of the genetic material, the DNA. DNA contains the instructions for all cellular activities. These instructions are encoded within sections of DNA devoted

to each specific function, the genes. Specific units consisting of many genes are called chromosomes. The number of chromosomes is defined for each species (e.g. 46 chromosomes in humans, 78 in dogs). An alteration in any part of the DNA that results in permanent changes in cell function is called a mutation. The agents that cause such mutations are known as genotoxins. The rationale for the testing of a biomaterial's genotoxic potential is the concern for the ability of a biomaterial to induce cell death based on genetic mutation or to initiate a malignant transformation (i.e. cancer). DNA modifications can have permanent inheritable effects on genes and on chromosomal structure or number (chromosomal aberration). DNA single or double strand breaks, base modifications or deletions, intra- or interstrand DNA-DNA or DNA-protein cross-links constitute the major lesions which can lead to cell death or result in mutational events which in turn may initiate cancer [28].

There are different levels of genetic variations that can occur as specified above. These different levels can be detected in vitro and in vivo (Table 10.4). Again, no single assay is capable of detecting all types of DNA damage; therefore, the utilization of different test systems is generally recommended. In vitro tests in each category can be conducted using microorganisms or mammalian cells and in vivo tests systems (mostly with rodents) are also available which are usually performed subsequently to the in vitro assays. In the following a number of established tests for genotoxicity detection are described.

A gene mutation test can be used for the detection of so-called point mutations. Point mutations are characterized by minor alterations of the DNA molecule (e.g. substitution, addition or deletion of base pairs) and these mutations can be detected by the so-called reverse mutation test (previously known as the Ames test) [29, 30]. This assay is commonly employed as an initial screening for genotoxic activity and, in particular, for point mutation-inducing activity. The principle of the test system is based on genetically-modified bacterial cells that are deficient in histidine biosynthesis. These bacterial cells do not have the ability to grow without the supplementation of histidine. If these bacteria are exposed to the test substance and regain the possibility to grow without the addition of histidine, this indicates that a material-/compound-induced genetic transformation most likely occurred.

A test method that can be utilized in vitro as well as ex vivo is the so-called unscheduled DNA-synthesis test (UDS). UDS is the term used to describe the replication of DNA during the repair of DNA damage and as such is distinct from the DNA replication which is confined to cell proliferation. The test is based on the incorporation of radioactive labeled thymidine into the DNA of non-dividing mammalian cells [31]. Thus, the higher the amount of incorporated radiolabeled thymidine, the higher the DNA repair activity and consequently the higher the genotoxicity of the tested compound. A further assay that detects DNA strand breaks is the Comet assay (also known as single cell gel electrophoresis). In case of a severe damage of DNA an increase of DNA fragments occur. These DNA fragments can be detected and visualized. The name Comet assay is based on the appearance of a characteristic streak similar to the tail of a comet. This assay can be utilized in vitro as well as in vivo [32].

Table 10.4 In vitro and in/ex vivo test systems in genotoxicity detection approximately assigned to the induced defect and the class of test method

Defect	Point mutations	Primary DNA damage			Chromosomal aberrations		
Test system	Bacteria-based reverse mutation test (e.g. Ames-test)	Unscheduled DNA synthesis (UDS)	Comet-assay	Mammalian cell-based gene mutation test (e.g. MLA)	Sister chromatid exchange (SCE)	Chromosomal aberration test (CAT)	Mammalian erythrocyte micronucleus test
	In vitro	In vitro and ex vivo	In vitro and in vivo	In vitro	In vitro	In vitro	In vivo
Analyzed aspect	Induction of bacterial growth in histidine-free medium	Incorporation of labeled DNA components at sites of reparation DNA defect	Damaged DNA structures	Induction of colony formation in the presence of selective agent	BrdU-labeling and subsequent microscopic analysis	Chromosome analysis	Detection of DNA segments outside the nucleus

Further detection methods for mutagenesis are the in vitro mammalian cell gene mutation tests (to which the well known mouse lymphoma assay/MLA belongs). The assays of this category work by a comparable principle in that heterozygous deficient cell lines (e.g. the mouse lymphoma cell line L5178Y) are utilized. These cell lines have heterozygous deficiencies in specific enzymes due to mutations. For example, L5178Y cells have deficiency in thymidine kinase (TK, the heterozygous mutated L5178Y cell line is denoted as $tk^{+/-}$). The principle of this test method is based on the fact that $tk^{+/+}$ and $tk^{+/-}$ cells arrest their growth totally in the presence of the pyrimidine analogue trifluorothymidine (TFT), whereas the $tk^{-/-}$ cells are not susceptible to the cytotoxic effects of TFT. Thus, homozygous mutant cells are able to survive and proliferate in the presence of the selective agent TFT, whereas non-mutant and heterozygous cells ($tk^{+/+}$ and $tk^{+/-}$), which contain TK, are not [33]. For testing the mutagenic activity of a compound or its extract, respectively, known numbers of the heterozygous $tk^{+/-}$ cells are exposed for a suitable period of time (with and without the selective agent TFT) and subcultivated to determine cytotoxicity and to allow phenotypic expression of potentially induced mutations. After approximately 2 weeks colonies are counted. The mutant frequency is derived from the number of mutant colonies in selective medium and the number of colonies in non-selective medium. Thus, the higher the colony formation of the $tk^{+/-}$ cell strain in the presence of the test compound and the selective agent, the higher the mutagenicity of the compound. The range of genetic mutations detected with the MLA is broad: the MLA measures either gene mutations or heritable chromosomal events, including genetic events associated with carcinogenesis [34].

A further possibility for testing genotoxicity in vitro is the chromosome aberration test (CAT). The purpose of this method is to identify compounds that cause structural chromosome aberrations in cultured mammalian (human) cells. Chromosome aberrations and related events appear to cause alterations in genes regulating the cell cycle of somatic cells and are thus involved in the induction of cancer in humans and experimental animals. The test principle is based on the exposure of accepted cell cultures (e.g. cell lines, strains, or primary cells may be used) to the test compound. At predetermined intervals the cells are treated with a cell cycle-arresting substance (i.e. colchicin, a poison that does not allow the last steps of the cell cycle to proceed; colchicin arrests the cells in metaphase). Since the chromosomes are highly condensed in the metaphase, a microscopic analysis for the occurrence of chromosome aberrations can be performed.

Interestingly, both the CAT and the MLA often yield similar results. Therefore the utilization of both in the course of biomaterial testing may not be reasonable. Besides these above mentioned in vitro test systems a number of other assays are utilized to analyze genotoxic and/or mutagenic effects. One example is the so-called sister chromatid exchange assay (SCE) [35]. This assay examines the reciprocal exchanges between two sister chromatids of a duplicating chromosome. Through the use of special staining techniques the sister chromatids display differential staining in case of an exchange, permitting enumeration of SCE. The molecular basis of these mutations remains unknown, although it is presumed to involve complex DNA

breakage and reunion processes [36]. The SCE assay is usually performed on human peripheral blood lymphocytes.

A genotoxicity testing method that is intended for in vivo application is the mammalian erythrocyte micronucleus test. Micronuclei are small nuclei, separated from and additional to the main nuclei of the cells (1/5 to 1/20 the size of the nucleus). The micronuclei develop during cell division by lagging chromosome fragments or whole chromosomes by chromosome breaks or dysfunction of the spindle apparatus. For this test method mammalian erythrocytes are utilized. When the erythrocyte progenitors (the erythroblasts) develop into polychromatic erythrocytes (intermediate stage in the development of erythrocytes), the main nucleus is extruded but the already present micronuclei remain. Thus, an increase in the number of micronucleated erythrocytes is an indication for a genotoxic potential of the compound tested [37]. This method is especially relevant for assessing the mutagenic hazard of the test compound. Cells derived from bone marrow or peripheral blood can be used for this analysis. Usually mice are used for testing the effects in peripheral blood. For testing in bone marrow both mice and rats are recommended.

10.3.5 Tissue Specific Aspects of Biocompatibility Testing

The potentially broad applicability of biomaterials requires that the construction of the medical device and the testing scheme has to be adapted to the intended implantation site. As we have seen before, the problems in hemocompatibility testing deviate strongly from those of tissue compatibility for example. Therefore, as example, materials that are intended for utilization as a vascular implant must be tested by different biocompatibility tests than a material that is produced for a metallic hip prosthesis. The reasons for this are due to the different cell types and protein subsets that the material comes into contact with. However, different tissues also show large discrepancies regarding their requirements. For example soft tissue and hard tissue (bone) implants: soft tissue-implants destined for utilization in extensive injuries such as deep burns and lacerations need to permit filling of large defects and allow the ingrowth of adequate cells and blood vessel formation. In contrast to this, the routinely utilized metallic implants are not intended for cellular ingrowth. Furthermore, materials for soft tissue implants have to be soft and flexible, similar to the tissue that is to be replaced, whereas the materials for hard tissue replacement must fulfill the mechanical tasks of bones and withstand high mechanical forces. Mechanical forces are an important aspect in the field of biocompatibility, and due to the intensity of these forces especially in the field of osteocompatibility and odontocompatibility (i.e. the material's compatibility to perform with bone and tooth tissues).

Some similarities in the response of an implant are exhibited by all tissue types. The initial response after the insertion of any biomaterial leads to an injury that is accompanied by blood clotting, inflammation, the recruitment of inflammation-involved cells and afterwards by the development of the provisional tissue (the

so-called granulation tissue). These first steps occur regardless of the materials and the implantation site. However, the character and the course of the following repair processes depend once again on the type of tissue and the localization of the implant.

10.4 Animal Experimentation

Animal studies are used to determine the in vivo compatibility of biomaterials/medical devices and are performed prior to human testing to help predict the human response. Animal studies should only be performed when the information required is essential to characterize the test material and when no suitable scientifically validated in vitro test method is available. Furthermore, animal experiments shall not be performed before appropriate in vitro tests have been carried out, and the results have been evaluated.

In the course of the already described biocompatibility issues different in vivo assays have also been mentioned. This includes tests for sensitization, irritation, intracutaneous reactivity, subacute, acute and chronic toxicity, hemocompatibility, genotoxicity, implantation, carcinogenicity, reproductive and developmental toxicity, and biodegradation. The selection and design of animal tests should be appropriate and address the specific scientific objectives of the study. Furthermore, minimizing the pain, suffering, distress or lasting harm that might be inflicted on the test animals is a fundamental requirement of all regulating organizations (e.g. [38]). These fundamental requirements have to be applied and are regulated by respective supervisory authorities. The type of authority and the specific needs for the application of animal studies depends on each country. The responsible investigator has to verify that animal care is appropriate and medical care is guaranteed prior to, during and after the experiments.

The use of appropriate animal models is an important consideration in the safety evaluation of medical devices. Different animal models are typically used for specific applications. Examples for subject-bound animal models are the sensitization test in guinea pigs (see Section 10.3.3), the testing of irritation and intracutaneous reactivity, intramuscular implantation and pyrogenicity in rabbits, and subacute, acute, and chronic toxicity in mice and rats. In addition, the in vivo assessment of the functionality of medical devices is also being performed in sheep, dogs, pigs, and calves. Sheep are commonly used for the evaluation of heart valves since the sizes are nearly proportional to human dimensions. Furthermore, pigs are often used for the in vivo testing of vascular grafts and artificial hearts are mainly tested in calves. Because of the homology to humans, the most appropriate animal models are primates. However, ethical concerns regarding the utilization of intelligent animals and the costs of purchasing and maintaining such primate colonies are substantial.

The acquisition of the animal experimental data depends on the type of test: for example the sensitization in guinea pigs can be detected on the skin by redness and swelling, the pyrogenicity (fever inducing capacity) of a material is usually

detected by rectal temperature control in rabbits. The analyses of body tissues that are exposed to materials are more complex and must be analyzed by histological means. For example, the tissues are analyzed for the occurrence of inflammatory cell populations. It should be noted, however, that whereas a positive response to a material (i.e. a pathological reaction in the animal) can be interpreted relatively easily, the absence of a response in the animal does not necessarily prove compatibility in humans. In conclusion, animal experiments are expensive, complex and difficult to interpret and require experienced personnel and expertise. However, if performed properly they are useful and necessary prior to human trials.

10.5 Alternatives to Animal Experimentation

As mentioned above the demand for the utilization of alternatives to animal studies is a high priority. All in vitro tests addressed in the previous sections are approved alternatives to animal testing since they give information about possible cellular impairment induced by the material or its extraction product prior to animal testing. However, these in vitro test systems are based on the utilization of one cell type (bacterial or mammalian) and do not reflect the complexity of the multicellular/multitissue-organism, the human body.

More recently, more complex in vitro test systems working with multiple (human) cell types have been added as alternatives. These are to a degree under consideration and partially validated and accepted by different regulating organizations. Examples for these are the human skin equivalent tests (e.g. EpiDerm^{TM}, EpiSkin^{TM}). These tests consist of normal, human skin cells, which are cultured to form a multi-layered model of human skin. The significance of these human skin equivalent tests has been established through validation studies by the European Centre for the Validation of Alternative Methods (ECVAM) in Ispra, Italy. These three-dimensional models allow the application of the test material directly to the "artificial skin" surface and permit the assessment of dermal toxicity via different parameters (cytotoxicity, histological examination, detection of inflammatory mediators etc.) [39].

Medical devices that come into direct or indirect contact with blood must be examined for pyrogens. Pyrogens are defined as fever-inducing substances. Pyrogens can be components of the cell wall of Gram-negative bacteria (endotoxins) or chemical compounds. In the past pyrogen testing was always performed in rabbits (see Section 10.4). Fluid extracts of the test materials were administered intravenously to rabbits and afterwards the animals ́rectal temperature was monitored over the course of several hours. A significant rise in temperature indicated the presence of pyrogens. However, this test is subjected to well-documented drawbacks, including marked species and strain differences in sensitivity [40]. In addition to this conventional pyrogen test system alternative methods are currently available. Alternatively to the rabbit test, the limulus amoebocyte lysate (LAL) test is used. The LAL test is based on the blood cells of the Horseshoe crab *Limulus polyphe-*

mus. It has been shown that the blood lysates of *Limulus* exhibit clotting after contact to the above mentioned endotoxin. Thus, the LAL test detects only one specific pyrogen, the endotoxin. As a replacement method for pyrogen animal studies the so-called human whole-blood test (also in vitro pyrogen test (IPT)) has been developed and validated. The materials to be examined are incubated in human whole blood containing a number of different, relevant cell types (see Section 10.3.2). The blood is then examined for its pro-inflammatory response, the release of the pro-inflammatory cytokine IL-1β by an enzyme-linked immuno-assay (ELISA). Beside the aspect of replacement of animal studies, the test method possesses further advantages: i.e. that the system is of human origin and can be used for all potential pyrogens (not only bacterial cell wall components like in the LAL test) [40, 41].

Other complex in vitro models utilizing human cells are accepted for research purposes but have not yet been certified for the testing of medical devices. An example of such a model is shown in Fig. 10.10. This in vitro model examines the capability of human endothelial cells (the cells that line the blood vessels) to form new blood vessels, a process that is called angiogenesis. Angiogenesis is a fundamental requirement in wound healing. This model system of angiogenesis in vitro makes use of human endothelial cells suspended in a matrix of extracellular proteins (fibrin and collagen) resembling the wound healing matrix. After the addition of specific angiogenesis-inducing factors in vitro capillaries evolve within a few days (Fig. 10.10a). If performed on a titanium surface (commercial pure titanium/cpTi) the cells develop the normal capillary phenotype (Fig. 10.10b). Cells on a Co28Cr6Mo-alloy exhibit striking deviations compared to the other two surfaces: the formation of in vitro capillaries is completely inhibited, whereas the cell viability is not impaired (Fig. 10.10c). Since angiogenesis is fundamental in
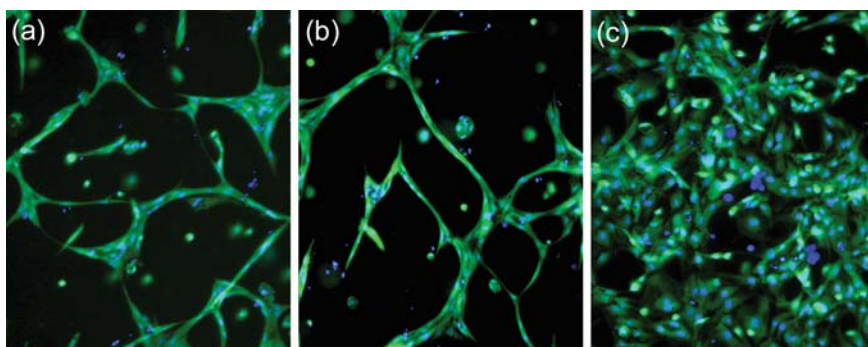
Fig. 10.10 A model of angiogenesis for use in vitro for biocompatibility studies. (**a**) human endothelial cells embedded into a three-dimensional matrix consisting of fibrin and collagen. After stimulation with angiogenesis-inducing factors the cells form in vitro capillaries within 5–7 days; (**b**) angiogenic-stimulated endothelial cells on a cpTi-surface; (**c**) endothelial cells on a Co28Cr6Mo-alloy exhibit striking deviations compared to the other two surfaces (digital overlay of a vital and a nuclear staining)

wound healing, the impairment of the angiogenic capacity of endothelial cells by a material could lead to the disturbance of wound healing and aseptic implant loosening [42].

Other complex in vitro models are also currently under validation for approval in the biocompatibility safety testing program. With increasing knowledge in this field and concerted, stringent validation processes the number of relevant approved alternatives to animal testing should increase within the next years. In conclusion, these alternative methods are beneficial for the replacement and reduction of animal testing and are often superior to animal studies. However, despite progress and considerable effort in this direction, currently no satisfactory in vitro test system has been devised to completely eliminate the requirement for in vivo testing.

References

1. Williams DF. Definitions in Biomaterials, Elsevier: Amsterdam, 1987.
2. ISO. Biological Evaluation of Medical Devices – Part 12: Sample Preparation and Reference Materials, 2007.
3. Campisi J. Cancer, aging and cellular senescence. In Vivo, 2000, 14(1): 183–188.
4. Pedrinaci S, et al. Protein kinase C-mediated regulation of the expression of CD14 and CD11/CD18 in U937 cells. Int J Cancer, 1990, 45(2): 294–298.
5. Freshney RI. Culture of Animal Cells: A Manual of Basic Techniques, 4th edn, Wiley & Sons: New York, 2000.
6. ISO. Biological Evaluation of Medical Devices – Part 5: Tests for In Vitro Cytotoxicity, 1999.
7. Peters K, et al. Cell type-specific aspects in biocompatibility testing: The intercellular contact in vitro as an indicator for endothelial cell compatibility. J Mater Sci Mater Med, 2008, 19(4): 1637–1644.
8. Decker T and Lohmann-Matthes ML. A quick and simple method for the quantitation of lactate dehydrogenase release in measurements of cellular cytotoxicity and tumor necrosis factor (TNF) activity. J Immunol Methods, 1988, 115(1): 61–69.
9. Cory AH, et al. Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. Cancer Commun, 1991, 3(7): 207–212.
10. Denizot F and Lang R. Rapid colorimetric assay for cell growth and survival. Modifications to the tetrazolium dye procedure giving improved sensitivity and reliability. J Immunol Methods, 1986, 89(2): 271–277.
11. Gillies RJ, Didier N, and Denton M. Determination of cell number in monolayer cultures. Anal Biochem, 1986, 159(1): 109–113.
12. Klein CL, et al. A new quantitative test method for cell proliferation based on detection of the Ki-67 protein. J Mater Sci Mater Med, 2000, 11(2): 125–132.
13. Assuncao Guimaraes C and Linden R. Programmed cell deaths. Apoptosis and alternative deathstyles. Eur J Biochem, 2004, 271(9): 1638–1650.
14. ISO. Biological Evaluation of Medical Devices – Part 4: Selection of Tests for Interactions with Blood, 2002.
15. Garner LA. Contact dermatitis to metals. Dermatol Ther, 2004, 17(4): 321–327.

16. Basketter D, et al. The chemistry of contact allergy: Why is a molecule allergenic? Contact Dermatitis, 1995, 32(2): 65–73.
17. Kanerva L. Cross-reactions of multifunctional methacrylates and acrylates. Acta Odontol Scand, 2001, 59(5): 320–329.
18. Dannaker CJ. Allergic sensitization to a non-bisphenol A epoxy of the cycloaliphatic class. J Occup Med, 1988, 30(8): 641–643.
19. Shanklin DR and Smalley DL. The immunopathology of siliconosis. History, clinical presentation, and relation to silicosis and the chemistry of silicon and silicone. Immunol Res, 1998, 18(3): 125–173.
20. Oleffe J and Wilmet J. Generalized eczema and an osteosynthesis screw. Arch Belg Dermatol Syphiligr, 1972, 28(3): 275–278.
21. Oleffe J and Wilmet J. Generalized dermatitis from an osteosynthesis screw. Contact Dermatitis, 1980, 6(5): 365.
22. Budinger L and Hertl M. Immunologic mechanisms in hypersensitivity reactions to metal ions: An overview. Allergy, 2000, 55(2): 108–115.
23. ISO. Biological Evaluation of Medical Devices – Part 10: Tests for Irradiation and Delayed Type Hypersensitivity, 2002.
24. Buehler EV. Delayed contact hypersensitivity in the guinea pig. Arch Dermatol, 1965, 91: 171–177.
25. Magnusson B and Kligman AM. The identification of contact allergens by animal assay. The guinea pig maximization test. J Invest Dermatol, 1969, 52(3): 268–276.
26. Eloy R, et al. Current and future issues in sensitisation testing. Med Device Technol, 2001, 12(7): 12–15.
27. Kimber I and Weisenberger C. A murine local lymph node assay for the identification of contact allergens. Assay development and results of an initial validation study. Arch Toxicol, 1989, 63(4): 274–282.
28. Moustacchi E. DNA damage and repair: Consequences on dose-responses. Mutat Res, 2000, 464(1): 35–40.
29. Ames BN. Identifying environmental chemicals causing mutations and cancer. Science, 1979, 204(4393): 587–593.
30. Maron DM and Ames BN. Revised methods for the Salmonella mutagenicity test. Mutat Res, 1983, 113(3–4): 173–215.
31. Williams GM. Carcinogen-induced DNA repair in primary rat liver cell cultures; a possible screen for chemical carcinogens. Cancer Lett, 1976, 1(4): 231–235.
32. Collins AR. The comet assay for DNA damage and repair: Principles, applications, and limitations. Mol Biotechnol, 2004, 26(3): 249–261.
33. Clive D, et al. A mutational assay system using the thymidine kinase locus in mouse lymphoma cells. Mutat Res, 1972, 16(7): 77–87.
34. Applegate ML, et al. Molecular dissection of mutations at the heterozygous thymidine kinase locus in mouse lymphoma cells. Proc Natl Acad Sci U S A, 1990, 87(1): 51–55.
35. Kato H and Shimada H. Sister chromatid exchanges induced by mitomycin C: A new method of detecting DNA damage at chromosomal level. Mutat Res, 1975, 28(3): 459–464.
36. Dillehay LE, Jacobson-Kram D, and Williams JR. DNA topoisomerases and models of sister-chromatid exchange. Mutat Res, 1989, 215(1): 15–23.
37. von Ledebur M and Schmid W. The micronucleus test. Methodological aspects. Mutat Res, 1973, 19(1): 109–117.
38. ISO. Biological Evaluation of Medical Devices – Part 2: Animal Welfare Requirements, 2006.
39. Monteiro-Riviere NA, et al. Comparison of an in vitro skin model to normal human skin for dermatological research. Microsc Res Tech, 1997, 37(3): 172–179.
40. Hartung T, et al. Novel pyrogen tests based on the human fever reaction. The report and recommendations of ECVAM Workshop 43. European Centre for the Validation of Alternative Methods. Altern Lab Anim, 2001, 29(2): 99–123.

41. Hoffmann S, et al. International validation of novel pyrogen tests based on human monocytoid cells. J Immunol Methods, 2005, 298(1–2): 161–173.
42. Peters K, et al. Software supported quantification of angiogenesis in an in vitro culture system. Examples of applications in studies of basic research, biocompatibility and drug discovery, in Zubar RV, ed. Trends in Angiogenesis Research, Nova Science Publishers Inc.: New York, 2005, pp. 103–123.

Chapter 11
Biomaterials for Dental Applications

Sarit B. Bhaduri and Sutapa Bhaduri

11.1 Introduction

Based on population projections, the number of teeth at risk for dental disease in 1980 was 2.8 billion for those over the age of 65, and is expected to increase to 5 billion teeth at risk by 2020 [1]. Dental materials are going to play an important role in the treatment of such ailments. As a matter of fact, dental materials have a long checkered history. From a historical viewpoint, many different materials, both natural and artificial, have been used in treating various oral disorders. Obviously for the materials to succeed, they have to be at least biocompatible and should be able to bond with teeth or bone. These materials should appear like natural tooth, if visible, and exhibit properties similar to those from various parts of a tooth.

While dental materials are, in general, classified as preventive materials or restorative materials, the primary use of dental materials is for restoration. Dental restorative materials are made up of different types of materials such as metals, ceramics, polymers, and composites. The selection of the type of dental restorative material is dependent on many factors, such as the patient, the dentist, the material, and characteristics of the tooth itself. Important parameters for choosing dental restorative materials include strength, hardness, wear resistance, chemical resistance, and esthetic appearance. Metals are primarily used where strength and durability are needed. Typical uses of metals include amalgams, casting alloys, and porcelain fused to metal (PFM) restorations. Ceramics are also used as in conjunction with metal as PFM or as stand-alone all-ceramic materials. They are used where chemical durability and esthetics are important. They are used as veneers, crowns, as protective layers etc. Polymeric materials are used as cements and as cavity fillings in conjunction with particulate ceramic filler materials. Polymers are useful because of esthetics and adhesion properties.

This chapter will review important features of metals, ceramics, polymers, and composites in restorative dentistry. The subsections will describe the compositions, the relevant materials, scientific aspects, and related materials properties.

Where appropriate, specialty processing techniques associated with manufacturing/fabrication will be described.

11.2 Historical Perspectives

The use of both artificial as well as natural materials in dentistry is well known for about 5,000 years [2]. Around 700 B.C. Etruscans used elephant or hippopotamus ivory to construct partial dentures fixed to natural teeth by gold wires. Cavities in teeth were filled with a range of materials such as gum, metals, ivory, etc. Fossil evidence indicates that 2,000 years ago, the art of dental implantology was already taking off by replacing lost teeth with alloplastic or homologous materials such as human and animal teeth, carved human and animal. These replacements were not only functional but they were used to meet esthetic needs. Ancient dental implant designs emerged in the Middle East, Western Europe, Asia, and Central and South America [3].

About 1,000 years ago, the Spaniard Alabucasim recommended the replacement of teeth by transportation (transfer of living tissue), indicating that this was an acceptable strategy for replacing missing teeth. Several hundred years ago in France and England, it was fashionable to replace lost teeth with transplanted teeth obtained from young people who were paid for their extracted teeth [4]. In 1774, Duchateau and de Chemont developed a process for making porcelain dentures. In 1808, Fonzi, an Italian dentist produced an individual porcelain tooth. In 1816, Taveau developed the first amalgam in France. A widespread use of vulcanized rubber dentures took place after Charles Goodyear invented the material in 1839. In 1843, Giovanni d'Arcoli showed how to use gold foils in filling cavities.

These are some of the very important milestones of dental materials developments, especially in restorative dentistry, and laid the foundation for the further developments of science and technology, as we know today.

11.3 Metals for Dental Application

Metals are used in dentistry because of their strength and durability. However, metals lack proper appearance and chemical durability. The current status is that virtually all metallic materials are suitable for use with minimal risk. However, complete elimination of "toxic" elements from all dental materials should be an everlasting goal.

11.3.1 Amalgams

Since 1895, dental amalgams have been used by dentists as direct filling materials [5]. Thus, these materials have a proven track record of more than 200 years. In

spite of concerns about the presence of Hg, it is still safe to use, easy to manipulate, inexpensive and durable. This is in spite of the competition from composites and glass-ionomer restorative materials. Some patients prefer amalgams to other alternatives because of their safety and cost-effectiveness.

Dental amalgams are formed by chemical reactions of Hg with intermetallic compounds in the Ag–Sn system as well as Cu. The reaction of liquid Hg with solid particles is known as amalgamation (also known as trituration). Mercury makes up about 45–50% of a typical amalgam composition. Once the amalgamation reactions are complete, the mixture is placed into the patient's teeth under pressure. The mixture solidifies and within a few minutes the amalgam increases in strength and hardness. In an ideal situation, an amalgam should retain its dimensions throughout the restoration period. However, in reality this is not the case. A variety of factors lead to dimensional changes causing several complications. For example, contraction results in microleakage, plaque accumulation, etc.

The various important phases that take part in the amalgamation process are shown in Table 11.1. In a typical low copper composition, the following reaction occurs [6]:

$$\gamma + \text{Hg} \rightarrow \gamma_1 + \gamma_2 + \gamma \, (\text{remnant})$$
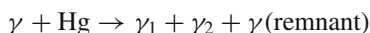
The microstructure consists of γ embedded in the matrix of γ_1 and γ_2. Cu and Zn, although present in minor amounts, play important roles. Cu improves the physical properties of amalgams and hence high Cu containing compositions were developed [7]. The high Cu containing amalgams do not contain γ_2 phase. This is because Sn preferentially reacts with Cu to produce the η phase.

The shape of the original γ particles and Cu content classify the amalgams. Depending on how they are processed, particles are referred to as lathe-cut (elongated particles) and spherical (produced by the atomization process). These are available in traditional as well as high copper compositions. Finally, the lathe-cut and spherical powders can be admixed. The properties of such compositions are given in Table 11.2. Most of the modern alloys are high Cu containing alloys.

The reaction between Hg and γ particles are performed in mechanical amalgamators. These motorized machines vibrate at a very high speed to mix the Hg and γ

Table 11.1 Symbols and stoichiometry of phases that are important in dental amalgams [5]

Phases in amalgam alloys and set dental amalgam	Stoichiometric formula
γ	Ag_3Sn
$\gamma 1$	Ag_2Hg_3
$\gamma 2$	$Sn_{7-8}Hg$
ε	Cu_3Sn
η	Cu_6Sn_5
Silver–copper eutectic	Ag–Cu

Table 11.2 Physical properties of dental amalgams [5]

Type	Compressive strength (MPa) (1 day)	Tensile strength (MPa)	Knoop hardness	Creep (%)	Dimensional change (μm/cm)
TL	430	52	146	2.05	8
TS	444	55	174	0.21	0
HCS	486	63	173	0.07	-7
HCL	477	45	174	0.17	5
HCB	434	50	155	0.24	-7
GA	383	57	–	0.17	16

TL – Traditional lathe cut, TS – Traditional spherical, HCS – High copper spherical, HCL – High copper lathe cut, HCB – High copper blend, GA – Alloy for gallium amalgam.

particles placed within a vial. For proper mixing, it is important to pay attention to the amalgamation time so that all the free Hg is consumed. Once the proper mix is obtained, it is condensed into the properly prepared cavity. Either hand condensation or mechanical condensation can be used.

As an alternative to the Hg-containing amalgams, gallium alloys were developed using the fact that Ga is a very low melting point metal [8]. Ga can be amalgamated with silver-tin-copper alloys in the same conventional fashion. The commercial compositions contain significant amounts of Pd to improve the corrosion properties. After triturating, the resulting microstructure consists of $CuGa_2$ and $PdGa_5$ surrounding the unreacted alloy particles within a mixture of Ag_9Iu_4.

11.3.2 Biocompatibility of Dental Amalgams

The usage of dental amalgam has remained a controversial issue. According to the U.S. Public Health Service, dental amalgams do not possess any immediate health hazard. However, studies found that mercury amalgams are unstable due to mercury's low vapor pressure and galvanic action, resulting leaking mercury vapor continuously into the lungs and saliva. [9]. There is no evidence in the published literature that shows that the minute amount of mercury can cause poisoning to patients, except for those who have allergic reaction to mercury [10].

11.3.3 Casting Alloys

Along with amalgams, various casting alloys play an important role in restorative dentistry. From an historical viewpoint, Taggart introduced the cast inlay fabrication in 1907 [11]. He practiced the "Lost Wax (Investment)" casting technique and since then, it remains as an important technique for the fabrication of restorations. The casting alloys can be subdivided into all-metal restorations, porcelain fused-to-metal (PFM) restorations and soldering alloys.

The alloys can also be grouped in terms of their compositions. American Dental Association (ADA) classifies them as High Noble (HN), Noble (N), and Predominantly Base Metal (PB) [12]. According to the periodic table, the eight noble metals are gold, silver, and the platinum group metals. However, in an oral environment, silver is not necessarily considered noble. According to the ADA specifications, HN must contain more than 40 wt% of gold and the rest can be either gold or platinum group metals. Noble group should contain more than 25 wt% of noble metals, while predominately base metals (PB) should contain <25 wt% of noble metals. Table 11.3 describes various alloy systems in all-metal and PFM restorations. This discussion will begin with NB and N and ending with PB materials.

For use in restorative dentistry, the casting alloys must satisfy many different criteria. These criteria include biocompatibility, corrosion resistance, tarnish resistance, allergenic reactions, esthetics, thermal properties, melting temperatures, extent of shrinkage, strength requirements, castability, bonding to porcelain, and economical aspects.

The casting alloys are further classified according to their mechanical properties. According to the ISO/DIS 1562 standard for casting gold alloys, several types of alloys are proposed [12]. These are:

- Type I (low strength) – for making castings subjected to very low stress (in inlays), with minimum yield strength of 80 MPa and a minimum p.c. elongation of 18%
- Type II (moderate strength) – for castings subjected to moderate stress (inlays, onlays, and full crowns) with a minimum yield strength of 180 MPa, and minimum p.c. elongation of 10%
- Type III (high strength) – for castings subjected to high strength (onlays, thin copings, and full crowns), the minimum yield strength is 270 MPa, and the minimum p.c. elongation is 5%

Table 11.3 Classification for full-metal and PFM restorations

Metal type	All-metal restorations	PFM restorations
High noble (HN)	Au—Ag—Pd	Pure Au (99.7 wt%)
	Au—Pd—Cu—Ag	Au—Pt—Pd
	HN metal-ceramic alloys	Au—Pd—AgAu—Pd
Noble (N)	Ag—Pd—Au—Cu	Pd—Au
	Ag—Pd	Pd—Au—Ag
	Noble metal-ceramic alloys	Pd—Ag
Predominately base metal (PB)	Cp—Ti	Cp—Ti
	Ti—Al—V	Ti—Al—V
	Ni—Cr—Mo—Be	Ni—Cr—Mo—Be
	Ni—Cr—Mo	Ni—Cr—Mo
	Co—Cr—Mo	Co—Cr—Mo
	Co—Cr—W	Co—Cr—W

- Type IV (extra high strength) – for castings subjected to very high stress (partial denture framework) with minimum yield strength of 360 MPa and the minimum p.c. elongation of 3%

Types I and II are mostly used for inlays and their use is becoming lesser with the esthetically appealing natural looking materials. Types III and IV have the capability to strengthen due to several solid state transformations resulting in their high strength.

As shown in Table 11.3, all-metal restorations do not have the characteristics necessary for the PFM applications. This is because PFM-capable alloys must satisfy additional criteria in order to be compatible with porcelain. These require that: (1) the alloys melt at temperatures above $1,100\,°C$; (2) they form thin stable oxides to promote atomic level bonding with porcelains; (3) they are compatible with thermal expansion coefficients in metal and porcelain (between 12.7–$14.8 \times 10^{-6}/°C$ for alloys and 10.8–$14.6 \times 10^{-6}/°C$ for porcelains; (4) they minimally creep or sag during firing of porcelains.

As seen in Table 11.3, at least three different classes of predominantly base metal (PB) alloys are used. These are cobalt-chromium, nickel-chromium, and titanium alloys. The former two are mostly used in metal-ceramic restorations and rarely used for all-metal restorations. All of these alloys were developed for restorative dentistry applications in the mid-1970s when the price of gold increased rapidly. Chromium provides tarnish and corrosion resistance of these alloys. Cobalt and nickel increases the Young's moduli of Ni-Cr alloys and can be further sub-divided into beryllium-containing alloys and those containing no beryllium. Be addition decreases the melting point of alloys, which is useful for PFM applications.

11.3.3.1 Titanium and Related Alloys

Titanium has widespread uses in dentistry, which necessitates a separate section. Titanium is light, strong, ductile, and corrosion resistant [13]. Titanium's high strength-to-weight ratio has been a major reason for titanium's popularity. In the early 1950s, the newborn jet aircraft industry led to the rapid growth in titanium development technology. Starting in the 1960s, industrial uses of titanium began to become significant. Titanium was found to be one of the most "corrosion-resistant" structural metals, and marine applications such as power plant heat exchangers and water jet propulsion systems started using them [13].

Health conscious patients and dentists alike are becoming more and more demanding in their quest to attain the "smile of the future." Both are demanding a high degree of comfort and 100% organic compatibility. While titanium can match these demands, other dental alloys can cause unpleasant side effects of chemical and physical reactions in the mouth. Thus, modern dentistry is increasingly turning to titanium as an alternative to other conventional dental alloys.

In comparison to gold alloys, titanium yields a wider range of strength properties at a specific gravity that is four times less than gold. These physical proper-

ties allow for stronger and lighter crown and bridge frameworks. For the sensitive mouth, titanium is the best solution, and is especially suitable for patients suffering from allergies. Because of its low weight, the new titanium prosthesis causes little inconvenience to the patient. Its low thermal conductivity means that hot and cold foods can be enjoyed without discomfort. For the patient, these benefits add up to a new confidence level, providing a greater quality and enjoyable lifestyle.

Titanium has been established to be "biocompatible" with the human body and is classified as "hypoallergenic". Titanium is totally resistant to attack by bodily fluids and is not recognized as a foreign material when utilized in implants such as artificial hips, pins for setting bones and other biological implants. In dentistry, titanium actually encourages bone to grow to it, and the rapid growth in popularity of dental implants can be attributed in large part to titanium. With the application of unique, new dental porcelains to titanium, dentist can prescribe the most advanced and biologically safe porcelain to metal crowns and bridges [12].

Commercially pure (CP) with 99% titanium, and small quantities of Fe, C, O, N, H has five different grades (1–4 and 7) depending on the oxygen and iron content. The strength goes up from grade 1 to grade 4 [12, 13]. The most significant advantage of Ti is the presence of a thin passivating oxide film on its surface, which provides it a very high corrosion resistance. However, the most important disadvantages of Ti are its reactive nature at the high melting point of $1,668\,°C$, which needs special atmosphere during casting.

Titanium undergoes phase transformations from α phase (HCP) at $885\,°C$ to β phase (BCC) structure. Alloying additions stabilize various phases. While vanadium stabilizes the β phase, aluminum stabilizes the α phase. Thus the most important alloy Ti—6Al—4 V is an α–β alloy. Table 11.4 lists the physical and mechanical properties of Ti and representative alloys.

As noted before, an oxide layer (TiO_2) of 3–5 nm that forms on the surface makes Ti and its alloys extremely corrosion resistance. This is called the passive condition of titanium. Both CP titanium and its alloys have been shown to be biocompatible. The chemical inertness of implant material plays a very important role in long-term clinical success of osseointegrated implants. It is well known that most implant materials leak ions into the tissues. Some undesirable biological effect include: (1) alteration of fibroblast and calcium deposition during osteogenesis; and (2) peri-implant bone formations and the deposition of calcium ions on titanium surface.

Table 11.4 Physical and mechanical properties of Ti and its alloys

Alloying designation	Microstructure	E (GPa)	YS (MPa)	UTS (MPa)
Cp–Ti	α	105	690	780
Ti–6Al–4 V	α/β	110	850–900	960
Ti–6Al–7Nb	α/β	105	900	1,000

11.3.3.2 Casting and Soldering

Other than the soldering alloys, the casting alloys are investment cast. The investment casting process can produce the fine features of restorations with high dimensional accuracy. ANSI/ADA specification 2 describes several types of investments (molds) into which alloys can be cast. The investments can be made of gypsum or phosphate bonded. Different types of casting machines are available for filling the mold. In laboratories, either a torch flame, or induction/resistance heating can be used to melt the alloys, followed by centrifugal casting of the molten alloys into a cavity. Alternatively, vacuum melting and Ar pressure casting can be performed. After casting and solidifying, the investment is removed, and the casting is given the final shape. For achieving desired mechanical properties, the castings can be further processed. Types I and II alloys do not harden, while Types III and IV alloys are hardenable alloys. Typical heat-treatment of alloys requires heating to temperatures between 200 and 450 °C followed by quenching in water.

Soldering enables a technician to join simple cast parts. Dental casting alloys that can be soldered are both noble and base metal alloys. The term "presoldering" is used for metals prior to firing the veneers and "post soldering" is used after veneering had taken place. Like the conventional soldering process in structural materials, soldering in dentistry needs filler and a flux, besides some suitable heat source. They are supposed to perform the exact tasks that their structural counterparts do. Gold-base fillers are normally used for noble metal alloys and silver-base fillers are used for base metal alloys. Traditional heat sources such as a torch can be used for soldering. Various lasers are now available for soldering Ti-alloys.

11.3.4 Wrought Alloys as Orthodontic Wire

Orthodontic wires are used in order to correct misaligned teeth. Usually, wires are made up of wrought alloys. As opposed to castings, they are usually cold worked and eventually formed by mechanical processes such as rolling, extrusion, and drawing.

One of the requirements of orthodontic wire is that they must have low elastic modulus, high yield strength, and good corrosion resistance. Besides, materials have to be sufficiently ductile for easier fabrication. Specifically, most wires should easily be soldered or brazed together. Four major alloys of orthodontic wires include austenitic stainless steels, cobalt—chromium—nickel, nickel—titanium, and beta titanium. Among these, even though stainless steel and cobalt chromium nickel alloy wires are far less expensive than Ti—Ni and beta Ti wires, former materials are not so popular. Table 11.5 describes the various materials and their properties.

The shape memory TiNi alloys call for some additional discussion because of their interesting and beneficial properties for orthodontic applications. These alloys are associated with such properties as "Shape Memory" and "Superelasticity".

Table 11.5 Wrought alloys and their properties

Materials	Composition	Yield strength	Elastic modulus	Properties
Stainless steel	18% Cr–8% Ni, <0.15 wt% C	205 MPa, annealed condition 1,380 MPa, cold worked	190 GPa	
Cobalt based alloys	39% Co–19% Cr–14% Ni–6% Mo–1.5% Mn	69 MPa, non-cold worked 1,240–2,240 MPa, cold worked	190 GPa (RT)	It can be easily deformed and shaped
Nickel titanium alloys	55 wt% Ni–45% Ti	70–560 MPa	40 GPa	Corrosion resistant
β-Titanium alloys	78Ti–11.5Mo–6Zr–4.5Sn	62 MPa	70–110 GPa	Good ductility, weldability

Buehler and Wang discussed the shape memory effect in an equimolar composition of Ti and Ni [14]. The shape memory effect refers to an alloy's ability to return to a previously defined shape when heated according to appropriate schedules. While many compounds with shape memory properties have been identified, TiNi is in the forefront in orthodontic applications [15]. The diffusionless martensitic transformations are able to generate high strains and/or forces in TiNi. In addition, TiNi is biocompatible.

Four temperatures usually describe the SMA effect in TiNi: martensite start (M_s), martensite finish (M_f), austenite start (A_s), and austenite finish (A_f). The low temperature martensitic phase is soft and deformable. As the temperature is increased, there is diffusionless transformation to the high temperature austenitic phase. The thermal cycling between these phases actually creates hysteresis. Furthermore, SMAs can result in superelasticity with very high recoverable strains as compared with traditional elastic materials. It is critical to pay some attention to the superelastic behavior. This requires deformation to occur by martensitic transformation rather than by plastic deformation. In TiNi, plastic deformation is more dominant between 100 and 150 °C. Based on thermal and mechanical measurements, a superelastic window in temperature can be obtained, where there is minimal remnant permanent strain, resulting in ideal superelasticity. The characterization of SMAs requires precision determination of aforementioned transformation temperatures using Differential Scanning Calorimeter (DSC).

In the past, the shape memory alloys were given shape at high temperature exposure. After cooling down, the clinician would manipulate the wire in the patient's mouth and it would get into the original shape of the wire. The superelastic wires have much more beneficial springback (the ratio of yield strength to elastic modulus), than its competitors. More recent alloys contain some amount of Copper. The addition of Cu approximately 5 wt% instead of Ni and 0.29 wt% of Cr makes the transformation purely shape memory without the presence of the intermediate-R phase and shows superelastic behavior at 37 °C, the temperature of the human body.

As discussed before, β-titanium orthodontic wires have been used as orthodontic wire. β phase is stabilized by the addition of Mo [16]. The favorable springback behavior and formability are also advantageous.

Although Ti—Ni shape memory alloys and 316L austenitic stainless steel are used in many dental applications, they contain Ni, which causes metallic allergy. Non-allergenic nickel-free stainless steel, in which mechanical strength and corrosion resistance are improved using nitrogen as a substitute for nickel, has drawn attention as a new biomaterial. However, because the material itself becomes extremely hard, depending on the forming and machining conditions, Ni-free stainless steel dental devices and technologies for manufacturing small-scale, complex-shaped dental devices from Ni-free stainless steel have not reached practical application. Technology development in this field is strongly desired.

11.3.5 Dental Implants

A dental implant is an artificial tooth root (synthetic material) that is surgically anchored into the jaw to hold a replacement tooth or bridge in place [17, 18]. The benefit of using implants is that they do not rely on neighboring teeth for support. They are permanent and stable. Implants are a good solution to tooth loss because they look and feel like natural tooth. Implant material is made from different types of metallic and bone-like ceramic materials that are compatible with body tissue. Today, dental implants are becoming more predictable. Studies have demonstrated success rates ranging from 80% to 92% for the maxilla over 5–10 years. Advances in surgical procedures, such as maxillary sinus lifts, lateral inferior alveolar nerve placement, and guided bone regeneration (GBR) allow the practitioner to place implants in patients that would not have received this treatment option 20 years ago [19].

In general oral implants can be categorized into three main groups:

- Endosseous implants
- Subperiosteal implants
- Transosseous implants

11.3.5.1 Endosseous Implants

Endosseous implants are the most frequently used implant today [20]. Based on their shape, function, surgical placement, and surface treatment, they could be further categorized into several sub-categories. However for the present purpose, only three main categories will be described. The three main categories are root-form, ramusframe, and blade-form.

Endosseous root-form implants are surgically inserted into the maxilla or mandible, whereby the implant body generally penetrates the cortical bony plate. In 1939, Strock first attempted to change the shape of a dental implant from that of a rootlike tooth to an implant with a threaded body resembling a wood screw [3].

Recognizing the need for implants to be biologically compatible, he made the implant from an alloy of chromium—cobalt—molybdenum (called Vitallium). He found that certain metals, when in contact with tissue fluids, produce a galvanic action. This ultimately corrodes the metal appliance. Vitallium was the only metal used at that time that produced no electrolytic activity when implanted into biological tissues. Among these, Strock's implants are the earliest documented long-term endosseous dental implants. Furthermore, he appreciated the need to direct axial forces and minimize nonaxial forces to the dental implants. Rootform implants are considered by some as the standard of care in oral implantology [21]. The implant, a screw shaped form like the original root, is screwed into the patient's jawbone. The screw's shape and size are based upon the tooth being replaced and/or the size and shape of the remaining jawbone. The implant is generally made out of titanium alloys, aluminum oxide, Vitallium, commercially pure titanium or sapphire [21]. Once the implant has been placed into the jawbone, a few months are required for the implant and the bone to fuse. This fusing creates a solid, secure implant to which an abutment can be attached. Before the prosthesis phase, a temporary crown or healing abutment is attached to the implant for up to eight weeks. Gum tissue grows around the temporary crown. After the implant has fused, an impression of the tissue shape and the implant is taken. From this impression a permanent tooth is fabricated. The permanent-fabricated tooth is then cemented to a custom abutment that is then screwed onto the implant. Root form implants are the closest shape and size to the natural tooth root. They are commonly used in wide, deep bone to provide a base for replacement of one, several or a complete arch of teeth.

Ramusframe implants belong in the category of endosseous implants. These implants are designed for the toothless lower jaw only and are surgically inserted into the jaw bone in three different areas: the left and right back area of the jaw (the approximate area of the wisdom teeth), and the chin area in the front of the mouth.

In 1968, Linkow introduced a flat titanium endosteal blade implant, which often serves as a means of using the narrow and/or shallow areas of remaining alveolar bone where dimensions do not permit the use of the cylindrical or root-shaped implants [3]. These implants primarily form a connective-tissue interface, a process called fibro-osseous integration. Connective tissue forming around the implant takes the form of collagenous fibers that are parallel to the dental implant surface. However, since these are also surgically placed into the bone we categorize them into the Endosseous implant category. Blade implants have a long track record, much longer then the Rootform implants. Their name is derived from their flat, blade-like (or plate-like) portion, which is the part that gets embedded into the bone. Blade implants are not frequently used any more, however they do find an application in areas where the residual bone ridge of the jaw is either too thin (due to resorption) to place conventional Rootform implants.

11.3.5.2 Subperiosteal Implants

The second type of dental implant is the subperiosteal, or on-the-bone, which has been used successfully for the past 30 or 40 years [22]. It incorporates a rigid,

plate-like element and is often used for areas without adequate bone for cylindrical endosteal implants. Subperiosteal implants are those that typically lie on top of the jawbone, but underneath the gum tissues. The important distinction is that they usually do not penetrate into the jawbone.

Subperiosteal implants were introduced in the 1940s [3]. Of all currently used devices, it is the type of implant that has had the longest period of clinical application [22]. Unlike endosseous implants, these implants are not anchored inside the bone. Instead they are shaped to ride *on* the residual bony ridge of either the upper or lower jaw. Subperiosteal implants have been used in completely edentulous (toothless) as well as partially edentulous upper and lower jaws. However, the best results have been achieved in treatment of the edentulous lower jaw.

While infection is a potential risk with subperiosteal implants, loosening over time is another problem and must be replaced in the upper jaw after five years. In the lower jaw, these implants last up to 10 years. Replacing these implants also requires a surgical procedure.

11.3.5.3 Transosseous Implants

Transosseous implants similar in definition to endosseous implants in that they are surgically inserted into the jawbone [22]. However, these implants actually penetrate the entire jaw so that they actually emerge opposite the entry site, usually at the bottom of the chin. This is also the site, where they are secured with a device similar to a nut and a pressure plate.

11.3.5.4 The Phenomenon of Osseointegration

In 1952, Dr. Per-Ingvar Brånemark, an innovative orthopedic surgeon, embarked on a series of basic pioneering studies unrelated to implants that could define the conditions necessary for regeneration of highly differentiated tissue after injury [23]. Using cylinders made of titanium to examine the repair mechanisms of bone, Brånemark reported that a direct functional and structural bond had occurred between organized vital bone and the inanimate alloplastic material. He called this process "Osseointegration". He suggested that osseointegration occurred only with the titanium and not with cylinders made of other materials. On the basis of these results, he designed and executed a series of investigations using titanium as a cylindrical threaded endosteal implant for replacing a single tooth or anchoring fixed bridgework [24].

By 1965, Brånemark's implants were first placed in patients. They consisted of a cylindrical screw-type implant made of pure titanium and induced osseointegration between titanium metal and the peri-implant bone. Degradation of the titanium in vivo is minimal because of a protective oxide layer that develops between the titanium implant and the bone. With its provisional acceptance by the American Dental Association in 1986, this innovation has greatly contributed to the acceptance, resulting in widespread use of dental implant treatment by the dental community [25].

Osseointegration is a mode of implant attachment with a bone anchorage, in contrast to a soft tissue-attachment implant. The former interface is rigid, whereas the latter tends to permit macroscopic movements. Various histology based definitions of osseointegration have not reached a general consensus because of difficulties in properly defining the resolution level of the direct bone-to-implant contact and the proportion of a bony contact necessary to call the implant osseointegrated [26, 27]. (Figs. 11.1 – 11.3).

Osseointegrated implants are most stable because they allow bone tissue to grow around them. Osseointegrated implants are made of titanium. Studies indicate that the body's tissues may better tolerate titanium than the chrome-cobalt metal that comprises other implants [28]. Because the bone thickens and grows attached to these implants, they become stronger with time and cause little or no bone re-absorption. Unlike other implants, they feel and function like natural teeth [29, 30].

As better scientific analyses and more clinical experience have been acquired, the definition of osseointegration evolved. A convenient working clinical definition is more useful and the term is currently best defined as a process whereby clinically asymptomatic rigid fixation of alloplastic materials is achieved, and maintained in bone during functional loading [31]. Osseointegration was the hallmark of success in implant dentistry in the 1980s. It was believed that an implant was successfully integrated when there was direct contact between bone and the titanium implant (at the light microscopic level) with no fibrous connective tissue interface.

(a) Root form   (b) Ramusframe Implants   (c) Plate Form Implants
    Implants
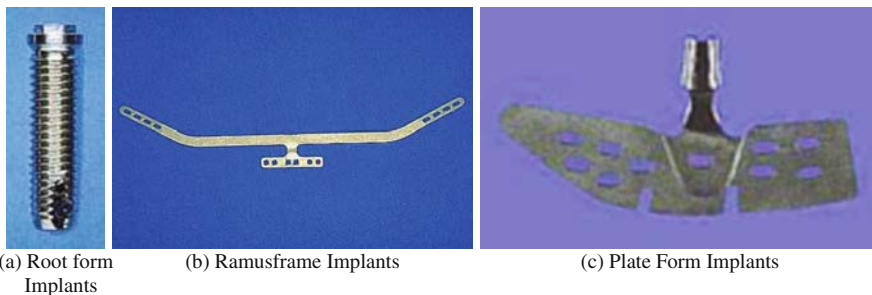
Fig. 11.1a–c Various forms of endosseous implants. **a**. Rootform. **b**. Ramusform. **c**. Plateform [23]

11.3.5.5 Materials Issues in Dental Implants

Ti, Ti—Al—V alloys, Co—Cr—Mo alloys and stainless steel are easy to manufacture and shape, have adequate physical properties, and can integrate with bone. That is why metals are the most common implant materials [13]. The major disadvantages are corrosion, fatigue failure and adverse stress distributions. The material of choice is titanium or Ti—V—Al alloy. Common misconceptions are that implants are made of "pure" Ti and they do not corrode. It is extremely difficult, and economically

Fig. 11.2 Subperiosteal
implant [23]

Fig. 11.3 Transosteal implant
[23]
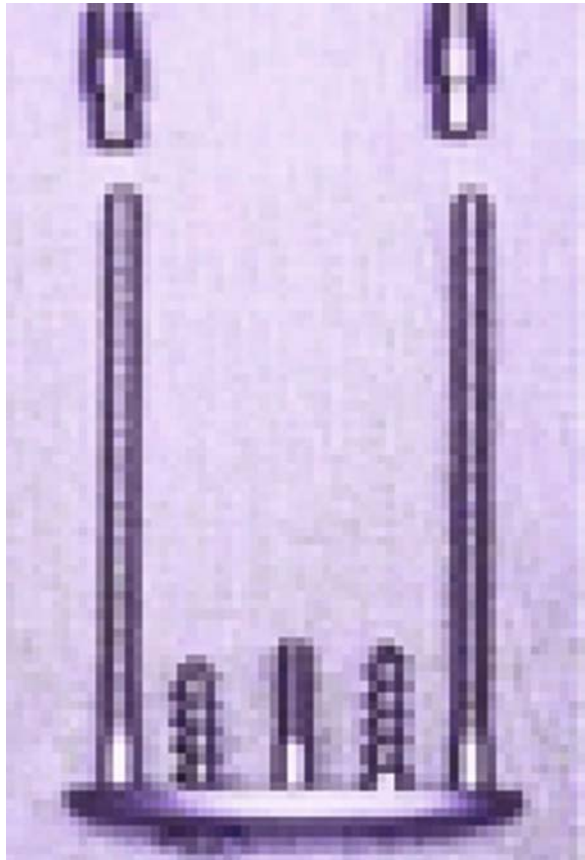
unviable, to remove oxygen from Ti. Thus the most common metal is "commercially pure" Ti. As noted before, the passive layer is typically 15 nm thick and prevents the metal from further corrosion. The passive layer on Ti and its alloys is unique in that it is stable under all normal physiological conditions. Furthermore, this layer is involved in the integration of the implant with bone. The favored alloy is Ti-6% Al-4% V because it has a low density, an elastic modulus half that of cobalt-chrome or stainless steel, has adequate fatigue properties and is biocompatible. Al and V are phase stabilizers, and the alloy can be heat treated to modify its properties. The major disadvantage is that it has a low torsional strength that can result in fracture of screw implants. These alloys are often said to osseointegrate with bone.

11.3.5.6 Surface Issues

Surface features that need consideration to improve material-tissue interaction include material purity, surface chemistry, surface charge, surface morphology, and surface coatings. Data demonstrate that rough implant surfaces have increased bone-to-implant contact. Greater forces to break the bone-implant interface compared to more smooth surfaces. In clinical trials, rough surfaced implants had significantly higher success rates compared to implants with more smooth surfaces. Various surface treatments that will be considered in the following section include etched titanium, titanium plasma spray (TPS), hydroxyapatite-coated (HA-coated) titanium, bioglass coating, and biomimetic surface modifications [44].

Etched Surface – titanium surface is etched with acid. This surface demonstrates increased roughness and therefore greater surface area. Associated with increased dispersion of blood cells around the implant body. Increase in blood-cell dispersion around an implant body has been shown to facilitate osseointegration. SEM and laser analyses have demonstrated the effectiveness of the acid-etching procedure by showing the superior roughness and therefore increased surface area achieved [32, 33].

Titanium Plasma Spray (TPS) – titanium plasma spray (TPS) coating on implants is a pure titanium powder that is applied to the implant using an advanced vacuum plasma-spray process. TPS-coated implants have a relatively large surface area. TPS-coated implants have been shown to possess exceptionally high mechanical retention properties in bone. (Fig. 11.4).

The most common technique to produce porous surface is plasma spraying titanium powder on the implant to create macroporous titanium surface layer [34]. The plasma sprayed surfaces were reported to have advantageous morphological features that make it more suitable to successful clinical applications. However, this technique has several disadvantages. First, the geometry of the interface between the porous coating and the substrate were found to cause stress concentrations that reduced the fatigue strength [35]. Second, high-temperature heat treatments often used to bond the porous coating and substrate was attributed to changes in microstructure and eventual reduction in strength [36]. Lamellar structure, micropores, and microcracks have also been observed inside the plasma sprayed titanium coating [37].

Fig. 11.4a–c Various surfaces of titanium used in dental applications. **a**. Acid etched surface. **b**. TPS coated surface. **c**. HA coated surface [37]

(a) (b) (c)
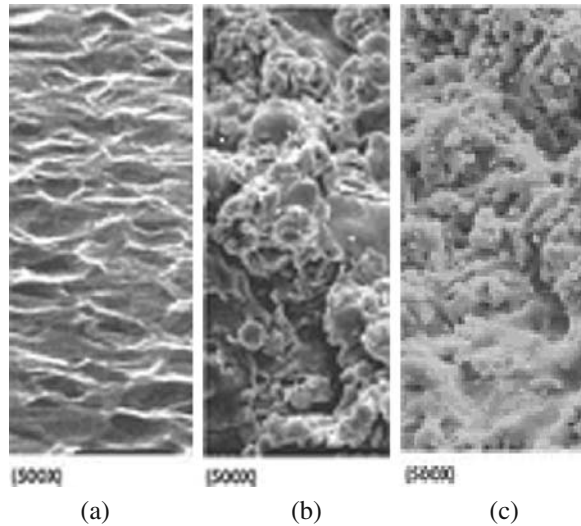
Hydroxyapatite-coated (HA-coated) titanium – studies demonstrated that the osseoconductive properties of HA-coated implants significantly accelerate osseointegration. This ensures a good tissue response and an optimal chemical bond between the coating and the bone. A very favorable combination of purity and crystallinity characteristics and a low dissolution rate. Furthermore, the increased roughness of the HA coating created by a plasma spray process helps lock the implant into the bone.

Excellent short-term clinical results were reported for hydroxyapatite coated implants [38, 39]. HA coated implants have higher integration rate, promote faster bone attachment and achieve direct bone bonding with higher interfacial attachment strength to none when compared to non-coated implant [40]. However, there are many controversies regarding their long-term prognosis. An 8-year old clinical study found that the survival rate was initially higher for HA-coated implants, but decreased significantly below that of titanium plasma-sprayed implants after 4 years [40]. Most of the delayed failures were associated with inflammatory disease. HA delaminating was obtained in both animal experiments and human clinical experience [41].

The resorption of the HA coating compromises long-term implant survival. The coating often contain impurities such as calcium phosphates and CaO, which decreases chemical stability and enhances degradation of the coatings. Delamination of the coatings is due to fatigue and/or thermal expansion coefficient mismatch between Ti/HA interface.

Bioactive Glass Coating – bioactive glass is a bioactive material that promotes the formation of bone in the human physiological environment. Some bioactive glasses also form bonds with soft collagen-based tissues. One barrier of using bioactive glass as coating is the high thermal expansion coefficients ($14–15 \times 10^{-6}/^{\circ}C$).

Tomsia and his colleagues demonstrated that bioactive glass could be tuned to the same thermal expansion coefficient as the Ti alloys it coats [42, 43]. Glasses in the system Si—Ca—Mg—Na—K—P—O were successfully applied on Ti6Al4V using conventional enameling technique. The thermal expansion coefficients obtained were close to that of Ti6Al4V of $9.6 \times 10^{-6}/^{o}C$. The bonding between the glass and coating was achieved by the formation of a Ti_5Si_3 interlayer. Further discussion of bioglass can be found in Sect. 11.4.6.

Biomimetic surface modifications – discoveries by Kokubo and his colleagues indicated that chemically treated Ti and it's alloys are bioactive. After being subjected to 5.0 M-NaOH treatment at 60 °C for 24 h and subsequent heat treatment at 600 °C for 1 h, they are soaked in simulated body fluid (SBF) [44, 45]. Titanium-based alloys such as Ti6Al4V, Ti6Al2Nb and Ti15Mo5Zr3Al formed the bone like apatite layer on the surface in SBF. This bonding accelerates the stabilization of implants and extends its lifetime before failure. The apatite formation on the surfaces of Ti and its alloys was assumed to be induced by a hydrated titania which was formed by ion exchange of the alkali ion in the alkali titanate layer and the hydronium ion in SBF. Through in-vitro studies Kokubo and colleagues showed that the alkali treated Ti provides the most favorable conditions for differentiation of bone marrow cells [46]. (Fig. 11.5).

In-vivo studies conducted in rabbit tibiae to determine the effects of chemical treatments and/or surface-induced bonelike apatite on the bone-bonding ability of titanium implants. Biomechanical examination results showed that treated implants exhibited significantly higher failure loads compared with untreated Ti implants at all time periods. They revealed that treated Ti implants directly bonded to bone tissue during the early post-implantation period, whereas untreated Ti implants formed direct contact with the bone only at 16 weeks [47–50]. The results of their study suggest that chemical treatments accelerated the bone-bonding behavior of Ti implants and enhance the strength of bone-implant bonding by inducing a bioactive surface layer on Ti implants. They concluded that alkali- and heat-treated titanium offers strong bone bonding and a high affinity to bone as opposed to a conventional mechanical interlocking mechanism. Alkali and heat treatments of titanium may be suitable surface treatments for cementless joint replacement implants. These results
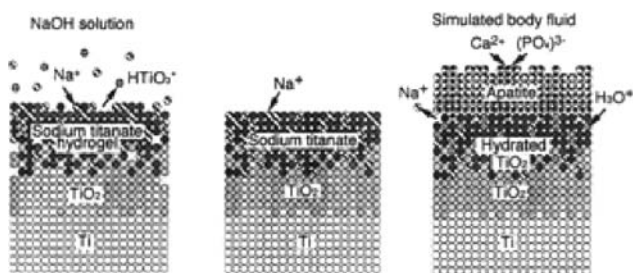
Fig. 11.5 Titanium surface and the formation of apatite as suggested by Kokubo [44]

imply that the Ti—OH groups on titania, proposed to be responsible for the apatite formation, are effective for apatite nucleation when they are arranged in a specific structural unit based on the anatase structure. Alkali- and heat-treated implants also showed direct bonding to bony tissue without intervening fibrous tissue [50]. On the other hand, untreated implants usually had intervening fibrous tissue at the interface between bone and the implant. The early and strong bonding to bone of alkali- and heat-treated titanium and its alloys without intervening fibrous tissue may be useful in establishing cementless stable fixation of orthopedic implants.

11.3.5.7 Problems with Dental Implants

Clinically, lack of osseointegration leads to implant mobility and subsequent failure. Therefore, a mobile implant is a failed implant. Non-functional, failed implants must be removed to prevent the associated bone loss from continuing. For long-term successful performance of all dental implant types the following general factors should be considered. These are biomaterials, biomechanics, dental evaluation, medical evaluation, surgical requirements and healing processes.

Implants fail for a variety of reasons. Some studies have related failures to biological or microbiological factors, while others attribute dental implant failure to biomechanical factors, biomaterial factors, or implant surface treatments and characteristics [32, 33]. Improper patient selection is a significant reason for failure. Obviously, a patient unmotivated to control plaque around natural teeth would not be a good candidate for dental implants. Implants are doomed to fail when placed in patients having insufficient quality and/or quantity of bone to support the implant fixture.

Systemic health of the patient is important when considering implants. Recent studies have demonstrated similar failure rates between well-controlled diabetics and non-diabetic controls or only slightly higher failure rates with Type 2 (non-insulin dependent) diabetics [34]. Uncontrolled diabetic patients are poor candidates for any surgical procedure. Inferior surgical technique is another possible cause of implant failure. During surgical placement, implant failure may result from inadequate irrigation of the surgical site or from using low torque and excessive drill speed during placement. Failure results from excessive temperature elevation in bone during placement, leading to necrosis of the supporting bone around the implant. Inadequate implant restorations also may contribute to implant failure. Restorations placed on endosseous implants may cause traumatic occlusion, leading to failure. Additionally, poorly restored implants may have overhangs or be over-contoured, which may lead to plaque accumulation and eventual failure.

The main factors in implant failure are infection and occlusal force stresses. The microbiota around stable vs failing implants seem to parallel the patterns observed around healthy vs diseased natural periodontal sites. Recent studies indicate that microorganisms associated with periodontal disease are found in higher proportions in failing implant sites. Periodic plaque removal and health maintenance should help prevent such failures.

With single-implant therapy, survival also depends on the position of the implant in the mouth. Some reported on a 24-month life-table analysis study on two-stage implant survival, using a variety of implant designs [32, 33]. They reported survival to be 100% in the anterior mandible, 92% in the posterior mandible, 94% in the anterior maxilla and 78% in the posterior maxilla. The lower survival rate in the latter area is probably the result of the cancellous nature of the bone and thin cortical plates. Small-diameter cylinders appear to do less well than large diameters or blades in hollow, cancellous bone and thin cortical plates. Various implant shapes are needed to overcome bone morphological limitations and meet restorative requirements. Unfortunately, implants for partially edentulous patients are most often needed in the posterior mandible and maxilla, reflecting the pattern of natural tooth loss.

11.4 Ceramics for Dental Applications

As opposed to metals, the strength and toughness values of ceramics are not as good. In spite of that, there is a long history of using ceramics in restorations. The two main reasons for the use of ceramics in dentistry are the ability of restorations to match the esthetics and functions of missing teeth [51–54]. Furthermore, ceramics are benign and do not result in any cytotoxicity.

It is obvious that ceramics can match the opalescent appearance of teeth that no metal can provide. Ceramics can also withstand the cyclic masticating forces that vary between 400 and 800 N. They can also provide the right balance and support between bone/soft tissue. While it is known that ceramics are less mechanically reliable than metals, more and more interesting developments are taking place in ceramics resulting in higher mechanical strength and toughness values [55, 56].

Most common uses of ceramics are inlays and veneers. In these simple applications, strength and fracture toughness values are less important than their esthetic attributes. Ceramics are also used as crowns, a structural replacement of a tooth. These are thin walled structures, which may be composed of metal-ceramic or all-ceramic materials. Some examples are depicted in Fig. 11.6. Like the various applications of metals in dentistry, dental ceramics can be classified according to several schemes, which may include their actual use, composition, processing methods, firing temperatures, microstructure, mechanical, and esthetic properties. Instead of classifying them in various sub-sections, it is simpler to categorize them into metal-ceramic and all-ceramic systems. The metal-ceramic systems are essentially feldspathic porcelains, while all-ceramic systems consist of glass-ceramics, glass infiltrated ceramics, and fully sintered monolithic ceramic materials. This simple classification can be further subdivided according to processing routes, indicated uses, microstructure, esthetic signatures, and mechanical properties. It is also worth mentioning that dental ceramic literature often refers to various commercial products. Therefore, as and when appropriate, these commercial names will be referred to in the text. (Table 11.6).

Table 11.6 Strength and toughness of tooth structures and dental ceramics [57]

Material		Flexural strength (MPa)	Fracture toughness (MPa.$\sqrt{m}$)
Porcelains	Feldspathic	60–110	1.1
	Leucite	120–180	1.2
Glass ceramics	Lab cast/Cerammed	115–125	1.9
	Premade/HIP	140–220	2.0
Glass-infiltrated	In-ceram Alumina®	400–600	3.8–5.0
	In-ceram Spinell®	325–410	2.4
	In-ceram Zirconia®	700–800	6–7
Monolithic ceramics	Lava Y-TZP®	1,000–1,200	9–13
	Procera®	500–600	4.5–5
Tooth structures	Dentine	16–20	2.5
	Enamel	65–75	1

11.4.1 Metal-Ceramic Restorations

The feldspathic porcelains, required for such uses, are based on patents by Weinstein et al. [58] and Weinstein and Weinstein [59]. The patents are about developing compositions with tailored thermal expansion coefficients and the process of their bonding to metallic alloys. The compositions are based in the SiO_2—Al_2O_3—K_2O ternary system. Specifically, SiO_2 and potash feldspar ($K_2O \cdot Al_2O_3 \cdot 6SiO_2$) are the basis for dental porcelains. Other additives such as opacifiers (TiO_2, ZrO_2, and SnO_2), appropriate pigments, fluxes (to lower melting point) are also added. The porcelains are normally classified in terms of their firing temperatures: high firing ($1,300\,°C$), medium firing ($1,100$–$1,300\,°C$), low firing (850–$1,100\,°C$), and ultra-low firing ($<850\,°C$).

The most important aspect of the development of feldspathic porcelain is the presence of leucite ($KAlSi_2O_6$). The presence of leucite increases the thermal expansion coefficient of porcelains to an extent that bonding between the porcelain and metal substrate can easily take place with matched thermal expansion coefficients. While it is easy to say that leucite can increase the thermal expansion coefficients of porcelains, it is not an equilibrium phase. Different thermal processing histories result in different cooling rates of leucite with concomitant variation in thermal expansion coefficients [60]. While slow cooling in a furnace increases the leucite content from 11 vol% to 56 vol% [61], an isothermal hold at $750\,°C$ 4–16 min (simulating post-soldering treatment), the leucite content in commercial porcelain increases from 6 vol% to 21 vol% [62]. The various thermal histories can also lead to cracking of porcelains due to the thermal expansion coefficient mismatch [63]. Microcracking can also happen due to the cubic-to-tetragonal transformation of leucite upon cooling [64].

The flexural strengths of feldspathic porcelains are quite low (between 60 and 70 MPa). However, when bonded with the metal substructure, the flexural strength increases significantly to approximately 300 MPa [12].

Fig. 11.6a–d Types of dental restorations. **a**. In-Ceram crown with standard crown design. **b**. Lateral view of In-Ceram crown demonstrating underlying tooth structures. **c**. Lateral view of ceramic crown on a titanium implant. **d**. Porcelain veneer resin bonded to tooth enamel

a.

b.

c.

Enamel
Dentine
Pulp
Core Ceramics
Venner Ceramic

d.
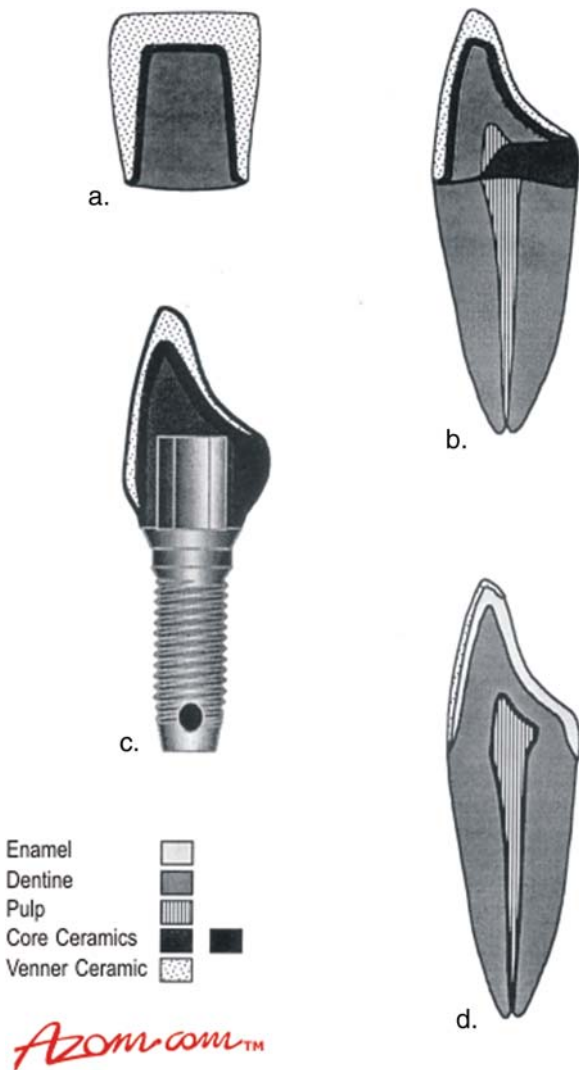
11.4.2 All-Ceramic Restorations

Glass-ceramics – these materials are composites of glass and ceramics, which can be processed into complex shapes using glass-shaping technology. These materials are, therefore, also referred to as castable/pressable glass-ceramics. The ceramic particulates are formed in a glassy matrix by carefully nucleating and growing ceramic crystals with suitable heat-treatments (also known as ceramming). The initial development of Dicor® glass ceramics took place in Corning Glass Works and was marketed by Dentsply International. Dicor® consists of a magnesium aluminosilicate

glass with 45% fluoromica ($K_2Mg_5Si_8O_{20}F_4$) crystals. The presence of these crystals enhances the mechanical properties of these glass-ceramic materials. In spite of the advantages of enhanced processibility and the potential for enhanced mechanical properties, the extent of mechanical property enhancement was not substantial. Furthermore, Dicor[®] materials do not have the ability to attain internal color. Even though, esthetically, they are better than metal-ceramics, with the development of other glass ceramics, Dicor[®] materials are no longer available on the market.

Another important glass ceramic is Empress 1[®], which is essentially a leucite reinforced feldspathic porcelain. The controlled nucleation of leucite enhances the mechanical properties while being produced by the conventional investment casting technique. Empress 1[®] has flexural strength values about the same as Dicor[®]. The translucency of this material is slightly less than Dicor[®] but can be tolerated in restorations.

A major advancement in glass ceramic development is the Empress 2[®], which is a lithium silicate based system. With proper composition choice and heat treatment, the microstructure consists of 70% of elongated lithium-disilicate crystals [65]. This creates a highly filled microstructure. The manufacturer recommends these be used for inlays, onlays, veneers, etc.

Monolithic ceramics – an example of this type of material is fully sintered alumina (Procera Allceram[®]). This material is 99.9% alumina and sintered at $1,600\,^\circ$C to produce a core, which is compatible with dental porcelain. While the mechanical properties are quite good, the sintering process results in a large shrinkage. Therefore, in order to obtain high dimensional tolerance, CAD-CAM system is used. Once the sintered coping is obtained, a matched porcelain is placed on the coping. Procera Allceram[®] is limited to use as a single unit restoration in the anterior or posterior regions.

The most significant advancement in monolithic ceramics is the introduction of toughened zirconias, specifically yttria-tetragonal zirconia polycrystals (Y-TZPs) [55, 56]. Doping zirconia with yttria results in a metastable tetragonal phase in the microstructure. During the propagation of a crack, there is martensitic transformation of tetragonal to monoclinic phase, which dissipates energy from the propagating crack. This increases the strength as well as toughness of monolithic zirconia twice that of monolithic alumina. While there are several dopants such as CaO, MgO and CeO_2 that are available for this purpose, yttria doping can produce the most esthetically pleasing restoration. CeO_2 doping can produce even better mechanical properties, but it has its own characteristic color and MgO doping needs very high temperatures of processing. Another interesting aspect of Y-TZP is that the mechanical properties are so good that they can be machined in Lava CAD/CAM system to produce complex bridges.

Glass infiltrated ceramics – these materials are based on partially sintered alumina, zirconia, and spinel. These materials are commercially known as In-Ceram Spinell[®], In-Ceram Alumina[®], and In-Ceram Zirconia[®]. They have surface pores, which are infiltrated with lanthanum aluminosilicate glass. Lanthanum additions decrease the viscosity of silicate melts, which assists in infiltration and simultaneously improving the strength of the ceramic core. The mechanical properties of

In-Ceram® materials are better than all other restorations other than Y-TZPs. From the esthetics point of view, In-Ceram® materials are opaque and they can be tinted easily.

11.4.3 Processing of All-Ceramic Restorations

A crown with the normal size and shape of a tooth, is very useful, when a tooth is broken down to a limit when filling no longer solves the problem. Traditional crowns have a porcelain-bonded enclosure filled with gold. They are sturdy enough to withstand about 200 pounds of pressure during use. Currently, a second generation materials are being used which are stronger and more durable than the original porcelain type materials with better wear, and more translucency to match the natural color of the teeth. All-ceramic restorations (e.g., crowns, inlays, onlays and veneers) are becoming quite popular when esthetics is important. A bridge is a complex version of a crown replacing one or more natural missing teeth, thereby "bridging" the space between two teeth. Bridges are cemented into place on the "abutment" teeth – the surrounding teeth on either side of the space or span. Unlike removable partial dentures, bridges cannot be taken out of the mouth by the patient. There are several criteria for selecting restorative systems and processes. Esthetics, strength, marginal fit, biocompatibility, cost and ease of fabrication, and patient comfort are among these criteria. A process and a relevant material should be considered most versatile when the combination can successfully produce a bridge.

   Most of the dental prosthetics that are produced manually require considerable technical skills and consume a substantial amount of time. Machining has become a viable option for most of the dental restorations [66, 67]. Computer-assisted design/computer assisted manufacturing (CAD/CAM) has become a viable option for making all ceramic restorations. Of late, CAD/CAM is referred to as the subtractive mode of rapid-prototyping technology whereby blocks of materials are machined to precise shapes. Commercially available systems include CEREC (Sirona), Procera (Nobel Biocare) and LAVA (3 M ESPE). CEREC mechanics are well suited for machining glass-ceramics, In-ceram alumina, spinel, and zirconia [68]. While LAVA machines can also machine Y-TZP materials. The prepared teeth are coated with an imaging liquid, which can hold imaging powder such as TiO_2 on the teeth. Using an infrared camera, an optical impression is captured and suitably modified for the proper fit. The data are transferred to a milling machine to mill a block of ceramic. After milling, the sprue is removed from the restoration. The main advantage of these machines is the elimination of a temporary restoration, resulting in a significant reduction of time and cost. They also reduce the human error in processing. The difference between the CEREC and LAVA machines is that the CEREC machines can fabricate one crown while the LAVA machine can produce bridges. This can be achieved because of the use of Y-TZP of high strength and toughness. An alternative to the LAVA system is the Cercon® Zirconia system, which requires

sending an impression to a laboratory for machining a pre-sintered porous Y-TZP. Once the machining is carried out, the restoration is sintered to full density taking into account the shrinkage.

Use of all ceramic materials requires a tooth preparation with a rounded shoulder or chamfer finish line. The amount of tooth reduction varies from system to system [69]. For example, the manufacturer of Empress 2® recommends 0.8 mm, while Procera® and In-Ceram® recommend 0.5 mm. Cementation with composite resin cements improves the success rate of Empress 1 and 2. Glassionomer cements are popular because they can be expanded and create stress, which will eventually lead to failure. Hybrid cements such as copolymers (polyacid modified resin cements and the resin-modified glass-ionomer cements (GIC) were developed as smart filling materials. However, biocompatibility remains an issue [70].

11.4.4 Selection Guide for All-Ceramic Restorations

In selecting materials, higher strength and fracture toughness should be used for canine and posterior regions of mouth. Besides, Weibull modulus is an important factors for successful restorations. If the scattering of tensile strength is high, Weibull modulus is low; therefore reliability might be a concern.

The strength of castable leucite reinforced glass ceramics is around 130 MPa and they are best suited for anterior crowns, inlays, onlays, and veneers. In-Ceram zirconia has a significantly higher strength and fracture toughness and can be used for anterior and posterior single unit crowns and anterior bridges. Lava form® (Y-TZP) has the highest strength and toughness. Table 11.7 shows selection guide for the use of all-ceramic restorations.

Table 11.7 Selection guide for all-ceramic restorations

	Inlays, onlays, veneers	Anterior crown	Posterior crowns	Anterior bridges	Posterior bridges	Translucency
Empress 1® leucite/field based castables	×	×	–	–	–	1
Empress 2®	×	×	×	×	?	2
Procera	×	×	×	?	?	3
In-Ceram alumina®	×	×	×	×	–	3
In-Ceram spinel®	×	×	–	–	–	2
In-Ceram zirconia®	×	×	×	×	×	4
Lava form®	×	×	×	×	×	2

''×'' means allowed use.
''–'' means not possible.
''?'' means yet to be determined.

11.4.5 Clinical Failure of All-Ceramic Crowns

Despite the wide variety of all ceramic restorative materials available today, they are still not highly recommended by dentist practitioner [70]. Usually cementation or internal surface flaws initiate most of the clinical failures. As an example, Dicor ceramic crowns luted with Zn-phosphate cement were reported to have poorer success rates than crowns cemented and bonded with composite resin cements. Internal surfaces have the highest tensile stresses and critical flaws. Etching and polymer coating of tensile surfaces can strengthen the strength of these ceramic materials by blunting the crack tip or by lessening the stress corrosion by barrier coatings [71–73].

11.4.6 Bioactive Glasses

Bioactive glasses are widely used in musculoskeletal applications, a smaller subgroup of which is dentistry. One of the important systems of bioactive glasses is derived from Na_2O—CaO—P_2O_5—SiO_2 ternary system with other substitutions. The system is commercially known as Bioglass®, which was developed by Hench et al. at the University of Florida [74, 75]. An important feature of these glasses is a lower SiO_2 content as opposed to the traditional soda-lime silica glass and much higher content of Na_2O and CaO. These compositions make the surfaces highly reactive when exposed to the body fluid. The most commercially important composition is known as 45S5 containing 45 mol% SiO_2.

Tomsia and colleagues modified the original compositions so that they are compatible to metallic implants in terms of thermal expansion coefficients [76–79]. Specifically, Na_2O and CaO were partially substituted by K_2O and MgO respectively. This results in thermal expansion coefficients compatible to implants. The processing of such glasses can be performed by the traditional glass melting process or by the sol-gel techniques.

As mentioned before, monolithic net-shaped implants have been in use in dentistry. Specifically, Endosseous Ridge Maintenance Implant (ERMI) was developed by Stanley et al [80]. Their 10-year survival rate is greater than those implants that do not osseointegrate. Bioglass® particulates have been placed around teeth with periodontal disease [81]. Osseointegration takes place at a rapid rate around those glass particles. This helps stabilizing the boundary between tooth and the periodontal membrane and the tooth is saved.

The other related material is actually a glass ceramic based on the CaO—P_2O_5—SiO_2 system. This material consists of crystallites of apatite $Ca_{10}(PO_4)_6$ (OHF_2) and β-wollastonite (CaO—SiO_2) with a residual CaO—SiO_2 glassy phase. These materials are referred to as A/W glass ceramics and were mentioned by Yamamuro and Kokubo [82–84]. The presence of crystallites enhances the mechanical properties (e.g., fracture toughness, strength as well as subcritical crack growth velocity). The mechanism of bioactivity of such glass ceramics is similar to the Bioglass®. By the same token, these can also be used in the cases where Bioglass® is used.

11.5 Polymers for Dental Applications

Polymeric materials with improved durability find important applications as dental materials. Denture bases and artificial teeth are two major areas where polymeric materials are used. Besides these, they are also being used in various applications such as crown and bridge facings, inlay patterns, implants, impressions, dies, temporary crowns, endodontic fillings, athletic mouth protectors, and orthodontic space maintainers etc [85].

11.5.1 Dentures

Dentures are usually used if an entire arch of teeth is missing. If there are a few teeth left, partial denture is needed. Generally, dentures have an acrylic base molding. The quality of the bonding of denture base polymers to synthetic polymer teeth is dependent on the temperature during processing. About 98% of the denture materials are constructed from methyl methacrylate polymers or copolymers. Other polymers include vinyl acrylic, polystyrene, epoxy, nylon, vinyl styrene, polycarbonate, polyurethane, silicones, rubber reinforced acrylics and butadiene reinforced acrylics, etc. Systems which are based on dimethacrylate monomers are commonly used for crowns and bridges, and now also for dentures.

The most important polymer system used for prosthodontic applications is based on polymethyl methacrylate (PMMA) powder and a monomer mixture (liquid) of methyl methacrylate and a cross-linking dimethacrylate monomer such as ethylene glycol dimethacrylate (EGDMA). The process of polymerization is crucial for the properties of the prosthodontic devices. The mechanical properties including craze resistance are dependent on the processing procedure and on the quality of cross-linking agent. At an ambient temperature, the material is brittle and transparent. At temperatures above the transition temperature, the material tends to soften and becomes viscoelastic.

The presence of surface cracks/pores and internal faults usually affect the tensile strength of the material. The failure of the denture material involves either impact or fatigue. The impact failure is due to rapid stress of the material and fatigue failure is due to cyclic flexing of the material generating cracks.

11.5.2 Dental Cements

Dental cements are used for fixing dental restorations such as crowns, orthodontic bands, serving as bases under any restorations, etc. They must exhibit sufficient viscosity to flow through the interfaces between hard tissue and the restoration [12]. Properties that are important include strength and stiffness, resistance to dissolution, and biocompatibility with pulpal tissue [85]. Further, they should be set within 25–50 μm thickness, and should release fluoride to prevent caries. The five most popular

systems include zinc phosphate, zinc oxide eugenol (ZOE), polycarboxylate (with zinc powder), glass-ionomer, and compomer cements.

Zinc phosphate and ZOE – zinc phosphate serves as an universal luting agent, which means that it has the right rheological properties to flow between two surfaces with the ability to solidify very fast [12]. It has good handling characteristics, and a proven longevity in the oral cavity when used in conjunction with well designed and well fitted restorations. The powder constituents for both zinc phosphate and ZOE are zinc oxide with minor additions of magnesium oxide. The main liquid component is either phosphoric acid or eugenol oil. Zinc phosphate is typically used for permanent cementation. The accepted mechanism for cementation is the initial attack of oxides by phosphoric acid, resulting in an aluminophosphate gel. The problem with zinc phosphate cement is that it is highly acidic and it may cause damage to the pulp. On the other hand, the advantage of ZOE is its biocompatibility, eugenol having a pH of 7 and hence benign to the pulp. Zinc oxide is hydrolyzed initially and undergoes chelation subsequently. ZOE can be used for either permanent or temporary fixation. There are several variations in compositions of ZOE based on their service demands. These cements are used as a base where a significant amount of dentin has been removed and pulp tissues are approximately 0.5–2 mm. Rather than placing a restoration near pulp, cements are used to protect the pulp.

Polycarboxylic acid cements – in polycarboxylic cements, the liquid component is polyacrylic acid, which is then mixed with zinc oxide [86]. Bonding to dentin can occur via chelation of carboxyl groups to calcium ions. The advantages include shorter gelling times (about half of zinc phosphates) and better bonding strength with the tooth structure (with respect to other zinc phosphates) [87].

Glass-ionomer cements – these use silicate glass powder instead of zinc oxide powder and an aqueous solution of polyacrylic acid (or a copolymer of itaconic acid, or maleic acid) [88, 89]. They bond to tooth structure well and release fluoride. Typical composites of glass particles are 20–36 wt% SiO_2, 15–40 wt% Al_2O_3, 0–40 wt% CaO, 0–35 wt% CaF_2, 0–5 wt% Na_3AlF_6, 0–6 wt% AlF_3, 0–10 wt% AlP. When glass particles are mixed with polyacrylic acid, hydrogen ions are removed from carboxyl groups of acid solution. Simultaneously glass particles undergo reaction, where Al^{3+}, Ca^{2+}, Na^+, F^-, and PO_4^{3-} are released into solution to form a silica rich gel on the particle surfaces. In order to use glass-ionomer cements, the dentin and enamel are acid etched with polyacrylic acid, then rinsed. Bonding to dentin occurs both by ionic and micromechanical forces. Glass-ionomers have the advantage of chemically bonding directly to enamel and dentin. There is an ionic bond formed between the calcium of the tooth and the set material. As enamel is richer in calcium, the bond to enamel is stronger than the bond to dentin. The bond strength of glass ionomer is much lower than that of bonding agent/composite systems, but in some situations, high bond strength is not the priority. In terms of toughness, they cannot withstand high stress concentrations that promote crack propagation. Besides, slow maturing contributes to low initial strength. Sometimes, glass-ionomers are reinforced to "cermets" (silver or silver amalgam powder mixed into original powder or pure metallic silver fused to glass powder) for increased strength, more radio opacity and increased abrasion resistance) [90]. Another type

of glass-ionomer cement is resin modified (hybrid ionomer), which overcomes the problem associated with low initial strength. Polymerizable functional groups can be added to impart rapid curing when activated by light or chemicals. This overcomes the problem associated with low early strength and moisture sensitivity that occurs during slow acid base setting reactions.

Compomer – compomers have the fluoride releasing ability of glass-ionomers and the durability of composites. These are typically one component systems consisting of fillers of sodium fluoride and polyacid modified monomers. The setting reaction is triggered by photopolymerization. The mixture does not contain water and hence is not self-adhesive like the glass ionomer cements. A separate dentin bonding agent is therefore needed.

Resin cements – the resin cements are used for attaching orthodontic brackets and fixed prostheses to tooth structure following the application of either enamel or dentin bonding agent. These cements are insoluble in any oral fluids. Although the resin cements have lower elastic modulus compared with zinc phosphate, fracture toughness is usually superior. Resin cements bond well with the dentin and make a strong attachment with the enamel.

11.5.3 Composite Dental Materials

Dental composites are becoming popular because of esthetic reasons and are used to rebuild or repair the teeth. The compositions consist of three parts – a resin, filler particles, and a coupling agent (typically a silane to bond filler with the resin). Composite resins have good esthetic properties and strength and are used in front teeth. The resin that is of key importance in dental composites is an aromatic dimethacrylate monomer, bisphenol A-glycidyl methacrylate (BIS-GMA) [91]. Polymer shrinkage, thermal expansion, water sorption, tensile strength, wear resistance, and modulus of elasticity are some of the properties that can affect clinical performance. In order to enhance further the longevity, filler particles such as quartz, colloidal silica, or silicate glasses containing Sr or Ba are incorporated into the resin matrix. These tooth-colored filling materials have reached a high degree of sophistication since their appearance on the dental scene in the early 1960s. A modern dental composite consists of a paste created by combining a mixture of dimethacrylate monomers and cross-linking agents with up to 80 wt% of silane-coated, ceramic particles (the filler), whose sizes range from 0.04 to 4 μm. This composite paste is packed into a dental cavity and the dentist exposes it for about 30s to intense visible blue light. The light activates a chemical initiator (camphorquinone, CQ) within the composite and the resins undergo free radical addition polymerization via their vinyl groups, turning the paste into a durable, solid filling. Fillers are placed in dental composites to reduce shrinkage upon curing. Physical properties of composite are improved by fillers, however, composite characteristics change based on filler material, surface, size, load, shape, surface modifiers, optical index, filler load and size distribution. Accordingly, the composites are

classified as traditional (containing large fillers). Hybrids (containing different particle sizes), flowable hybrids (containing different particle size with lower volume fraction of fillers), and homogeneous microfill (containing ultrafine/nanocrystalline fillers). Fillers less than 1 μm do not produce a rough appearing surface with aging. The nano particles fill between all other particles to further reduce shrinkage.

Composite fillings are less durable than amalgam for the large filling size. However, it is comparable for small to average size fillings. Usually composites filling are quite difficult to place especially for a back tooth, and therefore they are more expensive than amalgam filling. Composites are used in areas where the fillings are relatively small and there is plenty of tooth structure for support. They have similar strengths to amalgam but they tend to wear away more rapidly. Composite fillings are more natural looking, require less tooth reduction to place, and are bonded in place for a better seal. In terms of biocompatibility, composites are not totally biocompatible. Sometimes iron oxide, aluminum oxide, barium, and other materials leach out.

Polymerization shrinkage is one of the major disadvantages of composite materials. As the resin matrix polymerizes, the organized polymer molecules occupies less than its disorganized molecules. During cures, it shrinks considerably. The degree of shrinkage is determined by the percentage of filler in the composite (most hybrids are 60–65% filler by volume).

11.6 Closure

Current dental restorative materials can be used effectively for restoring teeth for functional or esthetic reasons. Diverse materials such as metals, ceramics, polymers, and composites are routinely used. Important properties include reliable mechanical properties, corrosion/wear resistance, biocompatibility, and physical appearance. No single material has been found that meets all the requirements. Furthermore, virtually all restorative materials have components with potential health risks. However, there is no scientific evidence that currently used restorative materials cause significant side effects. Available data do not justify discontinuing the use of any currently available dental restorative materials or recommending their replacement. The research is going along many different avenues. It is hoped that in the future, there will be substantial volume of research in this field.

References

1. http://www.ncbi.nlm.nih.gov
2. Hoffman-Axthelm W. History of Dentistry, Quintessence Publishing: Chicago, 1981.
3. Craig RG, O'Brien WJ, and Powers JM. Dental Materials: Properties and Manipulation, Mosby Company: St. Louis, MO, 1975.
4. Holt R. J Western Soc Periodont, 1986, 34: 49.
5. O'Brien WJ. Dental Materials: Properties and Selection, Quintessence Publishing: Chicago, IL, 1989.

6. Marshall SJ and Marshall GW. Adv Dent Res 1992, 6: 94.
7. Okabe T, Mitchell RJ, Wright AH, and Fairhurst CW. J Dent Res, 1977, 56: 79.
8. Okamoto Y and Horibe T. Br Dent J, 1991, 170: 23.
9. Mahler DB, Adey JD, Simms LE, and Marek M. Dental Mater, 2002, 18: 407.
10. Craig RG and Powers M. Restorative Dental Material, 11th edn, Mosby Inc: New York, 2002.
11. Taggart WH. Dent Cosmos, 1907, 49: 1117.
12. Phillips' Science of Dental Materials K. J. Anusavice, Sauder: St. Louis, MO, 2003.
13. Smith WF. Structure and Properties of Engineering Alloys, Mc Graw Hill: New York, 1981.
14. Buehler WJ and Wang FE. Ocean Eng, 1968, 1: 105.
15. Funakubo H. Shape Memory Alloys, Gordon and Breach Science Publishers: Reading, UK, 1984.
16. Burnstone CJ and Goldberg AJ. Am J Orthodont, 1980, 77: 121.
17. Nomura T, Shingaki S, and Nakajima T. J Long Term Eff Med Implants, 1998, 8(3–4): 175.
18. Brunski JB, Puleo DA, and Nanci A. Int J Oral Maxillofac Implants, 2000, 15: 15.
19. Heimke G and Wittal CG. Osseointegrated dental implants, in Wise DL et al., eds. Biomaterials Engineering and Devices: Human Applications, Vol. 2, Humana Press Inc.: Totowa, NJ, 1990, p. 65.
20. Smith DE and Zarb GA. J Prosthet Dent, 1989, 62: 567.
21. Sinha RK, Morris F, Shah SA, and Tuan RS. Clin Orthoped Related Res, 1994, 305: 258.
22. Sawase T, Hai K, Yoshida K, Baba K, Hatada R, and Atsuta M. J Dent, 1998, 26: 119.
23. Ratner BD, Hoffman AS, Schoen FJ, and Lemons JE, eds. Biomaterials Science, an Introduction to Materials in Medicine, Academic Press: San Diego, CA, 1996.
24. Albrektsson T and Zard GA. Int J Prosthodont, 1993, 6: 95.
25. Albrektsson T and Lekholm U. Dent Clin North Am, 1989, 33: 537.
26. Branemark P-I, Breine U, Lindstrom J, Adell R, and Ohlsson P. Scand J Plast Reconstr Surg, 1989, 3: 81.
27. Albrektsson T, Branemark P-I, Hansson HA, and Lindstrom J. Acta Orthop Scand, 1981, 52: 155.
28. Fugazzotto PA. Int J Oral Maxillofac Implants, 1997, 12: 17.
29. Swase T, Hai K, Yoshida K, Baba K, Hatada R, and Atsuta M. J Dent, 1998, 26: 119.
30. Steflik DE, Corpe S, Lake FT, Sisk AL, Parr GR, Hanes PJ, and Buttle K. Int J Oral Maxillofac Implants, 1997, 12: 443.
31. Zarb GA and Albrektsson T. Int J Periodont Rest Dent, 1991, 11: 88.
32. Shirkhanzadeh J. J Mater Sci Mater Med, 1998, 9: 67.
33. Shirkhanzadeh J. Mater Sci Lett, 1991, 10: 1415.
34. Yang YZ, Tian JM, Tian JT, Chen ZQ, Deng XJ, and Zhang DH. J Biomed Mater Res, 2000, 52: 333.
35. De Santis D, Guerriero C, Nocini PF, Ungersbock A, Richards G, Gotte P, and Armato U. J Mater Sci Mater Med, 1996, 7: 21.
36. Kohn DH and Ducheyne P. J Biomed Mater Res, 1990, 24: 1483.
37. Chang CK, Wu JS, Mao DL, and Ding CX. J Biomed Mater Res, 2001, 56: 17.
38. Block MS and Kent JN. Dent Clin North Am, 1992, 36: 27.
39. Guttenberg SA. Compend Contin Educ Dent, 1993, 14: 549.
40. Wheeler SL. Int J Oral Maxillofac Implants, 1996, 11: 340.
41. Biesbrock R and Edgerton M. Int J Oral Maxillofac Implants, 1995, 10: 712.
42. Tomsia AP, Moya JS, Carter WC, French R, Kerans R, Morinaga K, Niwa K, and Thomas G. Acta Mater, 2000, 48: 400.
43. Tomsia AP, Bloyer DR, Gomez-Vega JM, Saiz E, McNaney JM, and Cannon RM. Acta Mater, 1999, 47: 4221.
44. Kokubo T. Acta Mater, 1998, 46: 2519.
45. Takadama H, Kim H-M, Kokubo T, and Nakamura T. J Biomed Mater Res, 2001, 57: 441.
46. Nishio K, Neo M, Akiyama H, Nishiguchi S, Kim H-M, Kokubo T, and Nakamura T. J Biomed Mater Res, 2000, 52: 652.

47. Nishiguchi S, Fujibayashi S, Nakamura T, Kim H-M, Uchida M, and Kokubo T. J Biomed Mater Res, 2003, 67A: 26.
48. Fujibayashi S, Neo M, Nakamura T, Kim H-M, Uchida M, and Kokubo T. Biomaterials, 2003, 24: 3445.
49. Nishiguchi S, Kato H, Fujita H, Oka M, Kim H-M, Kokubo T, and Nakamura T. Biomaterials, 2001, 22: 2525.
50. Nishiguchi S, Kato H, Neo M, Fujita H, Oka M, Kim H-M, Kokubo T, and Nakamura T. J Biomed Mater Res, 2001, 54: 198.
51. Denry IL. Crit Rev Oral Biol Med, 1996, 7: 134.
52. Kelly JR, Nishimura I, and Campbell S. J Prosthet Dent, 1996, 75: 18.
53. Kelly JR. Annu Rev Mater Sci, 1997, 27: 443.
54. Kelly JR. Dent Clin North Am, 2004, 48: 513.
55. Heuer AH and Hobbs LW, eds. Science and Technology of Zirconia, The American Ceramic Society: Westerville, OH, 1981.
56. Meriani S and Palmonari C, eds. Zirconia '88 Advances in Zirconia Science and Technology, Springer Verlag: Berlin, Germany, 1989.
57. http://www.azom.com/details.asp?ArticleID=1682
58. Weinstein M, Katz S, and Weinstein AB. U.S. Patent 3,052,982, 1962.
59. Weinstein M and Weinstein AB. U.S. Patent 3,052,983, 1962.
60. Mackert JR, Butts MB, Morena R, and Fairhurst CW. J Am Ceram Soc, 1993, 69: C-69.
61. Mackert JR and Evans AL. J Dent Res, 1991, 70: 137.
62. Mackert JR, Twiggs SW, and Evans AL. J Dent Res, 1995, 74: 1259.
63. Anusavice KJ and Gray AE. Dent Mater, 1989, 5: 58.
64. Mackert JR, Twiggs SW, Russell CM, and Williams AL. J Dent Res, 2001, 80: 1574.
65. Holland W, Schweiger M, Frank M, and Rheinberger VA. J Biomed Mater Res, 2000, 53: 297.
66. Kelly JR, Luthy H, Gougoulakis A, Pober R, and Mormann WH. Proc Int Symp Computer Restorations, 1991, 253.
67. Estafan D, David A, David S, and Calamia J. Compendium, 1999, 20: 555.
68. Sohmura T and Takahashi J. Int J Prosthodontics, 1995, 8: 252.
69. Giordano R. Gen. Dentistry, 2000, 48: 38
70. Pospiech P. Clin Oral Invest, 2002, 6: 189.
71. Kelly JR, Campbell S, and Bowen HK. J Prosthet Dent, 1989, 62: 536.
72. Kelly JR, Giordano R, Poder R, Cima MJ. Int J Prosthodont, 1990, 3: 430.
73. Albakry M, Guazzato M and Swain MV. J Dentistry, 2003, 31: 181.
74. Hench LL. J Am Ceram Soc, 1998, 81: 1705.
75. Hench LL and Wilson J. An Introduction to Bioceramics, World Scientific: London, UK, 1993.
76. Gomez-Vega JM, Hozumi A, Saiz E, Tomsia AP, Sugimura H, and Takai O. J Biomed Mater Res, 2001, 56: 382.
77. Oku T, Suganuma K, Wallenberg LR, Tomsia AP, Gomez-Vega JM, and Saiz E. J Mater Sci Mater Med, 2001, 12.
78. Saiz E, Goldman M, Gomez-Vega JM, Tomsia AP, Marshall GW, and Marshall SJ. Biomaterials, 2002, 23: 3749.
79. Lopez-Esteban S, Saiz E, Fujino S, Oku T, Suganuma K, and Tomsia AP. J Eur Ceram Soc, 2003, 23: 2921.
80. Stanley HR, Ball MB, Clark AE, King JC, Hench LL, and Berte JJ. Int J Oral Maxillofac Implants, 1997, 12: 95.
81. Fetner E, Hartigan MS, and Low SB. Compend Contin Educ Dent, 1994, 15: 932.
82. Kokubo T, Ito S, Sakka S, and Yamamuro T. J Mater Sci, 1986, 21: 536.
83. Nakamura T, Yamamura T, Higashi S, Kokubo T, and Ito S. J Biomed Mater Res, 1985, 19: 685.
84. Yamamuro T, Hench LL, and Wilson J, eds. Handbook on Bioactive Ceramics: Bioactive Glasses and Glass-Ceramics, Vol. 1, CRC Press: Boca Raton, FL, 1990.

85. Davis JR and Associates, eds. Handbook of Materials for Medical Devices, ASM International: Materials Park, OH, 2003.
86. Smith DC. J Can Dent Assoc, 1971, 37: 22.
87. Wilson AD and Prosser HJ. Br Dent J, 1984, 157: 449.
88. Wilson AD and McClean JW. Glass Ionomer Cements, Quintessence: Chicago, IL, 1988.
89. Davidson CL and Mjor JA. Advances in Glass Ionomer Cements, Quintessence: Chicago, IL, 1999.
90. Goldman M. J Biomed Mater Res, 1985, 53: 771.
91. Bowen RL. U.S. Patent 3,006,112, 1962.

Chapter 12
Ophthalmic Biomaterials

Rachel L. Williams and David Wong

12.1 Introduction

Ophthalmology is a branch of medicine that has seen many technological advances in recent years. Lasers are used for imaging and treating many eye conditions and their capabilities are further enhanced by computers and culminate in the integration of devices with neuronal elements to produce "artificial vision." In virtually all endeavors to develop therapies involving eye surgery, biomaterials play a central part [6, 15] (Fig. 12.1). Contact lenses are worn on the corneal surface to improve visual acuity in place of spectacles. Corneal implants, or keratoprostheses have been designed to replace diseased or damaged cornea that have become opaque. Intraocular lenses are implanted following cataract surgery to replace the opaque crystalline lens. Viscoelastic substances have revolutionized cataract surgery enabling the first truly "key-hole" operations. Glaucoma filtration devices are use to produce a channel for the outflow of aqueous to prevent the increase in intraocular pressure. Tamponade agents are used to replace the vitreous and treat retinal detachments. Retinal implants are designed to transmit electrical signals to the brain via the retina and the optic nerve.

Biomaterials are exploited for their various physical properties to achieve the goals of surgery. The purpose of this chapter is not so much to enumerate the materials used, but to illustrate how treatment possibilities are being realized through an understanding of biological requirements and how these can be met by the appropriate choice of biomaterial. There are a number of essential requirements that need to be addressed by ophthalmic biomaterials which include the ability to deliver oxygen to the tissues, modification of refraction, protection of tissues during surgery, integration with the tissue, modulation of wound healing and tamponading of retinal breaks. This chapter will address each of these in turn and will discuss the disease processes, the requirements and the biomaterials solutions under each category.
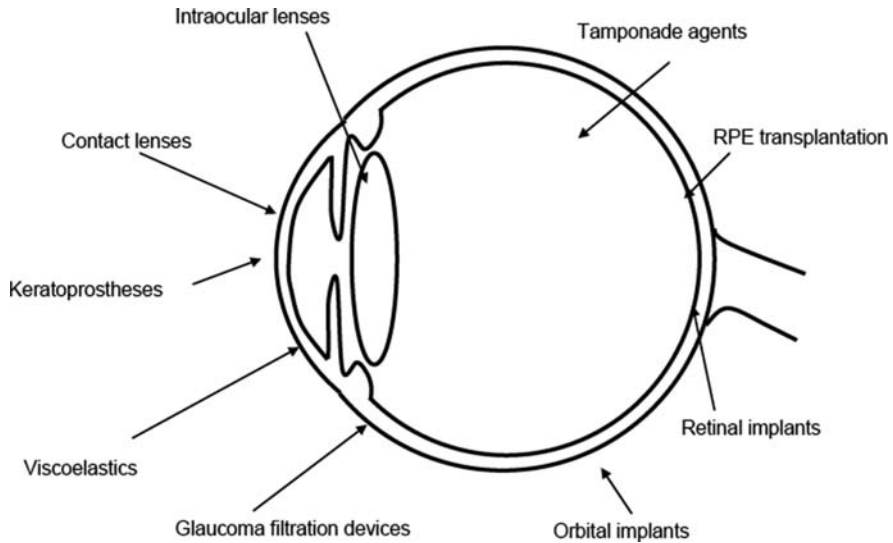
Fig. 12.1 The application of biomaterials in ophthalmology

12.2 Oxygen Delivery

A fundamental property of all viable biological tissue is respiration. The delivery of oxygen throughout the body is usually via perfusion with blood vessels. There are however a few notable exceptions where tissues do not have a direct blood supply in terms of arteries, veins or capillaries; these include the cornea in the eye and the cartilage in joints. Both tissues are bathed in fluids and the oxygen dissolved in the fluids is capable of meeting the metabolic demands of the tissues. In the case of the cornea, it is bathed on the internal aspect by aqueous humor and on the external surface by the tear film.

The aqueous humor is produced by ciliary processes inside the eye. These are highly vascularized finger like processes and the aqueous is formed by a combined process of active transport and diffusion. There are constituents in the aqueous with much higher concentration than in blood; this includes substances like vitamin C and glutathione. Other dissolved substances such as oxygen and glucose simply go down the diffusion gradient to supply the cornea.

The cornea is lined on the inner aspect with a single layer of cells called endothelium. The substance of the cornea is called the stroma and is made up of extracellular material and collagen fibers. The collagen fibers are arranged in lamellae or layers with precise spatial separation, the dimension of which is related to the wavelength of the visible spectrum of light. The endothelium keeps the cornea relatively dehydrated by an active transport mechanism. If for any reason the endothelium fails, the cornea will become edematous and lose its transparency. The external surface of the cornea is covered by a layer of epithelium. The cornea is also bathed with a film of

tears. The tear film presents a smooth refractive surface for light to enter into the eye and it transmits dissolved oxygen effectively to the corneal epithelium and stroma.

Covering the cornea by contact lenses potentially deprive the cornea of access to the oxygen supply dissolved in the tear film. The biomaterials used to make contact lenses deliver oxygen in different ways.

The number of people wearing contact lenses throughout the world is enormous and the materials used in their manufacture have evolved since the 1930s. There are a number of important properties required of a material to be suitable for use as a contact lens. These include stability, being non-irritant, robustness for handling, wettability to maintain tear film, resistance to spoilation by components of the tear film and most importantly oxygen permeability. Here we will consider particularly the way in which they influence oxygen delivery to the cornea.

The first readily available contact lenses were made of polymethylmethacrylate (PMMA) [2]. This had some very successful properties in terms of ease of manufacture to high tolerance, robustness for handling and stability. Its intermediate wettability allowed sufficient tear film coverage but its major drawback was the very poor oxygen permeability. This significantly restricted access of oxygen to the cornea. The route to overcome this problem was to make these PMMA lenses small so that they moved over the cornea during blinking thus allowing the tears to bathe the cornea. The problem with oxygen permeability led to the development of rigid gas permeable lenses.

Rigid gas permeable lenses are made by incorporating siloxane-based materials with the methacrylates for example, methacryloxy-propyltris (trimethyl siloxy silane) (TRIS) [21]. These materials had significantly higher oxygen permeability but the presence of the siloxane groups also increased their hydrophobicity thus reducing the tear film interaction and increasing the spoilation of the lenses. It is generally found that hard lenses and the original PMMA ones are less comfortable for the wearer and take longer for the eye to become accustomed to them.

Soft contact lenses are significantly more comfortable for the wearer. There were originally two types of soft lenses, hydrogel and polydimethylsiloxane. The first hydrogel lenses were developed from polyhydroxyethylmethacrylate (pHEMA) with a water content varying around 38–80% [21]. This provided excellent wettability and thereby increased the comfort factor. Although the oxygen permeability of these lenses was significantly better than PMMA it was afforded simply by the bulk water content and therefore increases in the oxygen permeability required higher water content and thinner lenses resulting in lenses that were more fragile and difficult to handle. Since the maximum oxygen permeability was limited by the water content of the gels it was determined that these lenses were not suitable for extended wear use because the health of the cornea could not be guaranteed. Most hydrogel lenses available now are copolymers of pHEMA with, for example, methacrylic acid (MAA), methylmethacrylate (MMA) or *N*-vinylpyrrolidone (NVP) and by adjusting the copolymer ratio it is possible to optimize the lens properties in terms of the water content, durability and strength. Many of these lenses are now marketed as disposable lenses and will be worn once and then thrown away.

Polydimethyl siloxane lenses have excellent oxygen permeability due to the bulky nature of the siloxane group and the high mobility of the polymer chains. They also have very good mechanical properties for ease of handling. Unfortunately the pendant methyl groups cause this material to be hydrophobic causing significant disruption of the tear film and also have a high susceptibility to lipid attachment and adhesion to the cornea. Attempts to increase the hydrophilicity of these lenses has been achieved by coating then with hydrogels or by gas plasma treatments.

Further developments in these devices then moved to exploring ways to incorporate components of both the hydrogels and the siloxane materials in an attempt to provide a material with both high oxygen permeability and high wettability. The development of silicone containing hydrogels has several difficulties including how to overcome the incompatibility for the hydrophobic and hydrophilic components during manufacture, which can lead to phase separation and consequent opacity. The key to achieving this has required the development of polymerization routes that allow the mixing of the hydrophobic and hydrophilic monomers. In contrast to the hydrogel lenses, which rely on the water content to increase oxygen permeability, these lenses rely on the silicone component and increases in the hydrogel component will decrease the oxygen permeability. Unfortunately these copolymers form with the silicone component at the surface; thus they are hydrophobic and therefore suffer from the same problems as the pure silicone lenses. To overcome this the surfaces can be modified to increase their hydrophilicity.

12.3 Refraction

The main conditions that we shall consider are myopia (short-sightedness), hypermetropia (long-sightedness), astigmatism and aphakia (without a lens). These are generally referred to as refractive errors. It means that the optics of the eye cannot focus a clear image onto the retina. To understand why, it is necessary to know how light is refracted by the various structures of the eye.

The main refractive power of the eye is provided by the cornea and to a lesser extent by the crystalline lens inside the eye. The cornea has an external surface covered by a tear film. Light rays travel from air (usually) and are refracted by curvature of the tear film overlying the cornea and enter into the eye. Once inside, the light rays are marshalled by the pupil, which varies in size allowing different amount of light through to the crystalline lens. The substance of the cornea and the aqueous has a refractive index of about 1.3, whereas the crystalline lens has greater refractive index of 1.4. The lens presents a front curved surface such that the light rays entering the lens will be further refracted. The lens has an even more steeply curved back surface. As the light rays exit the lens and enter the vitreous humor, there is a change in refractive index from 1.4 back to 1.3. The light rays are further refracted and travel through the vitreous cavity until they reach the retina. In a person with no refractive errors, the light rays will be refracted and be brought to

focus on the retina thus creating a clear image. When the optics of the eye fails to do this, refractive errors are said to occur.

The commonest refractive error is myopia. Myopia is often due to increased axial length; the eye is larger and the optics of the eye focuses the image of a distant object in front of the retina. By bringing the distant object nearer, the image can then be focused onto the retina. This is why myopia is also commonly called near or short sightedness. The "opposite" situation is when the axial length of the eye is too short, such that a distant object is brought to focus "behind" the retina. This is referred to as hypermetropia. Most young patients are capable of increasing the refractive power of the eye by "accommodation". By exercising the ciliary muscles, the lens changes its shape and becomes more rounded thus presenting a more curved surface to refract the light rays. Accommodation in the hypermetropic patient might bring a distant object into focus. There is however a limit to this accommodation and it is therefore more difficult to focus an image the nearer it is to the hypermetropic eye. That is why hypermetropia is also called long-sightedness, as it is easier for these patients to focus distant objects than those closer to the eye.

So far we have alluded to refractive errors caused by the eye having axial lengths that are too long or too short. Refractive errors can also be caused by the cornea being either too curved or too flat. The cornea presents an air/tear film interface and is therefore responsible for most of the refractive power of the eye. If the corneal surface is not perfectly rounded but is elliptical instead, then the refractive power in one axis may be greater than another. This creates what is called astigmatism. A point source of light in the distance cannot be brought into a single point focus inside the eye. A small amount of astigmatism is present in almost every person.

Another common cause of refractive error is due to cataract extraction. Cataract is the cloudiness of the crystalline lens of the eye. When a cataract is removed, the eye is no longer capable of focusing an image; replacement of the refractive power is required. This is achieved usually by placing an artificial intraocular lens in the eye. The normal crystalline lens is capable of changing focus as we alluded to previously. As yet, there is no intraocular lens that is capable of producing a significant amount of focusing.

In the case of refractive error, the requirement is simple. The aim is to focus the image on the retina. This is usually straightforward when both eyes have the same refractive error. There are some instances when the two eyes have significantly different refractive error. For example, a patient can be myopic in both eyes but only one eye develops a cataract. When the cataract is removed, one has an opportunity by using an intraocular lens to correct not just for the aphakia but also the myopia. If that is done, the un-operated eye will remain shortsighted, but the operated eye is in focus. This situation is called anisometropia, where the two eyes have significantly different refraction. The result of anisometropia is that the image size of an object cast on the retina can be significantly different between the two eyes. This difference in image size is referred to as aniseikonia. It has been estimated that the brain can cope with up to 30% difference in the image sizes but beyond this a patient might experience confusion and double vision. So the requirements are not only to provide refractive power to correct for errors, but also to balance the two eyes.

   Hard contact lenses rest on a film of tears. The main refractive surface then becomes the front surface of the contact lens rather than the tear film overlying the cornea; however, the tear film between contact lens and the cornea is also a refractive tool. Its front surface conforms to the back surface of the hard contact lens; its back surface conforms to the front surface of the cornea. The tear film has a refractive index close to that of the cornea. It therefore effectively neutralizes the refractive power of the front surface of the cornea. The front surface of the tear film is of course perfectly rounded as it conforms perfectly to the back of the hard contact lens. Therefore, the tear film acts as an astigmatic lens, overcoming the imperfection of the curvature of the cornea, converting an elliptical surface to a perfectly rounded surface. Hard contact lenses, in particular rigid gas permeable for example, silicone/acrylates and fluorosilicone/acrylates [3], are therefore useful for correcting astigmatism and can achieve this correction to limited extent. Soft contact lenses are more flexible and conform to the shape of the cornea. There is less of a tear film between the soft contact lens and the cornea. It is therefore not generally suitable for patients with astigmatism caused by elliptical curvature of the cornea.

12.4 Tissue Protection

Modern cataract surgery is often described as small incision microsurgery. Indeed, the cataract operation is one of the first keyhole type surgeries. The cataract has the shape and size of a "Smartie", being about 10 mm in diameter and 6 mm in thickness. In a young person the lens is semi-solid and capable of altering its shape largely as a result of the elasticity of the lens capsule. When the lens turns cataractous in old age it becomes very hard. Removing a lens in a child can be achieved by aspiration as the lens material is soft and this can be done via a small incision. After the age of 40, most lenses are too hard to be removed simply by aspiration. With the advent of phacoemulsification, small incision surgery was made possible. A special probe can be inserted into the eye via a small incision. This probe has a needle, which vibrates at ultrasound frequency providing the energy necessary to break down and liquefy the hard nucleus of a cataractous lens. The needle has an outer sleeve that provides an infusion and cools the needle and prevents the ultrasound energy from burning the surrounding tissue. Modern cataract surgery is therefore an elegant operation and is highly effective. The small incision ensures rapid visual recovery and minimizes the distortion of the cornea that can produce undesirable astigmatism.

   The use of ultrasound and the delivery of high energy into the anterior chamber are not without hazards. As described earlier the cornea is made up of collagen fibers arranged in layers. The transparency of the cornea is thought to be critically dependent on the spacing of the layers of collagen and when the cornea is waterlogged, it loses its transparency. The endothelium on the inner surface of the cornea plays a vital role in the active transport of water from the cornea. In adult life, the endothelial cells do not divide nor replenish themselves. With surgical trauma, such

as bending and distorting the cornea, endothelial cells are lost and a major concern is that the use of phacoemulsification can significantly increase endothelial cell loss. If the cell count falls below a critical level the endothelium can no longer maintain the clarity of the cornea.

Viscoelastics are used routinely in all cataract surgery. When injected into the anterior chamber, they coat the endothelium and absorb some of the ultrasound energy from the phacoemulsification probe. Furthermore, the viscoelastics are used to fill the anterior chamber to increase its depth such that instruments can be introduced without directly touching or damaging the endothelium. It is also injected between tissues as a spacer and provides a gentle way of separating them or holding structure back. The viscoelastics tend to stay together as a single bulk so any bleeding will not enter the substance of the viscoelastics; thus they help to maintain clarity and when the anterior chamber is filled it will not be obscured or stained by bleeding.

Viscoelastics are long chain polymers. Energy is dissipated through the relaxation processes in the long chain molecules of the viscoelastic substance thus they are capable of damping vibrations. Viscoelastics have several other important properties. They are described as cohesive, dispersive and viscous. Cohesive viscoelastic substances tend to have high molecular weights and high surface tension; they tend to stay together as a single bulk but they tend not to coat the tissue. Dispersive viscoelastic substances tend to have lower molecular weights and lower surface tension; they adhere to structures and they spread and coat tissues. In general, more viscous substances tend to flow less well. Once injected into the anterior chamber, they can be relied on to keep it inflated long enough to carry out some surgical maneuvers, even though the pressure in the eye tends to push the viscoelastics out. Combinations of viscoelastic substances with different properties are sometimes used. The so-called "soft-shell" technique uses first a viscoelastic with good dispersive properties to coat, adhere and protect the endothelium then, second, a viscoelastic with high viscosity and cohesive property is injected to keep the anterior chamber filled and inflated in order to carry out the phacoemulsification. Materials currently used in these products include hydroxypropylmethylcellulose, hyaluronic acid, chondroitin sulfate, polyacrylamide, collagen and various mixtures of these materials in an attempt to optimize the dispersive and cohesive properties [1, 16].

12.5 Tissue Integration

The integration of biomaterials with the tissue is an essential element of the treatment for a number of eye conditions. To give some examples in this section, we will deal with keratoprostheses for a temporary or permanent artificial corneal transplant, coral or hydroxyapatite orbital implants following removal of the eyeball and a "chip camera" as a retinal implant for artificial vision. It has to be pointed out that both keratoprosthesis and retinal implants are some way from being successful in the sense that neither are readily available as a viable treatment as yet.

12.5.1 Artificial Cornea Transplants

The cornea can turn opaque and cause blindness commonly from infection, trauma and some genetic eye condition. In the Western world, herpes simplex infection is common. Bacterial infections can lead to ulceration and the subsequent healing gives rise to cloudiness and loss of vision. Xerophthalmia is a major cause of blindness in Africa and Latin America. It is associated with vitamin A deficiency and general malnutrition. Infection and trauma leads to scarring of the cornea as part of the healing process. Scarring of the cornea give rises to loss of transparency and to the development of new vessels. Corneas that are severely damaged can be treated with corneal transplant (penetrating keratoplasty). Transplant of any human tissues are associated with rejection as the body becomes sensitized and mounts an immune reaction against the transplanted tissue. Rejection of the cornea is particularly a problem with eyes that have developed new blood vessels. Transplant requires donors and there has always been a shortage of donor material especially in third world countries. The sheer number of patients involved with corneal scarring diseases is such that it has long been imperative to find an artificial cornea graft or keratoprosthesis rather than to rely on donation from deceased donors.

The keratoprosthesis needs to be watertight since any leaks will give rise to portals for infection. By definition this requires the artificial cornea to integrate with the surrounding eye tissue, which are often abnormal and damaged by infection or trauma. For example, the eye often has abnormal lids that do not close properly and the conjunctiva may be dry and contracted. The keratoprosthesis will have to be transparent to allow light to enter the eye for sight. If the eye is severely damaged, it may also not have a crystalline lens so in these situations the keratoprosthesis will need to have refractive power to substitute for the lack of a lens in the eye. Furthermore, ideally the outer surface should support a functional corneal epithelium. The response of corneal endothelial cells to the inner surface of the keratoprosthesis also needs to be controlled.

Early designs consisted of PMMA cylinders to act as the optically clear window and various mechanical means of anchoring this to the surrounding tissues, for example, the Dohlman "collar-button" keratoprosthesis. Clinical results from these devices are limited mainly due to lack of integration of the surrounding tissue with the device. Attempts to improve this were made by incorporating the PMMA optic in autologous tissue such as tooth as in the Strampelli osteo-odonto-keratoprosthesis [10].

The next generation of keratoprosthesis involved a move towards more flexible materials that would improve the mechanical integrity of the tissue/implant interface. Initially these included various porous polymer skirts still attached to a rigid PMMA optic but more recently the advantages of flexible materials for the optics have also been realized, for example silicones and polyurethanes. The properties of the porous skirts require sufficient mechanical strength to withstand intraocular pressure and suturing to the surrounding tissue and they should encourage keratocyte ingrowth, collagen deposition and stabilization. Initially these were constructed from PTFE or carbon fiber reinforced PTFE (Proplast), which

were successful in terms of tissue ingrowth and suture placement. However, their attachment to a rigid PMMA optic was a recurring problem. Thus there was a focus of the research towards a one-piece, flexible design allowing tissue ingrowth directly into the device to produce a stable interface. Preliminary studies investigated the use of a device with a porous PTFE skirt, which was chemically bound to a transparent polyurethane elastomer optic with some success. Other combinations explored included silicone elastomer optics with carbon-fiber haptics and polypropylene/polybutylene skirts with PVA copolymer hydrogel central optics [5].

Major studies have evaluated the possible use of pHEMA in this application, in which the properties of the pHEMA can be controlled by the water content and they can be manufactured such that they consist of an outer poly HEMA sponge skirt which is fully integrated with a transparent optic [4].

In some situations, re-epithelialization of the keratoprosthesis surface is impractical owing to the underlying disease situation. If, however, appropriate ocular conditions apply, it would be a significant advantage if epithelialization of the surface could be achieved. The absence of this layer allows the entrance of bacteria, epithelial down-growth at the interface and inhibits spreading of the tear film. Current research to improve this process investigates the influence of surface properties and their modification on the epithelial cell adhesion and spreading [9].

12.5.2 Artificial Eye

The eye is sometimes removed as a whole because it contains a cancerous tumor or because of severe trauma. Severe injury can result in an eye that is not only blind but also painful. Healthy eyes have an intraocular pressure (IOP) as a result of the constant secretion of fluid (aqueous) by the ciliary processes within the eye. The fluid is drained away mainly via the anterior chamber angle; the drainage of the fluid against the resistance provided by the anterior chamber angle gives rise to the IOP. End-stage diseased eyes often have poor capacity to secrete fluid and the eye as a result is soft and appears collapsed. Quite apart from any visible damage, the sunken eye socket will be unsightly.

There are three main requirements when an orbital implant is placed in the socket after an eyeball is removed. First, the implant must replace the lost volume. Second, the implant should be capable of some "eye" movements. Third, the implant should be integrated to be secure such that it would not extrude from the eye socket.

Replacement of the volume is important when an eye is removed. Otherwise the socket appears sunken and unsightly; the upper lid will have a deep groove and the face will look asymmetrical. The eyeball is normally about 5–6 mL in volume. An orbital implant is used to fill the majority of this volume and a contact lens is used to cover this and improve the appearance.

The eyeball is attached by six "extra-ocular" muscles that move the eyeball as it looks in various directions. When the eye is removed surgically, the muscles are cut from the eye and the optic nerve is then divided. The eyeball is replaced with

an orbital implant, which is simply a spherical ball. This spherical ball is usually covered with donor sclera or with a polymer mesh. The muscles of the eyes are reconnected by sutures onto the surface of the covering so that when the other eye moves, the orbital implant also moves. When the patient wears a contact lens over the orbital implant, it gives the appearance that the two eyes are moving together. For small eye movements (looking say 30° in each direction), the "artificial eye" and the other eye usually move in a very realistic fashion but the movement of the "artificial eye" is limited in extreme directions of gaze.

Although orbital implants are attached to the muscles and covered by subcutaneous tissue and conjunctiva, they have a tendency to extrude. Extrusion occurs when the tissue overlying the implant wear away or breakdown [17]. This can occur when there is infection or if the implant is too big (to get the maximum volume replacements, surgeons tend to use the largest possible implant). The rate of extrusion depends on surgical technique and there being sufficient tissue to cover the implant and close the wound without putting the tissues under undue pressure. The foreign body response to the implant plays a significant role in the extrusion process. In the past, all implants were made of PMMA and the body tends to form a fibrous capsule and hold the implant in place. More recently, coralline hydroxyapatite (Fig. 12.2), alumina and porous polyethylene have been used [19, 7]. Their porosity allows the ingrowth of blood vessels and scar tissues, thus improving the integration of the implant with the tissue. After integration has occurred, a hole is sometimes drilled into the implant and a corresponding peg is made in the overlying contact lens. In this way, the orbital implant is directly connected to the contact lens and the range of movement of the artificial eye is increased [18].

Fig. 12.2 Hydroxyapatite orbital implant

12.5.3 Retinal Implants

There are a number of degenerative retinal diseases that cause blindness. Some such as retinitis pigmentosa has a genetic underlying cause and the photoreceptors and the underlying pigment epithelium are affected. In the last stages of the diseases, the blood vessels supplying the retina become thin and the optic nerve appears atrophic. Despite these advanced degenerative changes, many of the retinal ganglion cells remain viable. The ganglion cells form the inner layer of the retina and their axons normally carry signals to the brain via the optic nerve. In end-stage retinitis pigmentosa, the photoreceptors are atrophic and no longer react to light stimulation thus the eye is blind. With some of the ganglion cells remaining intact, it is possible to stimulate these cells directly electrically to produce the sensation of light. Access to the ganglion cells would be via the inner surface of the retina. This is known as the epiretinal approach. However, the normal retina is a network of cells that carries out signal processing locally before passing the information to the brain. This network is organized in such a way that it is particularly sensitive to the boundaries of light and dark, and to the orientation and the movements of such boundaries. This local processing is designed to extract the most "important" information from images cast on the retina and to pass this information on to the brain via the optic nerve. If retinal implants [23] are used to stimulate the ganglion cells directly, this input is bypassing most of the cells within the retina thus removing this local image-processing step. Therefore this processing work has to be carried out externally by a computer. An alternative approach is to use subretinal implants. By placing the retinal implants under the retina, these devices make use of the network of cells to carry out the local processing of the image.

Retinal implants need to make good contact with the retina to facilitate electrical conduction. The implants need to be anchored in a stable fashion so that good electrical contact is maintained despite movement of the eyes and the head. The anchoring or attachment of the implant should not be simply mechanical, but ideally involve cellular integration, with neurons (glial cells) growing into the implant to ensure good conductivity. The implant should not ideally produce a local inflammatory and fibrous tissue response. If the glial tissues in contact with the implant are replaced by scar tissues (fibroblasts-like cells laying down collagen) then the conductivity would be reduced and progressively larger currents would be necessary to stimulate the retina.

In the subretinal approach [20] an array of micro photodiodes is implanted under the retina sitting on the retinal pigment epithelium and assuming the role of the photoreceptors. The micro photodiodes therefore, process the incident light into electrical energy. The electrical energy is passed to the postsynaptic cells in the retina and thereby transferred to the optic nerve. The arrays of micro photodiodes are fabricated into discs in the order of $100\,\mu m$ thick and $3\,mm$ in diameter. These can be implanted subretinally either via the vitreous cavity or through the sclera. Prototype devices have been implanted in animals and have been well tolerated in terms of the tissue response to the materials although some degradation of the metal microelectrodes, usually manufactured from gold or silicon, has been reported. It

has also been demonstrated that some improvement in visual perception is possible although it is not certain that this is a result of electrical stimulation of the retina by the device since the surgery itself could cause transient stimulation due to release of neurotrophic factors. Further it is debatable whether sufficient electrical energy is produced by the currently available micro photodiodes to stimulate the retinal neurons and current research is developing secondary systems to amplify the response.

Fig. 12.3 Uncapsulated version of an epiretinal retinal prosthesis with the microfabricated antenna coil for data and energy transfer and the electronic components of the transponder system (*top*) and the microelectrode array with its drivers (*bottom*). Polyimide is used as the basic material. (We acknowledge Professor Peter Walter and the EPI-RET III consortium for providing this figure)

In the epiretinal approach the implant (Fig. 12.3) is placed on the inner surface of the retina and ganglion cells are directly stimulated. In this approach, therefore, the image has to be captured using a camera and processed to provide the properties of the image as a series of short current pulses. The camera is sometimes implanted in the lens capsule following lens extraction in a similar manner to cataract surgery. Intimate contact between the epiretinal implant and the ganglion cells is required for effective stimulation. It will be very important to maintain this during long-term implantation and thus it will be necessary to limit tissue ingrowth between the implant and the retina as a result of the response to the materials and/or the stimulation.

Animal studies have shown that these devices are well tolerated and have been shown to cause local activation of the visual cortex and preliminary clinical trials in blind patients have reported improved visual perception. Further technological advances in this area could lead to devices that could restore significant visual acuity to patients with degenerative diseases of the retina which are otherwise untreatable.

12.6 Modulation of Wound Healing

There are a number of instances where biomaterials are being used to modulate the wound healing response as part of a treatment process. This includes glaucoma filtration devices in which the control of fibrosis is used to allow fluid flow from the eye to control the IOP.

Glaucoma is a common ophthalmic disease, which results in the increase of IOP. This leads to damage to the optic nerve and significant loss of vision. Many glaucoma patients are successfully managed with a simple drug regime. For some patients, however, this does not control the IOP and it is necessary to use a surgical procedure to release fluid from the back of the eye, bypassing the normal route for fluid flow from the back to the front of the eye. Trabeculectomy is the procedure of choice but control of the fluid flow can be difficult and lead to hypotony of the eye. Implantation of glaucoma filtration devices is used to improve the control of the fluid flow.

Initially these devices were simple setons designed to maintain patency of the channel through the sclera/cornea into a subconjunctival bleb and involved materials such as silk, gold, platinum and magnesium. These universally failed often due to a chronic inflammatory response leading to occlusion of the channel. Tube designs were developed using materials such as polyethylene, silicone and PMMA. The success of these tubes depended on the fibrosis formed in the subconjunctival space to control the hypotony [14].

In 1969 Molteno developed a tube and plate design, which forms the basis of the current devices in clinical use. The idea of this design was that the plate would control the fibrosis around the subconjunctival bleb to maintain the patency of the fluid flow out of the eye. Current devices (Fig. 12.4) are predominantly made from

Fig. 12.4 Flexible plate
Ahmed valve (kindly
provided by Altomed Ltd,
UK)

silicone and various design modifications have been made to incorporate valves
within them to control the outflow of fluid [12].

12.7 Interfacial Tension and Tamponade

Retinal detachment is an important cause of blindness in the western world. It is the
final common pathway for many disease processes including diabetic retinopathy
and age-related macular degeneration. Retinal detachment is the separation of the
neurosensory retina from the underlying pigment epithelium. The commonest rea-
son for retinal detachment is that caused by a retinal break called rhegmatogenous
retinal detachment.

The center of the eye is filled with a transparent jelly-like substance called the
vitreous. The gel is made up of a network of long collagen fibers with glycoprotein
such as hyaluronic acid bound to the collagen. The hyaluronic acid in turn binds to
many molecules of water by non-covalent bonds leading to a gel which has 97% by
weight of water giving the vitreous a jelly-like, semi-solid consistency. The vitreous
moves as a single body when the eye moves, as it looks from one side to the other.
The vitreous is normally attached loosely to the back of the lens and the whole sur-
face of the retina and the optic nerve. With age and with certain conditions such as
abnormally high degree of myopia, the vitreous gel loses its homogenous consis-
tency and the collagen network collapses. The vitreous becomes detached from the
optic nerve and the posterior retina, whilst remaining attached to the retina and pars
plana anteriorly. As a result, the vitreous no longer moves as a single body. With
eye movement, the eye wall rotates and the vitreous posteriorly lags behind creating
what is referred to as dynamic traction. This vitreous traction is the cause of retina
tears.

The collapse of the vitreous and its separation from the optic nerve and the surrounding retina are referred as Posterior Vitreous Detachment (PVD). The posterior chamber is now made up of the semi-solid vitreous gel immediately behind the lens and water-like fluid vitreous in front of the retina. Eye movements in addition to producing traction on the retina also create currents in the fluid vitreous. After a retinal break is formed, the fluid behind the vitreous is channelled through the retina break and causes the retina to be detached. The fluid vitreous underneath the retina is now referred to as subretinal fluid. The separation occurs at the level of the photoreceptors such that the neurosensory part of the retina (all the nerve cells) is lifted away from the underlying epithelial part of the retina (retinal pigment epithelium). Once detached, the photoreceptors become apoptotic within hours. For this reason, the treatment of retinal detachment is relatively urgent otherwise there will be irreversible loss of vision. Thus a rhegmatogenous retinal detachment is caused by the combination of dynamic traction and currents in the fluid vitreous in the presence of a retinal break.

The treatment of retinal detachment aims at closure of the retinal breaks. The subretinal fluid can be drained away but this may be unnecessary because once the retinal break is closed, the subretinal fluid will be carried away by active transport by the retinal pigment epithelial cells. Retinal breaks are closed in one of two ways: by external or internal tamponade. The word tamponade carries the meaning of "plugging." External tamponade refers to buckling of the sclera. In this process the full thickness of the eye wall, including the choroids and the retinal pigment epithelium is "pushed in" to "plug" the retinal break. Scleral buckling is achieved by sutures and by the use of explants. The explants are made of either soft sponge-like silicone or more solid silastic material. Internal tamponade relies on the use of bubbles. The bubbles can be gas or liquid; they must be immiscible with water and form an interface with fluid vitreous. When injected into the vitreous cavity, they push against the retina "plugging" the opening thus preventing further liquid vitreous getting into the subretinal space.

The ability of a tamponade bubble to close retinal breaks depends crucially on the shape of the bubble. For example, if the bubble has a specific gravity close to that of the fluid vitreous, it will tend to remain spherical. This shape is not particularly useful because it makes only a small area of contact with the upper part of the retina. Thus positioning the bubble over the retinal tear requires the patient to position their head (and eye) in such a way in order to "float" the bubble against the correct part of the retina. Once the retinal break is closed, it needs to be "sealed" with retinopexy. Retinopexy is induced either by cryotherapy or laser photocoagulation. It causes controlled and localized tissue damage around the edge of the break. The healing response causes an adhesion of the retina to its surrounding structure and thereby a watertight seal. This healing response takes two weeks or so to form.

The effectiveness is therefore dependent on the shape of the bubble. This in turn is dependent on the interfacial tension against water and on the specific gravity of the gas or liquid bubble. Of the two, specific gravity is much more important in determining the shape of the bubble. In the case of liquid tamponade agents, the viscosity becomes an important consideration. There are two groups of tamponade

agents: the gases and silicone oils. They are all immiscible with water; the gases, however, are absorbable and will exit the eye. Silicone oils stay in the eye until they are removed usually 3–6 months after they have been injected although in some instances they are left in the eye permanently.

If an air bubble is used, the air will dissolve in the fluid vitreous and move down the concentration gradient into the blood stream and be carried out of the eye. The bubble may not plug the retinal breaks long enough for the retinopexy to take effect and the retina may become detached again. For this reason, gases such as sulfahexafluoride and perfluoropropane that are relative insoluble in water are used. This prolongs the retention of the bubble in the eye. Indeed, if a bubble of pure perfluoropropane is injected into the eye, dissolved nitrogen in the blood tends to move down the concentration gradient from the blood stream into the bubble. The bubble will therefore expand in size until equilibrium is reached when the egression of dissolved gases is balanced against the entry of dissolved nitrogen from the eye. This expansion of the bubble can give rise to a dangerous rise in the pressure, so insoluble gases like perfluoropropane are normally diluted with air to a non-expansile concentration such that the bubble would be in equilibrium avoiding a possible rise in eye pressure.

All gaseous tamponades will eventually dissolve and come out of the eye. This usually takes a few days and at most a few weeks depending on the solubility and concentration of the gas used. For example, a bubble made up of 12% perfluoropropane and air can last up to 10 weeks inside the eye. The bubble will, however, get smaller and smaller such that it is not useful for supporting the upper part of the retina for much more than 3 weeks. Specifically, in cases of Proliferative Vitreoretinopathy (exaggerated healing response), scar tissues can pull on the retina to form new retinal breaks and give rise to recurrence of retinal detachment. In these instances, it is desirable to achieve a wide support of the retina for a prolonged period. For this a non-absorbable tamponade agent might be used which normally means silicone oils.

It is important to compare the behavior of the gases and silicone oil tamponades [22, 8]. Gas or gas air mixtures have high interfacial tension with water (e.g., 80 dynes/cm for air). They are expected to adopt a very rounded shape. But because the specific gravity is so low (approximately 0.001 g/mL), the buoyancy is very high. Instead of a spherical bubble, gas and air bubbles have the shape of spherical caps. Silicone oil has a relatively low interfacial tension (35 dynes/cm), but a specific gravity (0.97 g/mL) close to water (1 g/mL). Silicone bubbles are almost spherical. For any given volume of a bubble, the more spherical the bubble, the less area of the retina it is expected to tamponade (Fig. 12.5).

Both groups of tamponade agents have their advantages and their limitations. A bubble of gas with its spherical cap shape means that most of the volume of the tamponade is used to support the retina but the bubble gets absorbed and duration of tamponade is limited. A bubble of silicone does not get absorbed and if the posterior segment of the eye is totally filled with silicone oil, then retinal breaks in all locations can be plugged. Randomized trials, however, failed to show silicone to be more effective than gas. One important reason for limitation of silicone oil is that
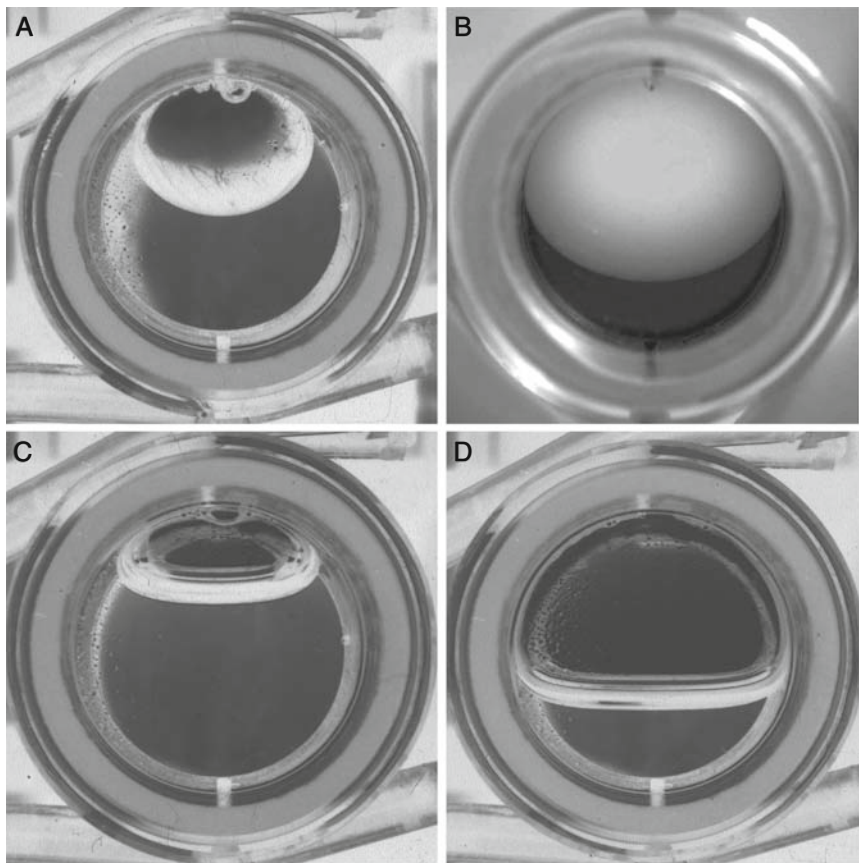
Fig. 12.5 Tamponade bubbles in a model eye chamber to illustrate the shape of the bubbler; (**A**) and (**B**) silicone oil bubbles, (**C**) and (**D**) air bubbles

it is not possible to totally fill the eye. Since the silicone oil bubble will maintain a spherical shape any under fill will result in a significant area of the lower part of the retina, which is not tamponaded.

Silicone oil was first introduced as a possible tamponade agent in 1958 [6]. There have been several multi-center trials of silicone oil tamponades that have reported that retinal reattachment is achieved in the majority of eyes using vitrectomy and a silicone oil tamponade and that the visual outcome is acceptable. There are, however, also several case studies that demonstrate specific problems associated with the use of silicone oil. These include inflammatory responses, retinal toxicity, lack of tamponading and emulsification.

For several years 1000cSt silicone oil was used clinically and after varying periods of time the oil was reported to emulsify in the eye. This caused three specific problems; first that the emulsion could not tamponade the retina, second the

emulsified droplets caused a clouding of the vision and third the emulsified droplets could stimulate adverse biological responses including inflammatory reactions and blocking of the fluid outflow from the eye. To reduce the risk of emulsification, 5000cSt silicone oil is now advocated for clinical use. However, the high shear viscosity of this oil significantly increases the difficulty of injecting the tamponade agent into the eye and its removal after the retina has reattached.

To tamponade retinal breaks in the lower segment of the retina it was necessary to find a tamponade agent that was heavier than water. Perfluorocarbon (PFC) with more than a five-carbon chain are liquids at room temperature. They are immiscible with water and having been originally investigated as blood substitutes have been shown to be non-toxic. PFCs are substances composed of a carbon backbone saturated with fluorine atoms. They are typically formed by fluorination of parent hydrocarbons. A range of PFCs is currently available for clinical ophthalmology applications

They are optically clear and their refractive index varies between 1.2 and 1.3. This property allows visualization of the retina through the PFC bubble. In practice it is desirable for an intraoperative tool to have a refractive index different to that of water (1.33), so that the interface between the agent and the infusion fluid can be seen during the vitreo-retinal procedure. As a tamponade, it is desirable for the agent to have refractive index closer to water so that the eye would undergo minimal refractive change post-operatively. The specific gravity of PFCs is in the range 1.7–2 g/cm^3 and thus they are significantly higher than water. So they will sink below water but also the buoyancy forces will be significantly higher, thus the bubbles will be flattened, increasing the contact between the retina and the tamponade agent. In addition, they are less hydrophobic than silicone oil and therefore will make more contact with the hydrophilic retina than silicone oils due to their interfacial energetics alone. However, their hydrophobic nature limits their passage through retinal breaks and increases the effectiveness of the PFC in providing intraocular tamponade to retinal breaks. PFCs have low viscosity (0.8–8.0 mPas) and although this property facilitates the injection and removal of the PFC through small-bore instruments during the surgery, it also increases their tendency to disperse into small bubbles.

Although PFCs have been shown to be non-toxic chemically the retina under them have been reported to suffer irreversible damage. It is believed that the high difference in specific gravity of these material and water could cause this damage either because the pressure on the retina is too high or that they are too effective at excluding water from the retinal tissue. Recent studies would suggest that the latter is more likely to be the cause.

The lack of short-term toxicity makes them ideal intraoperative tools for example, to unfold retina, the removal of subretinal fluid and the flotation and removal of dislocated intraocular lens components. At the end of the procedure they are removed and replaced with a tamponade agent that can remain in the eye for extended periods of time, for example silicone oil.

More recently semifluorinated alkanes (SFA) have been developed as tamponade agents [13, 11]. They are transparent liquids that are immiscible with water. In

terms of their interfacial energetics they are amphiphilic in that they have both a hydrocarbon end that is hydrophobic and a fluorocarbon end that is less hydrophobic. Measured against water their interfacial energetics is similar but slightly lower than that of PFCs. They have a lower specific gravity than the PFCs at between 1.2 and 1.7 g/cm^3 and thus, although they will sink below the remaining aqueous, the buoyancy forces will be less and therefore the bubbles will be more rounded than the same volume of PFC. These materials have low viscosities of around 2.5 mPas, making them easy to inject and remove through small-bore instruments but susceptible to emulsification.

The amphiphilic properties of the SFAs make them soluble in PFCs, hydrocarbons and silicone oil. It has recently been considered that it would be possible to take advantage of this property and mix SFAs with other tamponade agents, for example, silicone oil, to attempt to make a tamponade agent that has the advantageous properties of each material and overcomes the disadvantages of each.

The complete solubility of SFA in silicone oil can be produced in the laboratory. By mixing these materials in different ratios it is possible to produce fluids with different specific gravities. Several of these materials are available with a range of properties controlled by the ratio of the two components. These have the advantage of being heavier than water but with a lower buoyancy force in comparison with SFA or PFC alone. They also have a significantly increased viscosity in comparison with the pure SFA and PFCs making them potentially less susceptible to emulsification while at the same time they are less viscous than pure silicone oil, thus making them easier to handle. Thus these mixtures have the potential to overcome two of the disadvantages of pure SFAs and silicone oil. That is they will gently tamponade the inferior retina owing to the fact that their specific gravity is only slightly higher than that of water and they have a viscosity in a useful range. Unfortunately because of the chemical structure the interfacial energetics of the mixtures mimic those of silicone oil, i.e., highly hydrophobic rather than the less hydrophobic SFA. The interfacial energetics cause the SFA/silicone oil mixture to behave in a similar manner to pure silicone oil, that is they will make poor contact with the retina and they will not fit well into recesses formed by scleral explants. However it is expected that they will be as effective as silicone oil in the inferior section of the cavity and a thin layer of aqueous around most of the bubble could be beneficial in terms of the health of the retina.

12.8 Concluding Remarks

Biomaterials have made huge contributions to advances in ophthalmology with viscoelastics, intraocular lens and contact lenses being notable examples. There are areas where progress has led to recent developments, for example, the introduction of heavy liquids and heavy silicone oils for the treatment of retinal detachments. There are, however, many difficulties to overcome before artificial vision can be achieved through the use of retinal implants. These challenges can only be met

by the combined effort of clinicians, biologists, computer scientists and bioengineers. Undoubtedly, further understanding of the role that biomaterials have in these devices will lead to success in the future.

References

1. Andrews GP, Gorman SP, and Jones DS. Rheological characterisation of primary and binary interactive bioadhesive gels composed of cellulose derivatives designed as ophthalmic viscosurgical devices. Biomaterials, 2005, 26: 571–580.
2. Barr J. History and development of contact lenses, in Bennett E and Weissman B, eds. Clinical Contact Lens Practice, Lippincott Williams & Wilkins: Philadelphia, PA, 2005, p. 1.
3. Bennett E and Johnson J. Material selection, in Bennett E and Weissman B, eds. Clinical Contact Lens Practice, Lippincott Williams & Wilkins: Philadelphia, PA, 2005, p. 243.
4. Chirila TV. An overview of the development of artificial corneas with porous skirts and the use of PHEMA for such an application. Biomaterials, 2001, 22: 3311–3317.
5. Chirila TV and Hicks CR. Corneal implants, in Bowlin GL and Wnek GE, eds. Encyclopedia of Biomaterials and Biomedical Engineering, Marcel Dekker, Inc.: New York, 2004, p. 392.
6. Colthurst MJ, Williams RL, Hiscott PS, and Grierson I. Biomaterials used in the posterior segment of the eye. Biomaterials, 2000, 21: 649–665.
7. Custer PL, Kennedy RH, Woog JJ, Kaltreider SA, and Meyer DR. Orbital implants in enucleation surgery: a report by the American Academy of Ophthalmology. Ophthalmology, 2003, 110: 2054–2061.
8. Fawcett IM, Williams RL, and Wong D. Contact angles of substances used for internal tamponade in retinal detachment surgery. Graefes Arch Clin Exp Ophthalmol, 1994, 232: 438–444.
9. George A and Pitt WG. Comparison of corneal epithelial cellular growth on synthetic cornea materials. Biomaterials, 2002, 23: 1369–1373.
10. Hicks CR, Fitton JH, Chirila TV, Crawford GJ, and Constable IJ. Keratoprostheses: advancing toward a true artificial cornea. Surv Ophthalmol, 1997, 42: 175–189.
11. Hiscott P, Magee RM, Colthurst M, Lois N, and Wong D. Clinicopathological correlation of epiretinal membranes and posterior lens opacification following perfluorohexyloctane tamponade. Br J Ophthalmol, 2001, 85: 179–183.
12. Hong CH, Arosemena A, Zurakowski D, and Ayyala RS. Glaucoma drainage devices: a systematic literature review and current controversies. Surv Ophthalmol, 2005, 50: 48–60.
13. Kirchhof B, Wong D, Van Meurs J, Hilgers RD, Macek M, Lois N, and Schrage NF. Use of perfluorohexyloctane as a long-term internal tamponade agent in complicated retinal detachment surgery. Am J Ophthalmol, 2002, 133: 95–101.
14. Lim KS, Allan BD, Lloyd AW, Muir A, and Khaw PT. Glaucoma drainage devices; past, present, and future. Br J Ophthalmol, 1998, 82: 1083–1089.
15. Lloyd AW, Faragher RG, and Denyer SP. Ocular biomaterials and implants. Biomaterials, 2001, 22: 769–785.
16. Lloyd AW, Sandeman S, Faragher RG, and Denyer SP. Ocular implants, in Bowlin GL and Wnek GE, eds. Encyclopedia of Biomaterials and Biomedical Engineering, Marcel Dekker, Inc.: New York, 2004, p. 1098.
17. Oberfeld S and Levine MR. Diagnosis and treatment of complications of enucleation and orbital implant surgery. Adv Ophthalmic Plast Reconstr Surg, 1990, 8: 107–117.
18. Perry AC. Integrated orbital implants. Adv Ophthalmic Plast Reconstr Surg, 1990, 8: 75–81.
19. Shields CL, Shields JA, De Potter P, and Singh AD. Lack of complications of the hydroxyapatite orbital implant in 250 consecutive cases. Trans Am Ophthalmol Soc, 1993, 91: 177–189; discussion 189–195.

20. Walter P. Retinal implants, in Kirchhof B and Wong D, eds. Essentials in Ophthalmology –
      Vitreo-Retinal Surgery, Springer-Verlag: Heidelberg, 2004, p. 1.
21. Winterton L and Su K. Chemistry and processing of contact lens materials, in Bennett E and
      Weissman B, eds. Clinical Contact Lens Practice, Lippincott Williams & Wilkins: Philadel-
      phia, PA, 2005, p. 355.
22. Wong D, Williams R, Stappler T, and Groenewald C. What pressure is exerted on the retina
      by heavy tamponade agents? Graefes Arch Clin Exp Ophthalmol, 2005, 243: 474–477.
23. Zrenner E. Will retinal implants restore vision? Science, 2002, 295: 1022–1025.

Chapter 13
Hip Prosthesis

Afsaneh Rabiei

13.1 Introduction

Hip prosthesis or hip replacement surgery becomes necessary when the hip joint has been badly damaged from any cause such as arthritis, malformation of the hip since birth or abnormal development and damage from injury. Figures 13.1 and 13.2 show the anatomy of a natural hip. As can be seen, a natural hip is composed of a femoral stem (thigh bone) with a femoral head on top of it, that articulates against the acetabular cup in the acetabulum. As all other joints, there exists a cartilage between the acetabular cup and the femoral head to lubricate their movement and facilitate the articulation. In an arthritic hip joint, the cartilage has been damaged, narrowed or even lost by a degenerative process or by inflammation. Figure 13.3 shows an arthritic hip joint in which the cartilage has been damaged. This can make movement very painful for the patient. According to the Centers for Disease Control and Prevention, one out of every three Americans (an estimated 70 million people) is affected by one of the more than 100 types of arthritis. If a part of the joint is
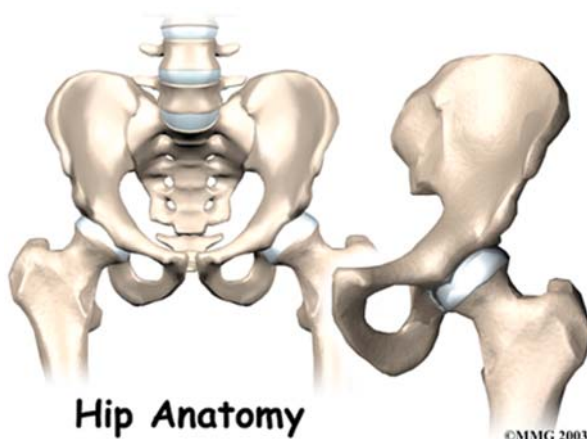
Fig. 13.1 Anatomy of a natural hip joint (image courtesy of Medical Multimedia Group, www.medicalmultimediagroup.com)
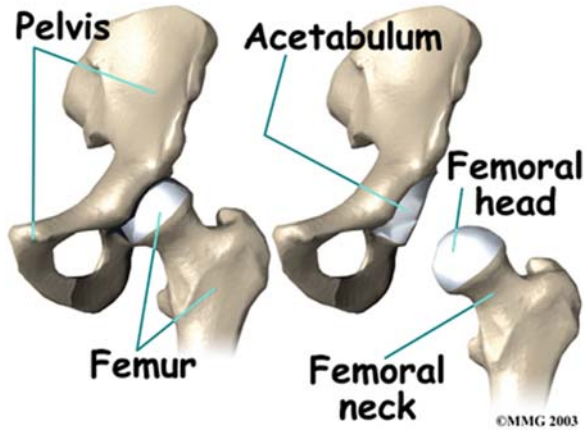
Fig. 13.2 Close up illustration of the anatomy of a natural hip joint (image courtesy of Medical Multimedia Group, www.medicalmultimediagroup.com)

damaged, a surgeon may be able to repair or replace only the damaged parts. On the other hand, if the entire joint is damaged a Total Hip Replacement (THR) or Total Hip Arthroplasty is suggested in which the damaged parts of the joint is being removed and replaced by an artificial prosthesis aimed for pain relief and restoration
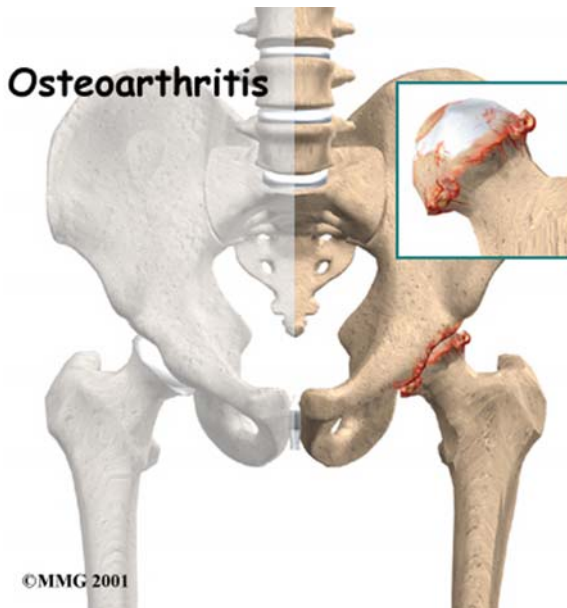
Fig. 13.3 Anatomy of an arthritic hip joint in which the cartilage has been damaged and even disappeared in some areas (www.medicalmultimediagroup.com)
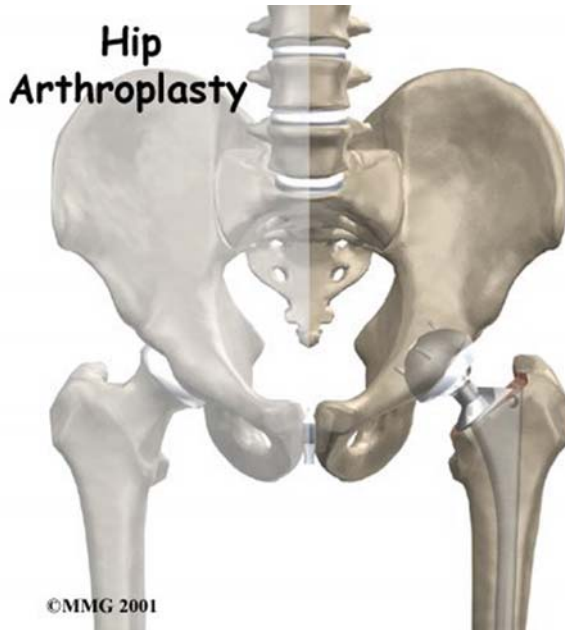
Fig. 13.4 Hip arthroplasty, in which the damaged parts of the joint have been removed and replaced by an artificial prosthesis (image courtesy of Medical Multimedia Group, www. medicalmultime-diagroup.com)

of movement. Figure 13.4 shows a hip arthroplasty, in which the damaged parts of the joint has been removed and replaced by an artificial prosthesis.

According to American Academy of Orthopedic Surgeons (AAOS), THR operation is one of the most successful joint reconstruction surgical operations in orthopedic surgery and consistently reduces or eliminates the pain in most patients. Total joint replacement is widely regarded as the major achievement in orthopedic surgery in the twentieth century. American Academy of Orthopedic Surgeons reported more than 200,000 THR surgeries being performed in the United States each year.

13.2 History of Total Hip Replacement

During the past 40 years many combinations of materials and designs have been considered and developed to achieve an ideal painless, stable and freely mobile joint implant with a longer lifetime. The first total hip replacement was performed by Gluck in 1890 using ivory, holding prosthesis with glue [1, 2]. Delbet in 1919, used a rubber component, but these early attempts were unsuccessful [2]. In 1939 Smith-Peterson described an interpositional (mould) arthroplasty, which initially was made of glass, but he modified it over a period of time to celluloid, Pyrex, Bakelite and finally Vitallium [2, 3]. Progress through the late 1930 and 1940 s was rapid, with

McKee and Watson Farrer (1966) describing a metal on metal hip with stainless steel material [2–5] and Charnley (1967) a metal femoral component and a low-friction high molecular weight polyethylene acetabular cup [2, 6]. By 1968 the superiority of cobalt-chrome over any other steel components was established [5]. But meanwhile, Charnley introduced the concept of low friction arthroplasty using both cemented femoral and acetabular components, with a plastic bearing surface and polymethylmethacrylate (PMMA) for fixation, and it is his prosthesis that remains the gold standard in hip replacement even today [2, 5]. In 1968 Ring pointed out disadvantages of using PMMA cements because of their poor bonding that caused loosening of the acetabular component and its eventual failure [5, 7]. He developed cementless cup components with long pelvic anchoring screws. In 1976 Dipisa, Sih and Berman [5, 8] suggested that loosening of THR might be partly caused by thermal necrosis of bone in contact with polymerizing PMMA. They reported a method to reduce bone cement interface temperature by pre-cooling the acetabular component, thereby increasing setting time by about 5.5 min [5].

In metal-plastic systems, movement at the interface of bone and cement and fractures of cement cause PMMA debris and accelerated wear of other components, leading to granuloma formation, bone resorption and final bone–cement disintegration [5]. As a result, scientists have started considering ceramic implants to achieve higher wear resistant properties. Alumina and zirconia balls are now in routine clinical use in Europe [5].

13.3 Various Components and Design of THR

As mentioned in the introduction, a THR includes the two major components of femoral and acetabular (or socket). The femoral component itself is including femoral head, femoral neck, femoral collar and femoral stem. Each one of these components has its own role in handling load and providing a smooth movement. In order to handle the applied load, the femoral component need to be connected to the adjacent bone using various fixations. Femoral fixation falls into two general groups of fixation using polymethylmethacrylate (PMMA) bone cement (cemented) or non-cemented categories. There are important design differences between the two varieties. Some people advocate a third type of artificial hip that uses cement on the ball side and non-cemented on the socket side (note that it is quite unusual to use a cemented socket with an uncemented ball). Figure 13.5 shows various components of a THR.

Successful joint surgery requires that the replacement joint be painless, stable and freely mobile with an acceptable service lifetime [5]. Ideally, the new joint must exactly match the geometry and performance of the natural hip. In any joint replacement surgery, there are three key variables: geometry, choice of material and operative procedure [2]. In this chapter we will focus mostly on the choice of material. Our primary focus will be on the overview of the many current designs and their combination with various biomaterials for hip arthroplasty.

Acetabular Cup
Outer Shell
METAL

Liner
POLYETHYLENE

Femoral Head
METAL

Metal-On-Polyethylene Bearing

Fig. 13.5 Various components of a THR (figure courtesy of Depuy Orthopedic)

13.3.1 Socket or Acetabular Cup

The artificial socket is used to replace the natural acetabular cup of the hip joint and to house the ball portion of the joint (Fig. 13.4). One of the most important factors affecting the durability of a hip implant is the quality of the acetabular cup. The hip joint is like a ball that fits into a socket. The socket part of the hip is called the acetabulum. The femoral head at the top of the thigh bone (femur) rotates within the curved surface of the acetabulum. In hip replacement surgery, the acetabular cup, usually comprised of a metal shell with a polyethylene liner, replaces the acetabulum (Fig. 13.4). The femoral component made up of the femoral stem and the femoral head (or ball) replaces the natural head of the femur.

The sockets for use with cement are made of a special type of polyethylene, which is very tough, and slippery particularly when wet. It has ridges on the outside and a wire marker so that its position can be seen on the X-ray. The ridges are designed to improve the fixation of the cup by the cement. It is implanted into the natural socket by first filling the cavity with the cement and then pushing in the artificial cup, which is held still while the cement sets.

The socket for use without cement is made of metal. It has a specially designed porous surface on its outer side to encourage bone growth into it and to join up with the socket much like broken bone heals. However, it is important that the artificial socket is held firmly in place while the healing process is going on and this can take 6–12 weeks. Machining the natural socket accurately and using an artificial socket that is precisely 1–2 mm bigger provides the initial fit. The metal socket is then forcibly "jammed" into place and further fixation can be obtained if necessary by

using additional screws. The next stage is to place a plastic insert within the metal shell against which the artificial ball will form the joint.

13.3.2 The Ball

The ball portion of the artificial hip consists of a metal stem or rod on top of which a metal ball is attached at an angle to mimic the shape of the top end of the human thigh-bone. The component for use with cement is fairly smooth on its outer aspect (Fig. 13.4). It is inserted into the thigh-bone after first filling the marrow cavity with bone cement and holding it still while the cement sets.

The implant for use without cement has a specially fabricated outer surface as in the case of the uncemented socket. The surface feels porous like a sponge and encourages bone to grow into it. This process takes about 6–12 weeks to complete. It is inserted after first machining the marrow cavity of the thigh-bone with special drills and then jamming in a component that is slightly larger.

13.3.3 Stem

In general there have been three types of femoral stems used in THR. The Charnely stem was manufactured from stainless steel. Later on, other materials such as titanium alloys (Ti-6Al-4 V) and CoCr alloys were used to manufacture the stem and other parts of THR. The choice of materials will be discussed in detail in the Sect. 13.4 of this chapter.

13.3.4 Fixation of THR

Excellent long-term results have been obtained with cemented total hip replacements. Traditionally, it is believed that an uncemented prosthesis is more suitable for the younger patient with higher bone density. Cemented hips, on the other hand, are believed to be the preferred method of fixation in older patients who have less dense bones. However, clinical experience has not supported this view [2]. Cemented fixation will provide an immediate stabilization of the prosthesis and allows the patient to walk on the new prosthesis the very next day, while in uncemented hips, most of the time, a lower weight-bearing status will be recommended for the patient for about 4–6 weeks after surgery.

While the original THR was cemented, the uncemented hip replacement was introduced in the 1980 s and is used mostly on the socket side of hip replacement and frequently on the stem side. The biological fixation between the prosthesis and the bone in cementless hips is a slow process and is achieved by press-fitting the implant into the bone. In such a case the inner diameter of the cavity made in the bone is 1 mm less than the outer diameter of the metal implant that is being inserted into the cavity. Another approach to encourage the biological fixation between the implant

©MMG 2002
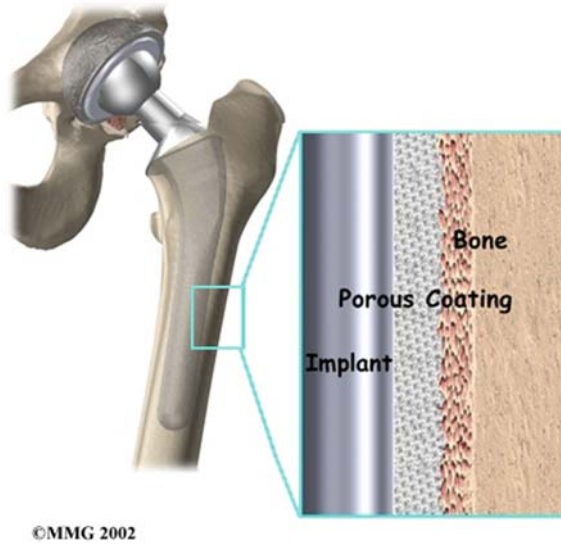
Fig. 13.6 Illustration of the bone-implant interface in an uncemented THR (image courtesy of Medical Multimedia Group, www.medicalmultimediagroup.com)

and the surrounding bone is by promoting the bone in-growth into the small pores at the implant surface, or mechanical supplemental fixation. Examples of porous surfaces are such as screws or pegs, for example on the cup to help attach to the bone. Figures 13.6 and 13.7 show some examples of cemented and uncemented hip prosthesis.

©MMG 2002
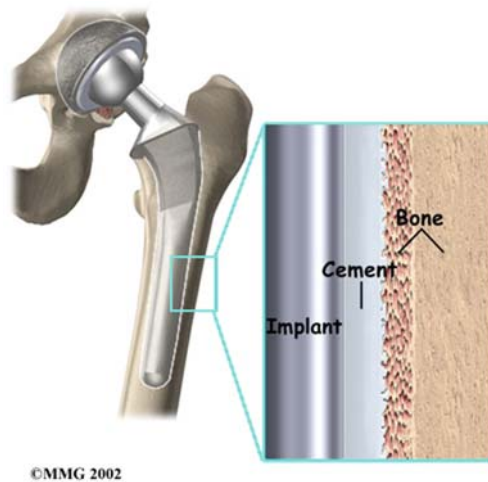
Fig. 13.7 Illustration of the bone-implant interface in a cemented THR (image courtesy of Medical Multimedia Group, www.medicalmultimediagroup.com)

    From the general and simplest viewpoint, contrasted to cemented implants, unce-
mented components can offer advantages of reduced operating time, reduced initial
trauma to endosteal bone surfaces, long term interface stability, less foreign material
(depending on implant design).

13.4 Various Materials for THR

The precise choice of implant is made by the specialist taking into account the
patient's age, lifestyle, activity level, whether the hip replacement is being done
for the first time or is a replacement for a previously carried out artificial hip and
also the specialists own experience and training.

    Only four broad material combinations have so far found favor in practical use:
a metallic femoral head in polymeric acetabular cup; a ceramic head in a poly-
meric cup; a ceramic head in a ceramic cup; and a metallic head in a metallic
cup [9]. The metals used for bearing surfaces are generally Co—Cr—Mo alloys, the
polymer Ultra-High-Molecular-Weight Polyethylene (UHMWPE), and the ceramic
high purity alumina, or in the case of ceramic on polymer joints only, in some
cases partially stabilized zirconia [9]. Modular design, where the stem and ball
are of two different materials, are very common in THR [10]. For example, hip
replacement implants featuring a titanium alloy femoral stem will have a Co—Cr
femoral head. Similarly, the UHMWPE socket of the common acetabulum replace-
ment can be implanted directly in the pelvis or be part of a modular arrangement
wherein the cup is placed into a metallic shell [10]. Table 13.1 lists some of the
femoral head to socket combinations that have been used for total hip replacement
arthroplasty. Cobalt base alloys are the most commonly used metals for current

Table 13.1 Materials combination in total hip replacement (THR) prosthesis

Femoral component	Socket component	Results
Co—Cr—Mo	Co—Cr—Mo	Early high loosening rate and limited use; new developments show lowest wear rate (THR only-in clinical use in Europe)
Co—Cr—Mo	UHMWPE	Widely employed; low wear
Alumina/zirconia	UHMWPE	Very low wear rate; zirconia more impact resistant
Alumina	Alumina	Minimum wear rate (components matched); pain-not in clinical use in the United States
Ti—6Al—4 V	UHMWPE	Reports of high UHMWPE wear due to breakdown of titanium surface
Surface coated Ti—6Al—4 V	UHMWPE	Enhanced wear resistance to abrasion; only thin treated layer achieved

UHMWPE, Ultrahigh molecular weight polyethylene
Source: [10, 14].

metal-on-polymer implants [10]. It is worth mentioning that the oxide surface layers on Ti alloy femoral heads result in excessive wear to the UHMWPE acetabular cups.

Although the metal on polymer hip prosthesis showed relatively low wear rate, it has been shown that the polymeric wear debris produced after a few years of using the prosthesis can cause adverse pathological reaction in the surrounding tissues, which can be a frequent cause for osteolysis and joint loosening [9]. It has been reported that cross linking the polymer chains can have a greater impact on lowering the wear rates [9, 11] compared to the application of very smooth ceramic femoral heads instead of metal heads [9, 12, 13]. The effect of various factors such as increasing the wear resistance of metal head by ion implantation and/or surface treatment as well as the impact of bearing clearances on lowering the friction in joint replacements are still open topics for scientist and biotribologists.

It has been reported [9, 15] that, under standard simulation conditions, the zirconia ceramic-on-UHMWPE prosthesis showed the highest wear rates (31 ± 4.0 mm^3/million cycles), followed by the metal-on-metal (1.23 ± 0.5 mm^3/million cycles), with the alumina-ceramic-on-ceramic prosthesis showing significantly lower wear rates of 0.05 ± 0.02 mm^3/million cycles. The alumina ceramic on ceramic wear rates has been 10- to 50-fold lower than some clinical retrieval studies, and the wear stripe observed on retrieved components was not reproduced under standard simulation conditions. The same references have reported that the mean size of the UHMWPE wear debris has been 300 ± 200 nm and 30 ± 2.25 nm for the metal particles and 9 ± 0.5 nm for the alumina- ceramic wear particles. The UHMWPE and cobalt-chrome wear particles were similar to those described in vivo; however, alumina-ceramic particles up to 3 μm have been identified in tissues around ceramic-on-ceramic prostheses [9, 15].

13.4.1 Alumina

In general, high density and high purity alumina (Al_2O_3 with more than 99.5% purity) is used in load bearing hip prosthesis and dental implants because of its combination of excellent corrosion resistance and biocompatibility, high wear resistance and high strength [10, 16–18]. Superb friction and wear properties of alumina occur when the grains are very small (<4 μm) and have a very narrow size distribution [10]. The presence of large grains can lower the wear resistance of alumina. Although some dental implants are single crystal sapphire, most alumina devices are very fine-grained polycrystalline α-Al_2O_3 produced by pressing and sintering at temperatures of 1,600–1,700 °C [10]. Small amounts of magnesia (MgO) is being added to alumina to aid sintering and limit grain growth during sintering [10]. Tables 13.2 and 13.3 summarize the physical and mechanical properties, along with the minimum requirements for alumina implants in accordance with ASTM (F603) and ISO (6474).

It has been reported [10] that the wear of alumina-to-alumina articulating surfaces in hip prosthesis is nearly 10 times lower than the metal-on-polyethylene

Table 13.2 Physical characteristics of Al_2O_3 bioceramics

	High-alumina ceramics	ISO 6474
Alumina content, %	<99.8	≥99.50
Density, g/cm^3	>3.93	≥3.90
Average grain size, μm	3–6	<7
Surface roughness (R_a), μm	0.02	…
Vickers hardness	2,300	>2,000
Compressive strength, MPa (ksi)	4,500(653)	…
Bending strength, MPa (ksi) (after testing in Ringers solution)	550(80)	400(58)
Young's modulus, GPa (psi×10^6)	380(55.2)	…
Fracture toughness (K_{IC}), MPa.$\sqrt{m}$ (ksi.$\sqrt{in}$.)	5–6(4.5–5.5)	…

Source: [10].

Table 13.3 Minimum physical characteristics of Al_2O_3 bioceramics in accordance with ASTM F 603

Alumina content, %	≥99.5
Density, g/cm^3	3.94
Median grain size, μm	≤4.5
Vickers hardness, GPa (10^6 psi)	18(2.56)
Compressive strength, GPa (ksi)	4(580)
Flexural strength, MPa (ksi)	400(58)
Elastic modulus, GPa (ksi)	380(55,100)
Weibull modulus	8

Source: [10].

surfaces. Low wear rates of alumina have led to widespread use of them in non-cemented cups press-fitted into the acetabulum of the hip, especially in Europe [10]. In such cases, the cups are stabilized by bone growth into grooves or around pegs. The mating femoral ball surface is also made of alumina, which is bonded to a metallic stem. Long-term results in general are excellent, especially for younger patients. However, stress shielding due to high elastic modulus of alumina may be responsible for cancellous bone atrophy and loosening of acetabular cup in older patients with senile osteoporosis or rheumatoid arthritis [10, 18]. As a result, it is important to consider the age of the patient, nature of the disease of the joint and biomechanics of the repair before any prosthesis is used including alumina ceramic [10].

13.4.2 Yttria Stabilized Zirconia

High purity dense Yttria stabilized Tetragonal Zirconia Polycrystal (Y-TZP) is also used as the articulating ball in THR prosthesis. This is a fine grain, high strength

Table 13.4 Minimum physical characteristics of yttria-stabilized tetragonal zirconia polycrystal bioceramics in accordance with ASTM F 1873

Zirconia content, %	$\geq$93.2
Yttria content, %	4.5–5.4
Density, g/cm^3	$\geq$6.0
Median grain size, μm	$\leq$0.6
Vickers hardness	1,200 HV
Flexural strength, MPa (ksi)	800(116)
Weibull modulus	10
Elastic modulus, GPa (10^6 ksi)	200(29)

Source: [10].

moderate high fracture toughness material manufactured by sintering at relatively low sintering temperatures (1,400 $^\circ$C) with an average grain size of 0.4–0.8 μm [10]. Minimum physical properties of Y-TZP bioceramics in accordance with ASTM F 1873 are listed in Table 13.4.

13.4.3 Polyethylene

Low density, high density, and ultra high molecular weight polyethylene (UHMWPE) are the three grades of commercially available polyethylene [19]. UHMWPE has less ductility and fracture toughness than those of other classes of polyethylene. But they have better packing of linear chains and high cross-linking levels which results in their increased crystallinity and mechanical properties. The high hardness and wear resistance of UHMWPE lead to their application in hip prosthesis. UHMWPE is used as the artificial socket in ball-and-socket joint replacement. The UHMWPE socket experiences cyclic loadings against the metal or ceramic ball. Hip simulator testing has shown that wear rates for UHMWPE made against Ti—6Al—4 V are 35% greater than that of Co—Cr—Mo alloy [10]. As a result of cyclic articulation under large force loadings against the metal ball, the passive oxide layer on the ball will be removed from the surface. The exposed metal may reform another protecting oxide layer or bond to the UHMWPE counterpart. In either case, further cyclic loadings will end up with more wear debris, which increases the surface roughness of the metal, and results in higher UHMWPE wear rates. Ultimately, the breakdown of the oxide layer creates the potential for abrasive wear, where the hard oxide debris acts as third body abrasive components [10]. The body treats the submicrometer-sized wear debris like a bacterium or virus, releasing enzymes to attack them. But the enzymes eventually kill the adjacent bone cells (osteolysis). This process causes bone resorption, loosening the implant, and eventually may even end up with the need for revision surgery for total implant replacement. There is therefore an increasing concern over the long term use of UHMWPE underscored by the recognition of the non-uniformity of the material, reports of the possible harmful

effects of UHMWPE sterilization, and the interaction of UHMWPE wear debris with the body fluids and tissues [10, 14]. Researchers are trying to find ways to decrease the wear rate of UHMWPE by improving UHMWPE properties, changing the ball material, and changing the processing of the replacement components.

13.4.4 Cobalt Based Alloys

Cobalt—chromium alloys are one of the first materials that successfully combined strength and biocompatibility for biomedical applications. Co—Cr alloys have very good corrosion resistance but they are brittle and difficult to fabricate. Co itself is not a particularly biocompatible material but the addition of 15–30% Cr will help the creation of a passivating oxide film that is very stable in the body. Allergic reactions to Co alloys in some cases have reported issues with implants. In general, metal sensitivity in some patients can lead to the death of the tissue around the implant and cause loosening of the implant. Toxic response or even cancerous changes in the tissues around the implant have been reported rarely [20, 21]. It is known that skin sensitivity to metals is rare and, as a result, the metal sensitivity is being developed over time in some cases and is not a preexisting condition. However, the occurrence of such allergies are less than 1% (reported for dental implants) and mainly affect women [10]. Contact allergic reactions to Cr from dental alloys are also reported in [22] but the occurrence of such reactions are rare [10]. While Co alloys are not 100% neutral in the body, the incidence of corrosion is very low and sensitivities seem to be limited to a small number of patients and they work well in replacement of bones in THR.

The four main Co alloys that are used mainly in medicine are:

- ASTM F 75, a Co—28Cr—6Mo casting alloy
- ASTM F 90, a Co—20Cr—15 W—10Ni wrought alloy
- ASTM F 799, a Co—28Cr—6Mo thermomechanically processes alloy with a composition nearly identical to ASTM F 75 casting alloy
- ASTM F 562, a Co—35Ni—20Cr—10Mo wrought alloy

Some properties of Co alloys developed for orthopedic implants are listed in Table 13.5.

Co—Cr—Mo alloys are mostly strengthened with carbides or nitrogen addition. They all have very good corrosion resistance. However, wear particles from metal-on-metal prostheses causes some potential health issues. Patients with metal-on-metal implants generally show increased systemic levels of cobalt and chromium, as well as higher levels of those elements in their blood and urine. This is unlike patients whose implants are metal-on-polymer, where the wear rate of the metal is much lower. Cobalt—chromium alloys can be cast, wrought, or fabricated using powder metallurgy. Wrought alloys are much more ductile than cast alloys and they

Table 13.5 Typical properties of cast and wrought cobalt-base alloys

ASTM desig-nation	Condition	Young's modulus		Yield strength		Tensile strength		Fatigue endurance limit (at 10^7 cycles, $R=-1$)	
		GPa	10^6 psi	MPa	ksi	MPa	ksi	MPa	ksi
F75	As cast/annealed	210	30	448–517	65–75	655–889	95–129	207–310	30–45
	P/M HIP[a]	253	37	841	122	1,277	185	725–950	105–138
F799	Hot forged	210	30	896–1,200	130–174	1,399–1,586	203–230	600–896	87–130
F90	Annealed	210	30	448–648	65–94	951–1,220	138–177	Not available	
	44% cold worked	210	30	1,606	233	1,896	275	586	85
F562	Hot forged	232	34	965–1,000	140–145	1,206	175	500	73
	Cold worked, aged	232	34	1,500	218	1,795	260	689–793[b]	100–1,159[b]

[a] P/M, powder metallurgy; HIP, hot isostatic pressing
[b] Axial tension, $R=0.05$, 30 Hz
Source: [10, 23].

can be cold worked, or forged. The addition of carbide or nitrogen as well as addition of small amount of alloying elements such as molybdenum, tungsten, chromium, manganese, or silicon to Co—Cr alloy can results in increased hardness of the alloy, thereby increasing the strength. Another approach to increase the strength of Co alloys is to reduce their grain size.

13.4.5 Titanium Based Alloys

Titanium and its alloys are of particular interest for biomedical applications because of their outstanding biocompatibility, low modulus of elasticity compared to other classes of implants such as stainless steel and Co alloys, as well as their little or not reaction with tissue surrounding the implant. Titanium alloys have the lowest moduli of 105–125 GPa, while stainless steels have moduli near 205 GPa and the cobalt—chromium alloys have the highest moduli of 240 GPa. Bone has a modulus on the order of 17 GPa (between 7 and 30 GPa in general depending the age and health condition of the person). The discrepancy between the modulus of bone and that of the alloys used for implant to support the structural loads causes the metallic devices implanted in the body to take a disproportionate share of the load. Consequently, the actual load experienced by bone will be proportionally lower due to the phenomenon known as "stress shielding". This phenomenon, which leads to the deterioration of bone quality, causes a decrease of bone thickness, bone mass loss, osteoporosis (bone resorption) that would eventually loosen the prosthetic device. A number of attempts have been made at decreasing the elastic modulus of Ti orthopedic alloys. Introduction of α-β titanium alloys with elastic modulus values approximately half that of stainless steels or Co—Cr—Mo alloys and the introduction of metastable β-titanium alloys with elastic modulus values in the range of 70 GPa are some of those. Table 13.6 present the elastic modulus of natural and synthetic joint materials.

Titanium alloys have very high corrosion resistance compared to that of other materials like stainless steel and Co—Cr—Mo alloys. Their high corrosion resistance is resulted from the formation of the stable oxide layer on the surface (spontaneous self-passivation) that has made them particularly good for total hip arthroplasty. Compared to stainless steel and Co—Cr—Mo alloys, the flexural rigidity and elastic modulus of Ti—6Al—4 V are lower, and their torsional and axial stiffness are closer to those of bone. As a result, they provide less stress shielding than either Co alloys or stainless steel.

Ti—6Al—4 V is more than 15% softer than Co—Cr—Mo alloys. As a result, when used in applications requiring articulation (THA), they will end up with more wear and wear debris. Significant wear of titanium femoral heads has been observed after articulation with ultrahigh molecular weight polyethylene (UHMWPE) both in wear simulation and in clinical retrievals. The low wear resistance of Ti alloys is related to the low mechanical stability of the passive surface film on them. Surface

Table 13.6 Elastic modulus value for natural and synthetic joint materials

Joint material	Elastic modulus	
	GPa	10^6 psi
Articular cartilage	0.001–0.17	0.000145–0.0247
PTFE	0.5	0.07
UHMWPE	0.5	0.07
Bone cement (PMMA)	3	0.44
Bone	10–30	1.45–4.35
TNZT alloys	55–66	7.9–9.6
"New generation" β-Ti alloys	74–85	10.7–12.3
Ti—6Al—4 V alloy	110	16
Zirconia	200	29
Stainless steel	205	30
Co—Cr—Mo alloy	230	33
Alumina	350	51

PTFE, polytetrafluoroethylene; UHMWPE, ultrahigh molecular weight polyethylene; PMMA, polymethyl methacrylate; TNZT, titanium-niobium zirconium-tantalum.
Source: [10, 14].

hardening through ion implantation has decreased the wear rate of titanium head implants. Some properties of Ti alloys developed for orthopedic implants are listed in Table 13.7.

13.4.6 Coatings

Coatings or ion implantation [24–27] are usually used to improve the surface properties and biocompatibility of implants as well as decreasing metallic wear and corrosion. Porous metal and ceramic coatings deposited on implants facilitate implant fixation and bone ingrowth [10, 24, 28, 29]. One simple method to allow tissue ingrowth into the implant is to modify its surface by implanting spherical beads [30] or wire mesh [24]. Coated implants are seen to exhibit superior hardness and wear resistance. Polymeric coating formulations are used to enhance biocompatibility and biostability, thromboresistance, antimicrobial action, dielectric strength and lubricity [10].

On the other hand, coating of hydroxyapatite (HA), $Ca_{10}(PO4)_6(OH)_2$, onto a bioinert metallic implant surface is an effective method of using this bioactive calcium phosphate compound in the human body [10, 28]. This will cause a direct bonding and fast stabilization of implant because of HA's similarity with the inorganic components of human bone [29–31]. Some biological advantages of HA coatings are enhancement of bone formation, accelerated bonding between the implant surface and surrounding tissues, and the reduction of potentially harmful metallic ion release. However, particulate debris at the bone prostheses interface with HA

Table 13.7 Properties of titanium alloys developed for orthopedic implants

Alloy designation	Elastic modulus		0.2% Yield strength		Ultimate tensile strength (min)		Elongation, %
	GPa	10^6 psi	MPa	ksi	MPa	ksi	
Alpha–beta alloys							
Ti–6Al–4 V	110	16	860	125	930	135	10–15
Ti–6Al–7Nb	105	15.2	795	115	860	125	10
Ti–5Al–2.5Fe	110	16	820	119	900	130	6
Ti–3Al–2.5 V	100	14.5	585	85	690	100	15
Beta alloys							
Ti–3Nb–13Zr	79–84	11.5–12.2	836–908	121–132	973–1,037	141–150	10–16
Ti–12Mo–6Zr–2Fe (TMZF)	74–85	10.7–12.3	1,000–1,060	145–154	1,060–1,100	154–160	18–22
Ti–15Mo	78	11.3	655	95	795	115	22
Ti–15Mo–5Zr–3Al	75–88	10.9–12.8	870–968	126–140	882–975	128–141	17–20
Ti–15Mo–2.8Nb–0.2Si–0.260 (21SRx)	83	12.0	945–987	137–143	979–999	142–145	16–18
Ti–16Nb–10Hf	81	11.7	736	107	851	123	10
Ti–35.5Nb–7.3Zr–5.7Ta (TNZT)	55–66	7.9–9.6	793	115	827	120	20

Source: [10, 14] and product literature from Stryker Howmedica Osteonics and Allvac, an Allegheny Technologies Company.

coated implants has been found to cause a foreign body response that is destructive to the surrounding tissues [32]. Dissolution rate of the implant must also be compatible with the rate of bone growth or otherwise the coating is not effective. The dissolution rate of crystalline HA has been observed to be very low, while that of the amorphous phase HA is considerably higher [32].

Various techniques have been used to deposit calcium phosphate on metal substrates such as sputtering, electron beam deposition, laser deposition, and plasma spraying [33–45]. So far, plasma spraying is the only technique being commercialized and most widely used because of its simplicity and cost effectiveness. Regardless of the processing technique, amorphous HA generally has a high dissolution rate in aqueous solutions. Therefore, the HA coatings need to be subsequently heat-treated in order to convert the ′as deposited′ amorphous phase into a crystalline phase [33–46]. However, the heat treatment increases the processing time and cost and can even cause residual stresses in the HA coating due to a thermal expansion mismatch between the coated layer and the metal substrate. These residual stresses may even lead to the cracking of the coating and eventual reduction in its bond strength [33–45].

Another type of coating for orthopedic implants is porous coatings, in which a porous layer of the base metal is coated onto the surface of the implant to increase the mechanical bonding between the implant and surrounding bone. This will happen by providing the empty spaces in porous coating into which bone can grow and mineralize, locking the device into place, which is called biological fixation. This type of coating has been discussed in Sect. 13.3.4 of this chapter in more detail.

In a new attempt, hydroxyapatite coatings with graded crystallinity were produced using ion beam assisted deposition with an in situ annealing process. This technique can minimize the processing time and cost as the deposition and annealing is performed at the same time. The graded crystallinity HA coatings deposited using this technique have reported higher nano-hardness and Young's modulus values as well as better bonding strength to the substrate [47–50] compared to HA coatings prepared by plasma spraying, sintering or even sputter deposition followed by post deposition annealing. It is anticipated that the functionally graded HA coatings can improve the implant-tissue bonding strength as well as provide fast relief for the patient who is receiving the orthopedic/dental implants by providing a higher dissolution rate associated with less crystallinity of the film at the top surface of the coating and lowering the dissolution rates through the thickness of the film towards the interface with the substrate.

13.5 Design Variation of THR

Design variation for THR prosthesis include the modular approach, straight stem, curved stem, platforms and no-platforms, holes and no-holes in the femoral stem

and so on [10]. The surgeon will choose the type of THR design for the patient based on his or her age, body condition, activity level and budget.

Artificial hip joint can easily last for 20–25 years. Although this period is reasonably long enough for many elderly patients, it is not long enough for younger recipients with a more active life style. In such cases hip replacement may end up with loosening or even failure that eventually requires revision surgery. The main conditions that cause the need for revision surgery are either related to osteolysis or in some cases infection. Osteolysis is a biological reaction to the debris from the bearing surfaces of implant as they rub against each other during daily life activities. Infection of the prosthesis occurs in only a very small portion of patients [51–53], but this dreaded complication results in major immobility due to pain, immobility, failure and loss of prosthesis, reoperation, and in some cases loss of limb or life [54]. The Osteolysis problems on the other hand, may not cause loss of life, but are still a major concern because of time, money and most importantly immobility associated with the revision surgery and its recovery period. It is notable that failure of prosthesis stems is a secondary event that follows stem loosening. Other factors that can promote stem loosening include bone resorption, degradation of bone cement, and unfavorable positioning of the prosthesis.

Improved joint prosthesis depends on progress in more reliable materials and design based on in-depth understanding of the important phenomena contributing to deep infection and loosening of joint implants [5]. For example, nowadays, due to systematic and/or local antibiotic treatment and other preventive measures, the incidence of deep hip prosthetic infections in Scandinavia appears to be less than 1% [5]. However, long term aseptic loosening of joint prosthesis is a quantitatively increasing problem [5]. Aseptic loosening linked with osteolysis caused by the wear debris of UHMWPE is the primary cause of failure of THR and accounts for almost three-quarters of all revision operations [55]. These motivated scientists to look for materials and designs with higher wear resistant in order to minimize osteolysis as well as antibiotic-loaded implants to lower the possibility of infection and eventually to minimize the need for revision surgery.

References

1. Gluck T. Autoplastik-transplantation implantation von Fremdkorpern. Klin Wochenschr, 1890, 27: 421–427.
2. Hughes S and McCarthy I. Science Basic to Orthopaedics, WB Saunders Company Ltd: Philadelphia, PA, 1998.
3. Smith-Peterson MN. Arthroplasty of the hip, a new method. J Bone Joint Surg [Br], 1939, 21B: 269–288.
4. McKee GK and Watson-Farrar J. Bone Joint Surg, 1966, 48B: 245–259.
5. Kossowsky R and Kossovsky N. Advances in Materials Science and Implant Orthopedics Surgery, NATO ASI Series, Kluwer Academic Publishers: Dordrecht, 1995, Vol. 294, pp. 103–133.
6. Charnley J. Total prosthetic replacement of the hip. Physiotherapy, 1967, 53: 407–709.
7. Ring PA. Complete replacement arthroplasty of the hip by the Ring prosthesis. J Bone Joint Surg, 1968, 50B940: 720–731.

8. DiPisa JA, Sih GS, and Berman AT. The temperature problem and the bone–acrylic cement interface of the total hip replacement. Clin Orthop, 1976, 121: 95–98.

9. Hutchings IM. Friction, Lubrication, and Wear of Artificial Joints, Professional Engineering Publishing: Bury St Edmunds, 2003, pp. 1–6.

10. Davis JR. Handbook of Materials for Medical Devices, ASM International, 2003.

11. Wang A, et al. Lubrication and wear of ultra-high molecular weight polyethylene in total joint replacement. Tribol Int, 1998, 31: 17–33.

12. Dowson D. A comparative study of the performance of metallic and ceramic femoral head components in total replacement hip joints. Wear, 1995, 190: 171–183.

13. Unsworth A, et al. Tribology of replacement hip joints, in Dowson D, et al., eds. Thinning Films and Tribological Interfaces, Tribology Series 38, Elsevier: Amsterdam, 2000, pp. 195–202.

14. Long M and Rack HJ. Titanium alloys in total joint replacement, a materials science perspective. Biomaterials, 1998, 19: 1621–1639.

15. Tipper JL, et al. Wear and functional biological activity of wear debris generated from UHMWPE-on-zirconia ceramic, metal-on-metal, and alumina ceramic-on-ceramic hip prosthesis during hip simulation testing, in Hutchings IM, eds. Friction, Lubrication, and Wear of Artificial Joints, Professional Engineering Publishing: Bury St Edmunds, 2003, pp. 7–28.

16. Doremus RH. Review: Bioceramics. J Mater Sci, 1992, 27: 285–297.

17. Boutin PM. A view of 15-years results using the alumina-alumina hip joint protheses, in Vincenzini P, ed. Ceramics in Clinical Applications. Elsevier: New York, 1987, p 297.

18. Christel P, Meunier A, Dorlot JM, Crolet JM, Witvolet J, Sedel L, and Britin P. Biomechanical compatibility and design of ceramic implants for orthopedic surgery, in Ducheyene P and Lemons J, eds. Biocermics: Materials Characteristics Versus In-Vivo Behavior, Vol. 523, Annals of NY Academy of Science: New York, 1988, p. 234.

19. Cooper S, et al. Polymers, in Ratner BD, ed. Biomaterials Science, An Introduction to Materials in Medicine, 2nd edn, Academic Press: New York, 2004.

20. Black J. Metal on metal bearings. A practical alternative to metal on polyethylene total joints. Clin Orthop Relat Res, 1996, 329 Suppl: S244–S255.

21. Meritt K and Brown A. Particulate Metal: Experimental Studies, in Morrey BF, ed. Biological, Material, and Mechanical Considerations of Joint Replacement, Bristol-Myers Squibb. Zimmer Orthopaedic Symposium Series, Raven Press, Ltd.: New York, 1993, pp. 147–159.

22. Mitchell EW. Summary and recommendations to the workshop, Workshop in Biocompatibility of Metals in Dentistry, Conference Proceedings, American Dental Association: Chicago, IL, 1984.

23. Brunski JB. Metals, in Ratner BD, Haffman AS, Schoen FJ, and Lemons JE, eds. Biomaterials Science: An Introduction to Materials in Medicine, Academic Press: New York, 1966, pp. 37–50.

24. Korkusuz P and Korkusuz F. Hard tissue-biomaterial interactions, in Yaszemski MJ, ed. Biomaterials in orthopedics, 2004, M. Dekker: New York. pp. 1–40.

25. Sovak G, Weiss A, and Gotman I. Osseointegration of Ti6Al4V alloy implants coated with titanium nitride by a new method. J Bone Joint Surg, 2000, 82-B: 290–296.

26. Sawase T, Wennerberg A, Baba K, Tsuboi Y, Sennerby L, Johansson CB, Albrektsson T. Application of oxygen ion implantation to titanium surfaces: effects on surface characteristics, corrosion resistance and bone response. Clin Implant Dent Relat Res, 2001, 3: 221–229.

27. Krupa D, Baszkiewicz J, Kozubowski JA, Barcz A, Sobczak JW, Bilinski A, Lewandrowska-Szumiel M, and Rajchel B. Effect of phosphorus-ion implantation on the corrosion resistance and biocompatibility of titanium. Biomaterials, 2002, 23: 3329–3340.

28. Lacefield WR. Hydroxyapatite coating. Ann NY Acad Sci, 1988, 523: 72–80.

29. Hench LL. Bioceramics. J Am Ceram Soc, 1998, 81: 1705–1728.

30. Suchanek W and Yoshimura M. Processing and properties of hydroxyapatite-bases biomaterials for use as hard tissue replacement implants. J Mater Res, 1998, 13: 94–117.

31. LeGeros RZ. Biodegradation and bioresorption of calcium phosphate ceramics. Clin Mater, 1993, 14: 65–88.
32. Wang S, Lacefield WR, and Lemons JE. Interfacial shear strength and histology of plasma sprayed and sintered Hydroxyapatite implants in vivo. Biomaterials, 1996, 17: 1965–1970.
33. van Dijk K, Schaeken HG, Wolke JGC, and Jansen JA. Influence of annealing temperature on r.f. magnetron sputtered calcium phosphate coatings. Biomaterials, 1996, 17: 405–410.
34. Ong JL, Lucas LC, Lacefield WR, and Rigney ED. Structure, solubility and bond strength of thin calcium phosphate coatings produced by ion beam sputter deposition. Biomaterials, 1992, 13: 249–254.
35. Chen TS and Lacefield WR. Crystallization of ion beam deposited calcium phosphate. J Mater Res, 1994, 9: 1284–1290.
36. Yoshinari M, Ohtsuka Y, and Derand T. Thin hydroxyapatite coating produced by the ion beam dynamic mixing method. Biomaterials, 1994, 15: 529–535.
37. Choi JM, Kong YM, Kin S, Kim HE, Hwang CS, and Lee LS. Formation and characterization of hydroxyapatite coating layer on Ti-based metal implant by electron beam deposition. J Mater Res Soc, 1999, 14: 2980–2985.
38. Singh RK, Qian F, Nagabushnam V, Damodaran R, and Moudgil BM. Excimer laser deposition of hydroxyapatite thin films. Biomaterials, 1994, 15: 522–528.
39. Cotell CM, Chrisey DB, Grabowski KS, and Spregue JA. Pulsed laser deposition of hydroxyapatite thin films on Ti–6Al–4 V. J Appl Biomater, 1992, 8: 87–93.
40. Ducheyne P, Raemdonck WV, Heughebaert JC, and Heughebaert M. Structural analysis of hydroxyapatite coating on titanium. Biomaterials, 1986, 7: 97–103.
41. Tsui YC, Doyle C, and Clyne TW. Plasma sprayed hydroxyapatite coating on titanium substrates. Part 1: Mechanical properties and residual stress levels. Biomaterials, 1998, 17: 2015–2029.
42. Zyman Z, Weng J, Liu X, Zhang X, and Ma Z. Amorphous phase and morphological structure of hydroxyapatite plasma coatings. Biomaterials, 1993, 14: 225–228.
43. Ji H and Marquis PM. Effect of heat treatment on the microstructure of plasma-sprayed hydroxyapatite coating. Biomaterials, 1933, 14: 64–68.
44. Chen J, Wolke JGC, and de Groot K. Microstructure and crystallinity in hydroxyapatite coatings. Biomaterials, 1994, 15: 396–399.
45. Brossa F, Cigada A, Chiesa R, Paracchini L, and Consonni C. Post-deposition treatment effects on hydroxyapatite vacuum plasma spray coatings. J Mater Sci Mater Med, 1994, 5: 855–857.
46. Chen J, Tong W, Coa Y. Feng J, and Zhang X. Effect of atmosphere on phase transformation in plasma sprayed hydroxyapatite coatings during heat treatment. J Biomed Mater Res, 1997, 34: 15–20.
47. Rabiei A and Thomas B. Processing and development of nano-scale ha coatings for biomedical application. MRS, fall meeting 2004, 845: 193–199.
48. Rabiei A, et al. A study on functionally graded HA coatings processed using ion beam assisted deposition with in situ heat treatment. Surf Coat Technol, 2006, 200(20–21): 6111–6116.
49. Rabiei A, et al. Microstructure, mechanical properties, and biological response to functionally graded HA coatings. Mater Sci Eng C-Biomimetic Supramolecular Syst, 2007, 27(3): 529–533.
50. Rabiei A, et al. A novel technique for processing functionally graded HA coatings. Mater Sci Eng C-Biomimetic Supramolecular Syst, 2007, 27(3): 523–528.
51. Ahnfelt L, Herberts P, Malchau H, and Anderson GB. "Prognosis of total hip replacement" A Swedish multi-center study of 4,664 revisions. Acta Orthop Scand Suppl, 1990, 238: 1–26.
52. Gristina AG, Naloy PT, and Myrvik QN. Mechanism of musculoskeletal sepsis. Orthop Clin North Am, 1991, 22: 363–372.

53. Morrey BF and Bryan RS. Infection after total elbow arthroplasty. J Bone Joint Surg Am, 1983, 65A: 330–338.
54. Bisno AL and Waldvogel FA. Infections Associated with Indwelling Medical Devices, ASM Press: Washington, DC, 1994, pp. 259–291.
55. Malchau H and Herberts P. Prognosis of total hip replacement: Surgical and cementing technique in THR: A revision study of 134,056 primary operations, in Proceedings of the 63rd American Academy of Orthopedic Surgeons, Atlanta, 1996.

Chapter 14
Burn Dressing Biomaterials and Tissue Engineering

Lauren E. Flynn and Kimberly A. Woodhouse

14.1 Introduction

The skin is the largest organ of the body, ranging in size from 1.5 to 2.0 m^2 in adults [1]. This highly-organized composite structure fulfils a wide variety of functions critical to the maintenance of homeostasis [2]. Burns, caused by thermal, chemical, or electrical injuries, can result in severe and irreparable damage to the skin, leading to wound contracture, scar tissue formation, and a loss of functionality. In the United States, between 60,000 and 80,000 patients are hospitalized annually for the treatment of serious burns [3, 4]. The average cost of patient care, reconstruction, and rehabilitation is extremely high, especially in severe or extensive cases [5]. Improvements in resuscitation techniques now facilitate the survival of patients with major burns extending over more than 90% of their bodies [6].

Conventional treatment strategies for wound closure rely on the use of autografts. Weaknesses of the autologous approach include the need for additional surgery, the creation of donor site defects, technical difficulty, expense, and higher patient risk [7, 8]. Allogenic materials have also been implemented, but are only suitable for temporary wound coverage, as they are associated with immune rejection and a failure to integrate into the host tissues [9]. Consequently, there is great need for the development of an economical tissue-engineered skin substitute. To date, numerous burn dressing biomaterials and engineered constructs have been investigated, but have failed to match the wound healing properties of grafted autologous tissues [7, 10–12]. With continued research, it may be possible to design a device that could promote the complete regeneration of skin, with rapid wound closure, functional recovery, and excellent cosmetic results.

14.2 Physiology of the Skin

In order to develop a successful tissue-engineering strategy to promote regeneration following burn injury, it is critical to have a comprehensive understanding of the structure and function of normal skin.

14.2.1 Basic Organization and Cellular Composition

The skin, or integument, is composed of multiple tissue and cell types arranged to form a complex, multifunctional organ (Fig. 14.1). The integumentary system is multilayered and can be divided into three interconnected regions: the epidermis, the dermis, and the hypodermis [2]. Located within these layers are the accessory structures or appendages, including glands, hair follicles, and nails, that are critical to the proper functioning of the skin [13]. The basal lamina located between the epidermal and dermal layers also plays a critical role in the maintenance of skin structure and function [14]. The primary cell types found within the skin include keratinocytes, melanocytes, Merkel cells, Langerhans cells, and fibroblasts [15].

14.2.1.1 Keratinocytes

Keratinocytes are the stratified squamous epithelial cells that are the primary cellular component of the epidermis, comprising 90–95% of all cells within this region [2]. These cells function as a protective barrier and also play an important role in skin immune function, secreting numerous cytokines that mediate inflammatory and immune responses, including interleukins and interferons [16].

The keratinocyte phenotype changes as the cells progress outwards from the basal layer towards the skin surface. This alteration corresponds to a change in
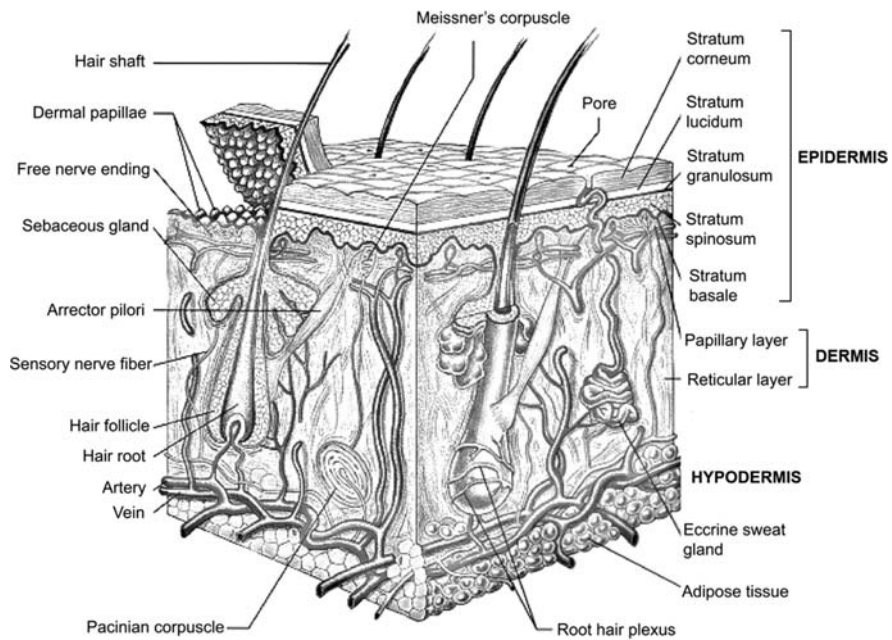
Fig. 14.1 Structure of the integumentary system. (Figure reprinted with permission from Marieb EN. Human Anatomy and Physiology. Copyright by Pearson Education Incorporated)

the state of differentiation of the keratinocytes from proliferative to cornified cells [17]. The structural changes in the cells can be correlated with variations in the expression of keratins, the main structural proteins produced by the keratinocytes [18]. These alpha-helical proteins can be classified as either acidic (type I) or basic (type II). During assembly, heterodimers, each composed of one acidic and one basic keratin, are arranged to form cytoskeletal intermediate filaments [19]. The keratinocytes in the basal layer express keratins 5 (acidic) and 14 (basic), whereas the more-differentiated cells in the outer layers express keratins 1 (acidic) and 10 (basic) [20, 21].

14.2.1.2 Melanocytes

Melanocytes are pigment-producing cells that are typically found in the basal epidermal layer in an approximate ratio of 1 melanocyte to every 10 basal keratinocytes [15]. In general, these cells do not divide and their survival is dependent on growth factors secreted by the keratinocytes, such as melanocyte stimulating hormone (MSH), basic fibroblast growth factor (bFGF), and endothelin-1 [22, 23]. Melanocytes have many cytoplasmic processes, termed dendrites, which facilitate the delivery of the pigment melanin to the epidermal keratinocytes [24]. A single melanocyte will supply melanin to approximately 35 keratinocytes through the transfer of unique organelles termed melanosomes, which produce and store the pigment within the cells [25]. The specific mechanism of this transfer is currently unclear, but may involve keratinocyte-melanocyte membrane fusion, or melanosome exocytosis followed by keratinocyte coated-pit endocytosis or phagocytosis [26]. The two main types of melanin are eumelanin, which is brown to black in color, and pheomelanin, which is yellow to red in color [25, 27]. Skin color is dependent on both melanosome size and density within the keratinocytes and can be affected by factors such as ultraviolet light and MSH [28].

14.2.1.3 Merkel Cells

Merkel cells are specialized neuroendocrine cells that are found in the epidermis and dermis in varying site-specific densities. The highest concentrations of these cells are in the outer root sheaths of the hair follicles and in the epidermal ridges of the deep epidermis [29]. In some regions of the skin, these cells are arranged into units termed tactile discs or touch receptors. Upon compression, Merkel cells can function as mechanoreceptors, secreting neuropeptides that stimulate the dermal nerve endings [2]. The cells can be identified microscopically by their characteristic cell-surface microvilli and dense-core cytoplasmic granules that contain various neuropeptides [30]. Epidermal Merkel cells have been shown to produce nerve growth factor (NGF), which can bind to NGF receptors expressed on dermal Merkel cells [31]. These dermal cells often form synaptic contacts with the axon terminals present in the dermis. Merkel cells can also be connected to keratinocytes via desmosomal junctions and may play a stimulatory role in the proliferation and differentiation of these cells [2].

14.2.1.4 Langerhans Cells

Langerhans cells are antigen-presenting dendritic cells that function in the immune response to pathogens and cancerous cells within the skin [32]. These cells are primarily found in the non-basal region of the epidermis, but can also be found within the dermis. Overall, the Langerhans cell population is relatively small, representing 1–5% of the total number of epidermal cells. Antigens diffusing through the epidermis are likely to contact one of the many dendrites of the Langerhans cells [15, 33]. Endogenous antigens, such as those related to epidermal cancer, are also recognized [34]. The Langerhans cells engulf the antigens via phagocytosis and then migrate across the basal lamina and underlying dermis into the regional lymph node [35]. Following intracellular processing, the antigens are presented at the cell surface bound to major histocompatibility complex (MHC) class II molecules. Interaction of the T-cells in the lymph node with the MHC/antigen complexes, combined with the release of stimulatory cytokines by the Langerhans cells, results in T-cell activation [32, 36, 37]. Langerhans cells are also involved in the initiation of T-cell-dependent B-cell immune responses and secondary T-cell responses [33]. For wound healing treatment strategies, it is important to note that these cells have been shown to play a role in the rejection of grafted allogenic and xenogenic skin [38].

14.2.1.5 Fibroblasts

Fibroblasts are the main cellular component of the dermis and are responsible for the production and secretion of numerous extracellular matrix (ECM) molecules found in the skin, such as collagen, elastin, fibronectin, decorin, tenascin, laminin, and various proteoglycans [39]. To allow for turnover of the ECM, these cells also secrete matrix metalloproteinases (MMPs) [40].

14.2.2 The Epidermis

The epidermis, the outermost region of the skin, provides many of the barrier properties of the organ. This avascular, multi-layered epithelium has very sparse ECM and is composed primarily of cells. The thickness of the epidermis is site-specific, with thicker layers found in the regions of the body that experience greater frictional forces, such as on the palms of the hands and the soles of the feet. The deeper layers of the epidermis are arranged into the epidermal ridges that increase the contact area between the epidermis and the dermis [2, 15, 24]. To support the continual regeneration of the epidermis, the keratinocytes differentiate in a process termed keratinization or cornification [17]. The progressive stages of differentiation can be observed in the epidermal layers: the stratum germinativum, stratum spinosum, stratum granulosum, stratum lucidum and stratum corneum (Fig. 14.2). In humans, it takes approximately 14 days for a cell leaving the innermost epidermal layer, the stratum germinativum, to reach the outer stratum corneum. These
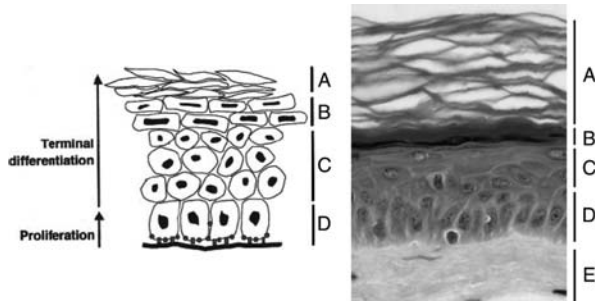
Fig. 14.2 A diagrammatic and histological view of the epidermal layers. (**A**) The stratum corneum. (**B**) The stratum granulosum. (**C**) The stratum spinosum. (**D**) The stratum germinativum. (**E**) The underlying dermis. The integrin junctions between the basal keratinocytes and the underlying basal lamina can be seen in the diagram on the *left*. (Diagram on *left* adapted with permission from Janes et al. [43]. Copyright 2002 by John Wiley and Sons Limited. Histological image on right reprinted from www.lab.anhb.uwa.edu.au with the permission of Dr. L. Slomianka)

cells then typically reside in the stratum corneum for an additional 14 days before they are shed from the body [15].

14.2.2.1 Stratum Germinativum

The stratum germinativum, also referred to as the stratum basale, is the innermost epidermal layer, composed of a monolayer of cells that are firmly attached via hemidesmosomal junctions to the basal lamina that separates the epidermis from the dermis [2, 14]. This cell layer primarily consists of morphologically-identical, columnar basal keratinocytes, interspersed with melanocytes and Merkel cells. The basal keratinocytes can be classified as one of three types based on their proliferative capacity: stem cells, transient-amplifying cells, or post-mitotic differentiating cells [41]. Approximately 10% of the cells that reside in the basal cell layer are stem cells, which are slow-cycling cells with the potential for continual cell division [42, 43]. When a stem cell divides, one daughter cell remains a stem cell in the basal layer, while the other daughter is destined to differentiate. The cells that will differentiate, termed transient-amplifying cells, will divide a limited number of times before differentiating and moving up in the epidermal layers. The post-mitotic differentiating cells account for 5–10% of the basal keratinocyte cell population and can be identified based on the expression of early differentiation markers, such as keratin 10 and involucrin [44, 45].

14.2.2.2 Stratum Spinosum

Adjacent to the stratum germinativum, the stratum spinosum consists of 8–10 rows of polyhedral keratinocytes that have begun to differentiate. Melanocytes and Langerhans cells can also be found interspersed throughout this layer [15]. The keratinocytes in this stratum are interconnected by desmosomal junctions, which are

supported intracellularly by bundles of cytoskeletal proteins termed tonofibrils [46]. This network structure gives the cells a characteristic "spiky" appearance following histological processing. The phenotype of the spinosum keratinocytes varies depending on their position within the layer. As the cells progress outwards, they lose their polygonal structure and develop a larger and flatter shape [2]. All of the keratinocytes still contain cellular organelles, including melanosomes secreted by the melanocytes [27]. In the upper layers of the stratum spinosum, lamellar granules can be observed within the keratinocytes. These granules contain proteins, such as keratohyalin, and other components that are involved in the final stages of keratinization [17]. Within this layer, the differentiating cells begin to synthesize keratin types 1 and 10. It is interesting to note that during wound healing the expression of these two proteins is downregulated in favor of keratin types 6 and 16 [47].

14.2.2.3 Stratum Granulosum

The 2–3 rows of flattened, nucleated cells in the stratum granulosum can be identified by the uniform presence of basophilic keratohyalin (lamellar) granules [2]. As previously mentioned, the proteins included in these granules function in the formation of the outer cornified cell layer. Keratohyalin has been shown to form an intracellular matrix that envelops and strengthens the cytoskeletal keratin filaments [48]. This keratin network is further reinforced by keratin aggregation and inter-filament disulfide bond formation, promoted by the protein filaggrin, whose precursor profilaggrin is found within the granules [2, 49]. Two other granular proteins, involucrin and loricrin, are involved in the formation of the cornified cellular envelope [50, 51]. The abrupt change from granulocytic phenotype to that of a terminally-differentiated corneocyte, is mediated within the outer layer of the stratum granulosum. These cells produce apoptotic enzymes that facilitate the destruction of the nucleus and all cellular organelles [52].

14.2.2.4 Stratum Lucidum

The stratum lucidum, found primarily in the thick skin of the palms and soles, consists of several layers of dead, clear, flattened keratinocytes that are enriched in keratin [24].

14.2.2.5 Stratum Corneum

The outermost layer of the epidermis, the stratum corneum, is composed of multiple rows of dead, terminally-differentiated keratinocytes that are continually sloughed off and replaced. These cells are responsible for the majority of the barrier properties of the epidermis and, consequently, the skin. The thickness of the corneal layer varies from 15 to hundreds of layers, depending on the site within the body [2, 15]. Variations in thickness can also be correlated to differences in sex, age, and disease state [24, 53]. In general, the corneocytes are large, flattened, polyhedral cells that are interconnected via desmosomal junctions at overlapping boundaries. Over 80%

of the cell content is composed of high-molecular mass keratins [17]. As in the other epidermal layers, changes in cell phenotype and function can be observed across the layer. The outer cell layer has flatter cells, which have more rigid, cornified cellular envelopes [51]. These exterior cells also secrete proteolytic enzymes that degrade the desmosomal junctions to allow for desquamation. Due to the presence of residual interconnections, corneal cells are typically shed in large groups [54].

14.2.3 *The Dermis*

The dermis is rich in ECM components that impart strength, elasticity, density and compliance to the skin. Within this region, a vast network of fibrillar and amorphous connective tissue supports the dermal cellular components, vascular and neural networks, and the appendage structures [55, 56]. Interactions between the dermis and the epidermis via the basal laminar region are critical to maintaining the proper phenotype and functioning of both regions [14, 57, 58]. In order to satisfy the metabolic needs of both the dermis and the avascular epidermis, the dermis is highly vascularized. The intricate networks of dermal blood vessels also function in the homeostatic regulation of the body temperature [59]. Neural networks of sensory and autonomic fibers present in the dermis function in the control and communication systems of the skin [60]. Lymphatic channels facilitate the controlled removal of water and other molecules from the region, and play an important immunogenic role [61]. Although the dermis contains fewer cells than the epidermis, fibroblasts, smooth muscle cells, pericytes, monocytes, macrophages, dermal dendrocytes, and mast cells can be found within this region [2].

As previously mentioned, dermal connective tissue is composed of both fibrous and non-fibrous components, which are continually undergoing turnover. Collagen is the most abundant protein in the dermis, accounting for approximately 75% of the dry weight of skin [2, 62]. The most common types of collagen found in skin are types I, III, and V, which can be classified as fibrillar collagens. More specifically, approximately 80–90% of the collagen in the dermis is of type I, whereas 8–12% is of type III. Collagen type V functions to regulate fiber diameter through interactions with types I and III. Fibril-associated collagen type VI is also found throughout the dermis, associated with type I collagen [63–65]. This network of collagen fibers imparts tensile strength to the skin. Elastic fibers, including elastin, form a network that extends throughout the dermis, spanning from the basal laminar region into the hypodermis. These fibers impart elasticity to the skin, allowing it to return to its original structure after experiencing distending forces [62, 66].

The non-fibrous connective tissue components of the dermis include proteoglycans (PGs), glycosaminoglycans (GAGs) and glycoproteins. These matrix components form the gel-like ground substance that surrounds the fibrous networks. PGs and GAGs found in the skin include biglycan, decorin, versican, perlecan, syndecan, heparan sulfate, and chondroitin-6 sulfate [2, 67–69]. Hyaluronan is also expressed in the embryonic dermis and during wound healing [70]. The ability of the PGs and

GAGs to bind water molecules contributes to the regulation of water loss, as well as to the maintenance of the dermal volume and compressive properties [71, 72]. These molecules also function to bind cytokines and mediate cellular attachment to other matrix components [73]. Glycoproteins, including fibronectin, laminin, vitronectin and tenascin, function in dermal cell-matrix and matrix-matrix interactions [2]. Overall, based in part on connective tissue structure, the dermis can be subdivided into two regions, the papillary dermis and the reticular dermis.

14.2.3.1 Papillary Dermis

The papillary dermis is the more superficial dermal region that is composed of small bundles of loose connective tissue, including collagen fibrils and immature elastic fibers [65]. This region is named for the papillae or projections of the dermis into the epidermal ridges. The matrix in this region, which hosts the majority of the dermal cells, is composed primarily of type III collagen [15]. The subpapillary plexus of arterioles and post-capillary venules serves as the lower separation boundary with the reticular dermis. The capillaries that supply the epidermis and upper dermal layers extend from this plexus into the papillary dermis [59]. Sensory and autonomic axons, as well as receptors, can also be found in this region [60].

14.2.3.2 Reticular Dermis

The lower and larger portion of the dermis, termed the reticular dermis, is composed of an interconnected collagen network, intertwined with a mature system of elastic fibers [55, 65, 66]. This region, which is connected to the papillary region by extended fibrous bundles, imparts the dermis with strength and resilience. Blood vessels, appendages, and dermal cells are interspersed throughout this connective tissue meshwork. The collagen bundles and elastic fibers increase in size in the deeper reticular layers approaching the hypodermis, which is distinguished from the dermis by the appearance of adipose tissue [2, 24].

14.2.4 The Dermal-Epidermal Junction Zone

The dermal-epidermal junction (DEJ) zone is characterized by a specialized basal lamina that is critical to the maintenance of proper skin structure and function [14]. This dynamic layer of interconnected proteins is approximately 100 nm thick and has been shown to have a critical role in skin remodeling, wound healing, and embryonic development. While the DEJ structurally separates the dermis from the epidermis, the proteins found within the basal laminar layer also mediate the attachment and interactions of these two regions [74]. For skin tissue engineering purposes, it is important to note that the DEJ zone maintains the overall structural integrity of the skin. Disorders affecting this region result in weak skin that is highly prone to shearing or tearing when exposed to applied forces [58]. The basal lamina also has critical functions in regulating keratinocyte cell polarity, proliferation,

differentiation, and migration. Further, this matrix also influences the barrier properties of the skin and can restrict the movement of molecules based on size and charge [14].

Three different types of anchoring complexes stabilize and strengthen the DEJ zone: hemidesmosomes, anchoring filaments, and anchoring fibers. Using microscopy techniques, these structures can be observed in four distinct regions within the DEJ [75]. The first identifiable region consists of the hemidesmosomes in the plasma membranes of the basal keratinocytes. These cell-matrix attachment structures are supported intracellularly by bundles of intermediate keratin filaments [17]. The next two regions, the lamina lucida and the lamina densa, contain the anchoring filaments primarily comprised of linkages between hemidesmosomal integrin $\alpha_6\beta_4$ and DEJ laminin type 5 [76–78]. The anchoring filaments firmly attach the basal cells to the underlying matrix, including the rich network of collagen IV fibers found in the lamina densa. The DEJ region adjacent to the dermis, termed the sublamina densa, contains collagen type VII anchoring fibrils [79]. These specialized complexes, originating in the lamina densa, firmly attach the DEJ to the dermis by either extending into anchoring plaques in the dermis or looping back into the densa region. The fibrils that form loops are interconnected with dermal collagen type I fibers, further stabilizing the structure [57, 58]. Overall, the DEJ maintains the integrity of the skin and regulates the dermal-epidermal interactions.

14.2.5 The Hypodermis

The hypodermis, also referred to as the subcutaneous layer, is located directly below the dermis. The primary component of this region is adipose tissue [2, 24]. Although a distinct border exists between the dermis and the hypodermis, the two regions are structurally and functionally interconnected through the rich vascular, neural, and lymphatic networks that supply the regions [59–61]. In addition, many of the appendage structures, including hair follicles and sweat glands, extend into the subcutaneous region [2]. The primary functions of the hypodermis include energy storage and release, insulation, mechanical protection, and the maintenance of body contours [80].

14.2.6 The Appendages

The epidermal appendages, including sweat glands, sebaceous glands, hair follicles and nails, arise from protrusions of the epidermis into the dermis [24, 81]. Consequently, these structures are lined with epithelial cells, which through growth and differentiation, are involved in the process of re-epithelialization following epidermal injury [82]. Moreover, the appendages have critical roles in the maintenance of homeostasis.

14.2.6.1 Sweat Glands

Sweat glands can be classified as either eccrine or apocrine, based on their structure and function. Eccrine sweat glands are composed of three parts: the coiled gland, the intradermal duct, and the intraepidermal portion. The coiled gland, located deep within the dermis, is lined with secretory cells that produce eccrine sweat. This sweat is delivered to the epidermal surface via the intradermal duct that extends through the dermal region, connecting to the intraepidermal portion, and ultimately terminating as an open pore [83]. These glands primarily function in the regulation of body temperature, through cooling by sweat evaporation, but are also involved in the elimination of wastes. In general, eccrine sweat glands are found over the majority of the skin surface and are most concentrated on the palms of the hands and the soles of the feet [84]. Although similar in structure to eccrine glands, apocrine glands have coiled regions that are approximately 10 times larger in diameter and intradermal ducts that terminate into the uppermost region of a hair follicle. Although the exact function of apocrine sweat glands remains unclear, it is believed that they function as scent glands related to sexual attraction. More specifically, these glands, which begin functioning during puberty, secrete small amounts of viscous sweat that becomes odorous when exposed to bacteria on the outer epidermis. Apocrine glands are most concentrated in the axillae and anogenital regions [24, 85].

14.2.6.2 Sebaceous Glands

Sebaceous glands are oil-producing glands that are found throughout the skin surface in varying sizes and densities, with the exception of the palms of the hands and the soles of the feet. In general, these glands are connected to hair follicles to form pilosebaceous units. A sebaceous gland consists of one or more branched lobules (acini) connected to a hair follicle by a short duct [13, 86]. The glands produce sebum, a substance composed of fats, cholesterol, proteins, and inorganic salts. Sebum protects the skin and hair, preventing these structures from becoming dry and brittle. This oily substance also aids in the prevention of excessive water loss due to evaporation from the skin surface [87].

14.2.6.3 Hair Follicles

While primarily vestigial in humans, the principal function of hair is to provide protection from injury and environmental conditions. Hair, consisting of a free shaft and a root structure, is supported in the skin by the hair follicle. Follicles extend at oblique angles into the deep dermis or hypodermis. The outermost layer of the follicle, the external root sheath, is composed of epithelium that is continuous with the epidermis. The inner root sheath surrounds the hair root, extending from the base of the follicle, termed the bulb, to the skin surface [88]. Progenitors in a region of the bulb referred to as the matrix, proliferate to form the hair shaft in alternating periods of growth (anagen) and rest (telogen). Also located at the base of the bulb

is the dermal papilla, an indentation of loose connective tissue, rich in vasculature, which supports the hair growth. Above the bulb, the follicular wall thickens to form the bulge, which contains stem cells believed to function in follicular regeneration and the replenishment of the matrix progenitors [43, 89, 90]. Smooth muscle bundles originating in the dermis, termed the arrector pilori, attach to the follicle at the bulge region and function in pilo erection [88]. The duct of the sebaceous gland associated with the follicle opens up into the channel above this region, closer to the epidermis [86].

14.2.6.4 Nails

Nails are modified keratin-rich epidermal cells that protect and support the tips of the fingers and toes. While primarily serving esthetic purposes in humans, deterioration in the condition of nails can often be indicative of more general health problems [24]. Nails are composed of a nail root, a nail plate, and a free edge. Growth of nails occurs from the matrix, which primarily underlies the nail root. Proliferating epidermal cells in the matrix are pressed tightly together into layers to create a plate that is slowly forced outwards, thereby forming the nail [91, 92].

14.2.7 Functions of the Skin

The skin, with its unique structure, performs a multitude of functions critical to the maintenance of total body homeostasis [2, 15]. Some of the key functions of this vital organ are detailed in Table 14.1.

Table 14.1 The functions of the skin

Function	Details
Thermoregulation	Controlled by the regulated dilation of the dermal blood vessels and the secretion of sweat
Protection	From a multitude of factors including:
	Mechanical shocks
	Thermal shocks
	UV irradiation
	Chemicals
	Water loss
	Ionic loss
	Infection
Metabolic organ	Vitamin D production
	Disposal of water, carbon dioxide, and toxins
	Glandular secretions from sweat and sebaceous glands
Sensation and communication	Touch, pain, heat and cold

14.3 Development of the Integumentary System

In order to design a successful strategy for skin regeneration following burn injury, it is important to have an understanding of the events that occur during the embryological development of the integumentary system. Epidermal development commences during the third week of gestation, shortly following the completion of germ layer formation by gastrulation. The single ectodermal layer surrounding the embryo is subdivided into the epidermal ectoderm, the neural ectoderm and the neural crest, through a process termed neurulation [93]. Underlying the epidermal ectoderm is the loose mesenchymal tissue that will ultimately develop into the mature dermis [94]. It is important to note that the basal lamina of the DEJ, which is fully developed by the ninth week of gestation, plays a critical role in skin morphogenesis [14].

14.3.1 The Epidermis

The development of the epidermal layers commences during the 4th week, with the proliferation of the epidermal ectoderm to form a secondary layer of flattened cells, termed the periderm. Signaling with the underlying dermis is critical to the formation of the epidermal layers during all morphogenic phases. By the 11th week of gestation, an intermediate epidermal layer can be observed between the germinative cells in the basal layer and the peridermal cells. During this period, the epidermal ridges also begin to form. As the basal cells continue to proliferate and differentiate, the appearance of the mature epidermal barrier commences around the 20th week of development. After this point, many of the peridermal cells undergo apoptosis and become detached from the fetal epidermis [94–96].

Unlike the keratinocytes, which arise directly from the superficial ectoderm, many of the other cells found in the epidermis migrate in from other regions [95]. Melanoblasts, precursors of the melanocytes, are derived from the embryonic neural crest and first migrate into the developing dermis during the 6th week. Mature melanocytes can be first detected in the epidermal region during the 10th week of gestation [25, 95]. The origin of Merkel cells is also believed to be the migrating neural crest cell population [2]. Langerhans cells, which differentiate from precursors in the bone marrow, migrate into the epidermis and first appear around the seventh week of development [60].

14.3.2 The Dermis

In general, the dermis is derived from the mesodermal cells underlying the epidermal ectoderm [95]. During the early stages of development, the neo-dermis is composed primarily of interconnected cells surrounded by a loose matrix of hyaluronan and glycogen [97]. Differentiation of the mesenchymal progenitors into mature fibroblasts commences during the third month of gestation. As the newly-formed

fibroblasts synthesize ECM components, the dermal matrix becomes more fibrillar. Vascularization of the dermis also begins during this time period [24, 94–96].

14.3.3 The Appendages

The appendages develop from regions of the epidermis that proliferate and extend into the dermis to form columns of epithelial cells surrounded by mesenchymal tissue. It is believed that both autocrine and paracrine signaling with the underlying mesoderm are critical in the formation of these ectodermally-derived projections, which first start to appear during the fourth month of gestation [24, 81, 95]. Ultimately, the development of each of the different types of appendages may be attributed to variations in the secretion and timing of inductive factors, which affect the proliferation, differentiation, and organization of the epithelial columns. Some factors believed to be important in follicular development include transforming growth factor-β (TGF-β), sonic hedgehog (SHH) protein, noggin, and members of the bone morphogenic protein (BMP) family [95].

14.4 Burns

14.4.1 Burn Classification

Burns are traumatic events that can severely disrupt the proper structure and functioning of the skin. Burns can be classified as either first degree, partial thickness, or full thickness based on the extent of the damage to the integumentary system (Fig. 14.3). Table 14.2 lists the general characteristics of each of these types of burns [98–100].

Superficial Dermal   Deep Dermal   Full Thickness
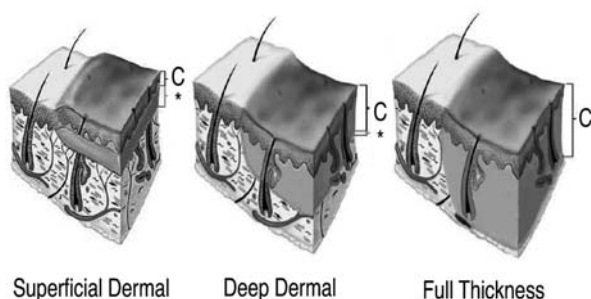
Fig. 14.3 The classification of partial and full thickness burns based on the extent of the damage to the integumentary system. "C" represents the necrotic tissues in the zone of coagulation and "∗" represents the regional edema in the tissues underlying the burn injury. (Figure adapted with permission from www.burnsurgery.org)

Table 14.2 The classification of burns

Burn classification	Characteristics
First degree	Loss of the epidermal layer
	Red, dry appearance
	No blistering
	Painful
	Will heal spontaneously
Partial thickness	Loss of the epidermal layer
	Partial loss of the dermal layer
	Red or pink appearance
	Blistering present
	Surface may be wet or waxy
	Painful
	Can be classified as either superficial or deep based on the degree of dermal damage
	Superficial dermal:
	Damage only to the papillary dermis
	Will heal spontaneously from the surrounding epidermis and the epithelial cells lining the appendage structures in the dermis
	Deep dermal:
	Damage extending into the reticular dermis
	Large amounts of necrotic dermal tissue (eschar)
	Healing impaired due to deep dermal damage
	Should be treated as full thickness burns
Full thickness	Loss of the epidermal layer
	Loss of the dermal layer
	Damage to the hypodermis that may extend to involve the underlying structures including muscle, tendon and bone
	White or charred appearance
	Not painful due to neural destruction
	Eschar widely present
	Granulation and contraction of the wound to form scar tissue

14.4.2 Principles of Burn Wound Healing

Normal wound healing is characterized by three stages of healing: the inflammatory phase, the proliferative phase, and the remodeling phase [101]. One of the primary events in the inflammatory phase is the accumulation of platelets at the site of injury, leading to blood coagulation and clot formation, which provides a preliminary matrix for cellular infiltration. Inflammatory cytokines secreted by these aggregating cells regulate the regional blood flow and have a multitude of stimulatory effects, including the promotion of leukocyte invasion and angiogenesis. The second phase of wound healing, the proliferative phase, generally commences within 24 h. Rapid re-epithelialization to achieve wound closure is facilitated through the migration of epidermal cells from the wound edges and from the residual appendage structures in the injured region (Fig. 14.4) [101–103]. This process is stimulated by
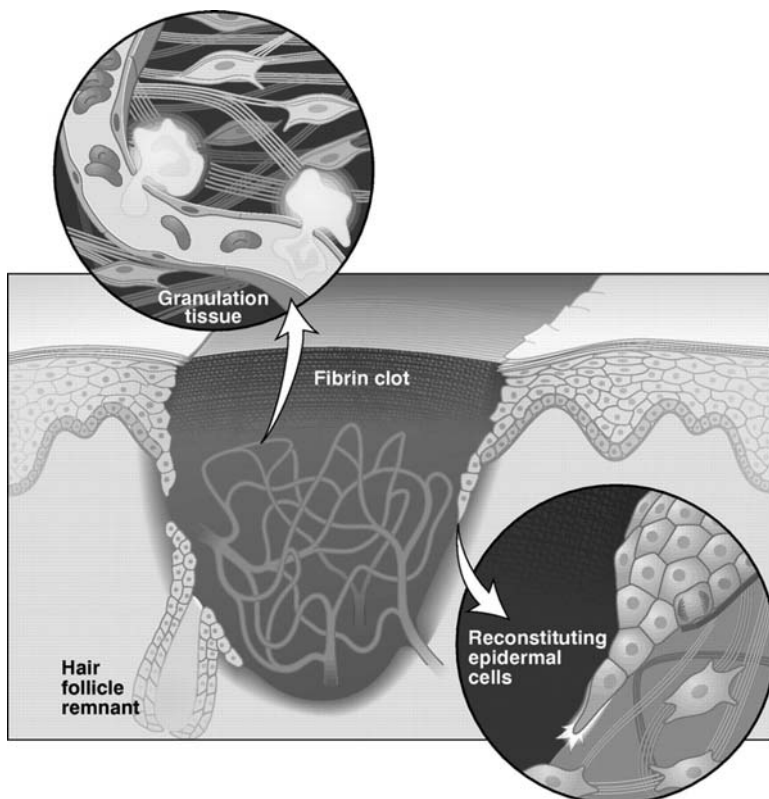
Fig. 14.4 The proliferative phase of the wound healing process in skin. A preliminary fibrin matrix for cell migration is formed during the blood coagulation and clotting processes following the initial injury. Invading inflammatory cells secrete growth factors that promote angiogenesis and dermal fibroblasts deposit extracellular matrix components, forming granulation tissue. Re-epithelialization occurs through the migration of epidermal cells from the wound edges and from the residual epidermal appendage structures, such as hair follicles. (Figure reprinted with permission from Martin [103]. Copyright 1997 by the American Association for the Advancement of Science)

growth factors, including TGF-β and epidermal growth factor (EGF). Angiogenesis also occurs and is promoted by the secretion of FGF by the macrophages that are localized in the wound site [101, 104, 105]. A loose ECM network, deposited by the fibroblasts that have invaded from the surrounding tissues, encompasses the newly-formed blood vessels. This provisional matrix, termed granulation tissue, provides a substrate for cell migration, facilitating cellular repopulation and wound closure. The final stage of wound healing, the remodeling phase, requires months to complete. In the initial period, the fibroblasts assume a myofibroblastic phenotype and, stimulated by TGF-β and other factors, cause the contraction of the wound, expediting closure [101, 104]. Ultimately, the dermal ECM is modified with newly-synthesized matrix components, resulting in scar tissue formation [101–103].

Severe burn wounds can be classified into three regions of varying tissue injury: the zone of coagulation, the zone of stasis, and the zone of hyperemia [100, 106]. The zone of coagulation is characterized by the complete dissolution of all levels of protein structure. Further, cell death within this region, which extends downwards from the original site of injury, is uniform and irreversible. Clinically, the zone of coagulation presents as the burn eschar [107]. Adjacent to the zone of coagulation is the zone of stasis, a region in which there is a progressive deterioration in the local blood flow. Although the majority of the cells within this zone are initially alive, many die due to ischemia. Numerous factors combine to cause the impaired blood flow, including clotting events, burn edema, vasoconstriction, and heat damage to the cells and proteins [100]. Without intervention, the tissues in this region will die, thereby expanding the zone of coagulation [69]. The zone of hyperemia is located at the periphery of the injury, adjacent to the unaffected body tissues. Within this region, there is minimal damage to the protein structures and cells. In general, complete healing of these tissues is possible. Due to local inflammatory responses and the release of numerous vasoactive molecules, increased vascular flow rates and the dilation of the local vessels characterize this zone [7, 100].

In the case of burn injuries, the process of wound healing can be dramatically modified or impaired. Skin damage can be extensive, with the complete destruction of all of the skin layers over a large surface area of the body. First degree or superficial partial thickness burns will generally heal normally, with re-epithelialization occurring from the wound edges and from the undamaged appendage structures. In contrast, in deep partial thickness or full thickness burns, the appendages are destroyed and re-epithelialization can only occur from the wound periphery. The process of cell migration can be further delayed if the burn injury cauterizes the local blood vessels, inhibiting the initial cascade of events during the inflammatory healing phase that creates a temporary fibrin matrix for cellular invasion [82]. Hence, as the coverage process may require an extremely long time frame and the dermal regenerative potential is limited, severe wound contraction and scarring are frequent outcomes. Hypertrophic scarring and impaired functionality are very common in burn patients. In some cases, keloid scars, which extend beyond the original boundaries of the injury, can be observed. In general, contracted and scarred tissues can be painful, as well as having poor functionality and cosmesis [108–110]. Complications associated with burn wound edema can also impact the wound healing process. Increased vessel permeability and the action of numerous inflammatory cytokines, such as prostaglandins, can lead to severe swelling that impairs the regional blood flow, particularly in the zone of stasis, thereby causing ischemic cell death [111, 112].

14.4.3 Immune System Response to Burn Injury

Suppression of the immune system has long been recognized as a major contributing factor to the high mortality rates that are associated with burn injuries. Patients with severe burns are extremely susceptible to infections that can result in death due

to septic shock or multiple organ failure [113, 114]. Clinical and research findings indicate that the impaired immune response is related to the suppression of both the cellular and humoral systems. More specifically, burn patients have been reported to have reduced lymphocyte populations, including decreased numbers of T-cells, B-cells, and natural killer cells [115–117]. In general, following burn injury, there is a failure in the T-cell response related to both diminished cell number and cell functionality [118]. Dysfunctions in the neutrophil cell population and the complement system further compound the immunosuppressed condition [119].

Numerous factors appear to contribute to the impairment of the immune system in burn patients. While the precise mechanisms of action are not fully understood, the systemic inflammatory response (SIR) associated with burn injury is hypothesized to play a role in the development of the immune dysfunction [111, 120]. Research continues to be conducted in order to elucidate all of the elements that are produced in this response, including cytokines and toxic by-products such as endotoxin, tumor necrosis factors (TNF) and leukotrienes, which could be immunosuppressive in nature [121]. One compound of interest, cutaneous burn toxin (CBT), is suspected to play a major role in the inhibition of granulocyte production, thereby accounting for the reduced lymphocyte populations. CBT is composed of polymers of lipids and proteins that are formed from the damaged tissues at the burn injury site [120]. Proper debridement of the wound can help to reduce the impact of this toxin. It is also important to note that the presence of necrotic tissue, impaired blood flow and nutritional deficiency can all have an impact on general health and immune function [122].

14.4.4 Complications

While the skin is the focus of this chapter, the systemic effects of burn injuries must also be recognized. As previously mentioned, infection and SIR are critical problems in burn patients, frequently resulting in major organ failure and death. Following injury, the release of a multitude of inflammatory cytokines and toxic by-products can severely impair all of the organ systems, including the cardiac system, renal system, and digestive system [123]. Complications can also arise from abnormally low blood pressures due to hypovolemia, attributed to a loss of blood from the damaged vessels surrounding the wound site [112]. Another major cause of mortality is the severe respiratory damage and distress associated with inhalation injuries in burn patients [124].

14.5 Conventional Treatment of Burns

14.5.1 Treatment of Minor Burns

As previously mentioned, first degree and superficial partial thickness burns will spontaneously heal within a short time period, ranging from 7 to 21 days [82, 98].

Basic first aid treatment of the wound will help to expedite the healing process and prevent infection. The injured site should be gently and thoroughly cleaned with lukewarm water and mild soap. Cool, moist compresses can be applied to reduce pain and swelling. While it is the subject of some controversy, it is generally accepted that the blisters of superficial burns should be left intact, to reduce the chances of bacterial infection and minimize discomfort. Spontaneous resorption of the blisters should commence within a week. With the exception of injuries to the face, a clean, absorbent, non-adherent dressing, such as lint-free gauze, should be applied to cover the wound. This dressing should be changed daily or as required. The application of topical antibiotics or other agents is generally not recommended, as they can impede the wound healing process and have other negative side effects [125]. Burn wounds to the face are typically left uncovered, although a thin layer of topical antibiotic can be applied for prophylactic purposes and to prevent against discomfort due to tissue dryness [106]. To control the pain that is frequently associated with superficial burns, oral analgesics may be indicated [126].

14.5.2 Primary Treatment of Severe Burns

Following stabilization, patients with deep dermal and full thickness burns should be sent to the hospital for immediate assessment and care. The maintenance of airway, breathing, and circulation is of primary importance to patient survival. During transport, the burn wound should be covered with clean towels, which can easily be removed for examination purposes. The patient should be kept warm and treated for shock [127]. Intravenous (IV) infusion should be commenced as early as possible to maintain the circulating volume and reduce the risk of damage to the cardiac, renal, and other major organ systems [128]. Upon arrival at the hospital, a team of burn care specialists will assess the severity of the burn, examining the depth of the injury and estimating the percentage of the total body surface area (TBSA) affected [129]. Morphine, delivered via IV, is often prescribed for pain management purposes [126].

Immediate burn wound treatment is required to aid in healing and reduce the risk of patient mortality. The wound site should be washed with warm saline solution and debrided to remove dead epithelial tissue and foreign contaminants. In general, as the seepage of blister fluid can provide an ideal medium for bacterial growth, blisters are evacuated and trimmed. This process also improves joint mobility and aids in the visual assessment of the burn [106]. In cases of circumferential burns, such as around the neck, torso or limbs, a decompression escharotomy may be required to relieve the tension in the tissues. More specifically, as burn edema progresses, the inelastic eschar at the surface does not expand well, causing increased pressures that can result in ischemic conditions, impaired respiration, or tissue rupture. These situations can be prevented by the escharotomy, which involves a strategic, longitudinal incision along the length of the affected tissues [130]. Following debridement, topical antimicrobial agents, such as silver sulfadiazine, may be applied to the surface of the wound to slow or reduce the occurrence of bacterial infection [131]. As the

blood supply to the burn wound is often severely impaired, the use of systemic antibiotics may be therapeutically ineffective [132]. In conservative approaches to burn treatment, an occlusive dressing or other wound covering, discussed later in the chapter, is then placed over the affected tissues prior to further treatment, such as autografting, if required.

Through advances in burn research, it is now accepted that the early surgical removal of the burned tissues is of benefit to the patient, in terms of pain management, immunocompetence, and wound healing. Excision of the necrotic tissues has been shown to reduce the total time of hospitalization, the chances of infection, the amount of surgical intervention required, and the instances of patient mortality. Improved functionality and cosmesis following wound healing have also been reported [8]. By limiting the systemic exposure to the inflammatory cytokines and toxic byproducts found in the necrotic tissues, it is possible to minimize immunosuppression and reduce the risk of major organ failure [122]. In general, the procedure involves the excision of all of the necrotic tissues, followed by the immediate application of either an autograft or a temporary wound covering. The operation is performed by tangential excision using a specialized dermatome. The necrotic tissue is removed in layers until only viable tissue remains, detected when bleeding is observed [8, 133]. Full thickness burns can be excised to either the hypodermis or the muscle fascia. While autograft acceptance is more limited on the subcutaneous adipose tissues, when successful, a superior esthetic result is achieved with this more conservative approach [8]. Overall, the major restriction to the surgical excision of burn wounds is the potential for severe blood loss, which limits the amount of tissue that can be removed at one time [9].

14.5.3 Autografting: The Current Gold Standard

The use of autologous grafted skin represents the gold standard treatment for deep dermal and full thickness burns. To date, this strategy is the only method available to achieve complete and permanent wound closure in extensive or deep burns [1, 8, 104]. Partial thickness grafts, ranging in thickness from 0.2 to 0.5 mm, are harvested from healthy donor sites using a dermatome [9]. Ideally, the mechanical properties, dimensions and pigmentation of the skin in the donor and recipient regions are similar. As only a small portion of the dermis is extracted, the donor site defect will heal spontaneously without scar tissue formation in 2–3 weeks [107]. The autografted tissue is surgically attached to a freshly excised, uncontaminated wound site. As it is extremely thin, the graft can survive the initial ischemic period prior to vascularization via diffusion of nutrients and waste. Following the surgery, the bond between the grafted skin and the wound bed is extremely weak and must be protected from disruption due to shearing, edema, or hematoma. As healing progresses, the graft fully integrates into the wound site, inhibiting contraction and restoring functionality. In most cases, graft acceptance is high and scar tissue formation and contraction are reduced, achieving an acceptable esthetic result [8]. However, the deep

dermal hair follicles, sweat and sebaceous glands do not regenerate in the transplanted tissue. Reinnervation and limited restoration of skin sensation is possible [11]. Disadvantages of the use of autologous tissues include the creation of a donor site defect, patient risk, susceptibility to infection, and the need for multiple surgical procedures [8].

The development of new methods of permanent wound closure is mandated for the treatment of patients with severe burns involving a large surface area of the body. During a single surgical procedure, sufficient autologous tissue can often be harvested to treat burns ranging from 30% to 50% TBSA. In more extensive burns, the use of autografting is limited by a lack of donor tissue [4]. Treatment in these cases may be facilitated by a modification to the autografting technique involving the use of a tissue mesher [134]. Small linear incisions are made in the harvested tissue to allow for the expansion of the graft. In practice, the meshed skin can be expanded 1.5–9 times its original size. When expanded more than three times, it is common to cover the site initially with a meshed allograft to improve healing and provide protection during re-epithelialization [9]. While meshing allows for improved drainage and greater coverage, this skin is more prone to infection and has esthetic limitations. Due to contraction and scar tissue formation, the meshing pattern on the skin surface is permanent. Further, tissue availability is still limited using this technique. As medical knowledge and technology continue to improve, an increased number of patients are surviving from extremely large surface area burns (>90% TBSA). In these cases, autograft availability is minimal and generally insufficient. Following the completion of healing, it is possible to re-harvest the donor and previously-grafted sites. This process can be repeated a limited number of times, and increases patient pain and susceptibility to infection, as well as treatment time and expense [8, 9]. Temporary wound coverage materials must be utilized during the interim healing period. Ultimately, alternative means to autografting may be required to achieve complete wound closure and severe scarring may be inevitable [135].

14.5.4 Biological Alternatives for Temporary Wound Coverage

As previously discussed, due to the limitations in the availability of autologous tissues, substitute materials to provide temporary wound coverage following early excision are required. Conventional biological approaches to achieve this goal involve the use of allografts and xenografts. However, in order to achieve permanent wound closure and healing without severe scarring and contracture, these natural materials must ultimately be replaced with autologous tissues [12].

14.5.4.1 Allografts

To date, allogenic strategies have focused on the use of cadaveric skin and amnion.

Cadaver Skin

Despite numerous limitations, allogenic cadaver skin remains the most successful and commonly used substitute for temporary wound coverage in severely burned patients [7]. Allogenic skin aids in the restoration of the barrier properties, thereby protecting against excessive water and electrolyte loss, infection and pain. Moreover, allografts have been shown to initially reduce any existing bacterial contamination, stimulate re-epithelialization from the periphery, improve the ultimate esthetic appearance of the patient, and increase autograft acceptance [136]. However, in spite of the immunosuppressive nature of burn injuries, graft rejection within 1–3 weeks is inevitable. During this process, the wound site can become severely inflamed and prone to infection [137]. There are also concerns associated with the transmission of viruses, cost, and tissue availability [9]. The tissues can be utilized while fresh, preserved in glycerol at 4 °C, lyophilized, or cryopreserved. Fresh tissues have been shown to more strongly stimulate angiogenesis, have superior adherence properties, and provide greater protection against bacterial invasion. However, glycerol preservation is relatively economical, decreases the tissue antigenicity, and facilitates the establishment of tissue banking [138, 139].

Amnion

The use of fresh human amniotic membranes has been reported for decades and remains popular in regions of the world where the use of cadaveric skin is limited due to expense and availability [140]. Human amnion has low antigenicity and is reported to readily adhere to the wound site, reducing patient pain, accelerating the wound healing process, and repressing bacterial growth. In addition, amnion has been shown to stimulate angiogenesis, ultimately improving autograft acceptance [141]. Despite these positive attributes, the use of amnion is limited by the fragile nature of the tissues. Within several days of application, the amnion becomes fragmented and dissolves, necessitating almost daily replacement. Hence, the quantities of tissue required for large TBSA burns would be exceedingly high. Further, an occlusive dressing is required to prevent desiccation of the membranes and maintain the skin barrier properties [7, 142]. However, human amnion may be of use to expedite wound healing in superficial partial thickness burns, autologous donor sites, or highly-meshed autologous tissues [143].

14.5.4.2 Xenografts

The primary motivation for the use of xenogenic tissue is the limited supply and high costs associated with allogenic cadaver skin [9]. Frozen porcine grafts are most commonly investigated due to the similarity of the tissues with human skin [144]. Despite the potential for unlimited availability, numerous concerns restrict the use of xenografts in burn wound treatment. The primary problems are associated with the antigenic nature of the tissues. All grafts are rejected within a 14-day period [145]. There are also serious safety and sterility issues related to animal to human viral

transmission [146]. Furthermore, xenografts have been shown to be less adherent than allografts and are associated with higher rates of infection and necrosis [136].

14.6 Burn Dressing Biomaterials and Tissue Engineering

Through the creation of unique biological substitutes, tissue engineering holds great promise for the development of novel treatment strategies for numerous diseases, disorders, and traumas. Unlike conventional therapies, through tissue engineering investigators seek to restore or regenerate the affected regions of the body, with the ultimate goal of creating functioning, healthy tissues that are fully integrated into the host system [147, 148].

To date, significant research has been conducted into the fields of burn dressing biomaterials and tissue engineering. In general, the two design strategies for skin substitutes are to create materials for either temporary wound coverage or permanent wound closure [10]. Wound coverage materials have been widely studied in both the research and clinical settings due to their greater simplicity in design. The principle behind these materials is to create an interim barrier until sufficient autologous tissues are available for grafting. Wound coverage materials have been shown to alleviate patient pain, reduce the likelihood of infection, and promote wound healing [135, 149]. In contrast, the objective of wound closure strategies is to design a material that will become incorporated into the wound site, promoting healing and restoring skin functionality without tissue autografting [135]. Such a material would be of great benefit to patients with limited autograft availability, and could also eliminate the creation of donor site defects and the need for multiple surgical procedures.

14.6.1 Design Criteria

There are numerous criteria for the development of a successful skin substitute. Ideally, the construct should minimize patient pain and promote wound healing without scar tissue formation or contracture. The complete structural and functional restoration of the integumentary system and all of its components are long-term objectives. Some specific design considerations include adherence, barrier properties, mechanical properties, biodegradability and immune response, surgical handleability, and expense [107, 150–152].

14.6.1.1 Adherence

To reduce the occurrence of infection and minimize fluid accumulation at the tissue interface, it is extremely important for the skin substitute to readily adhere to the wound site [153]. The rapid and uniform attachment of the construct to the wound bed attenuates patient pain and creates a hypoxic environment that stimulates wound

healing. Hydrophilic materials that adhere to the wound site without inflicting damage on the underlying tissues are of particular interest [107].

14.6.1.2 Barrier Properties

Wound coverage or closure materials should restore the barrier functions of the integumentary system. The construct should protect the fragile tissues at the wound site, inhibit bacterial infection, and prevent desiccation [136]. The flux of water across the material must be controlled, to prevent excessive water and electrolyte loss, while also limiting fluid accumulation at the wound site [154]. Protein loss should also be restricted and normal rates of heat conduction should be observed across the barrier [136].

14.6.1.3 Mechanical Properties

Ideally the substitute should have mechanical properties similar to those of healthy human skin. The use of components that are too rigid could cause damage to the surrounding tissues, resulting in inflammation and subsequent device failure [155, 156]. Skin equivalents must be pliable, flexible, elastic, and durable. It is important to have a strong material, resistant to applied forces, which will conform and adhere to the wound site [7]. The material porosity is also an important design consideration. As previously mentioned, the construct must provide the barrier properties of the skin, including the controlled flux of water. However, a porous material is desirable to enable cellular infiltration and repopulation of the wound site from the surrounding tissues. To facilitate cell migration, a minimum pore size of $10\,\mu m$ is required and the material must be cell adhesive [107].

14.6.1.4 Biodegradability and Immune Response

The skin substitute should be both non-toxic and non-antigenic in nature, evoking a minimal immune response [157]. Moreover, the device should stimulate wound healing and host cell invasion from the periphery. As healing progresses, the controlled degradation and replacement of the construct with healthy host skin would be ideal [152]. This regeneration would eliminate the long-term problems associated with the restriction of growth at the wound site, a severe concern for burned children, allowing the healed regions to expand normally with the individual [135]. The products of the degradation process must also be compatible with the host system.

14.6.1.5 Surgical Handleability

To be a clinically viable alternative, the skin equivalent must be able to be sterilized, cut into a variety of shapes and sutured into position with relative ease by the surgeons in the operating room [107].

14.6.1.6 Expense

An ideal skin substitute should be relatively inexpensive and widely available for use [152]. The materials should not be limited and the fabrication process should be straightforward and reproducible. The construct should have long-term stability so that it can be preserved and stored for an extended time period [157].

14.6.2 Skin Substitutes

A wide variety of skin substitutes have been investigated in the laboratory and clinical settings to promote regeneration, rather than repair, following burn injury. The constructs can be classified as epidermal substitutes, dermal substitutes, or composite materials that include an epidermal and a dermal component [1].

In each of the strategies, both synthetic and natural materials have been studied. Synthetic materials can be selected that are readily available, uniform, sterilizable, biodegradable, and cost-effective. Moreover, modulating synthesis conditions can often control the physical and chemical properties of these materials [158–160]. Natural constructs have been shown to promote cell infiltration and host integration, reduce scar tissue formation, and, depending on processing, trigger minimal immune response. Scaffolds derived from the ECM can degrade within the body and promote matrix remodeling, thereby augmenting the regeneration of damaged or missing tissues [161, 162].

The incorporation of allogenic cells is a design feature in many skin substitutes. Cultured allogenic keratinocytes and fibroblasts derived from neonatal foreskins initiate a minimal immune response in the immunosuppressed burn patients [104]. Unlike the Langerhans cells present in the allografted tissues, these cells do not express MHC class II antigens. As a result, the production of allogenic T-cells is not stimulated via this means, although the secretion of cytokines by the cells, such as α-interferon, can cause a more limited reaction [10]. Overall, due to the immunological properties of the cultured keratinocytes and fibroblasts, these cells can be included in the constructs with low concern for major immune response or sensitization.

14.6.2.1 Epidermal Substitutes

Various epidermal substitutes are available that can be used for temporary wound coverage, restoring the barrier properties of the skin. When applied to superficial partial thickness burns or autologous donor sites, these constructs can often expedite the wound healing process [135]. In deep dermal or full thickness burns, it is generally recognized that the reconstruction of the underlying dermis is required prior to the replacement of the epidermis, in order to achieve acceptable functional and cosmetic results [1].

Occlusive and Semi-Occlusive Dressings

Many different occlusive and semi-occlusive dressings have been developed and marketed for the temporary wound coverage of superficial burns and donor sites. Although a wide variety of materials have been investigated, all of these constructs have a similar function. Occlusive dressings create a moist, hypoxic environment at the burn site. This has been shown to reduce patient pain and promote wound healing [163–165]. Ideally, these constructs should restore some of the barrier properties of the skin, regulating the flux of water and preventing against infection [166]. However, many of the substitutes are associated with high rates of bacterial proliferation in the moist wound bed, resulting in deep dermal damage and tissue necrosis. Further, the accumulation of fluids below the occlusive dressings can injure the surrounding tissues and lead to device separation. Other concerns are related to poor adherence and the inability to observe the wound site during the healing process [167–169]. Table 14.3 outlines some of the specific advantages and disadvantages of some common occlusive and semi-occlusive dressings.

Cultured Autologous Keratinocytes

Techniques have been developed to facilitate the expansion of autologous keratinocytes from small biopsy samples. Keratinocyte sheets can be grown in vitro on feeder cell populations of irradiated murine 3T3 fibroblasts, supplemented with growth factors [189]. This process is expensive, labor intensive, and time consuming. The 3–5 weeks required to produce $1.8\,m^2$ of confluent keratinocytes from a $2\,cm^2$ biopsy sample limits the clinical applicability of this strategy [10]. Genzyme Corporation specializes in the growth of autologous keratinocytes, marketed under the trade name Epicel^{TM} [190]. Using the enzyme dispase, the undifferentiated monolayer of cells can be removed as a sheet from the tissue culture dish. This separation process can impair the subsequent proliferation of the keratinocytes [191]. The fragile sheet can be transplanted onto the wound bed, but is susceptible to damage and desiccation. Acceptance of the autologous cells is dependent on the condition of the wound bed, with a partially intact dermal layer required to avoid complete sloughing. If the grafting is successful, the resulting epidermis is unstable and has limited barrier properties [1]. Within 6 days of transplantation, the cells begin to differentiate into the normal epidermal layers. However, the absence of the DEJ can lead to blister formation and separation. With extended time, the basal lamina characteristic of the DEJ can begin to form [10]. The skin grown from these cultured cells will be incomplete, lacking a full cellular complement and all appendage structures [132].

Cultured Allogenic Keratinocytes

Allogenic keratinocytes have been investigated as a possible means of avoiding the culturing time delay associated with the use of autologous cells [10]. As previously discussed, the immunologic nature of the allogenic keratinocytes limits the concerns

Table 14.3 Occlusive and semi-occlusive burn dressing biomaterials

Material	Advantages	Disadvantages
Hydrophilic polyurethane films (Opsite™, Omiderm™) [154, 170–173]	Generally impermeable to bacteria Low rates of infection Reduces pain Increases re-epithelialization rate Elastomeric	Low water permeability results in fluid accumulation in the wound bed Poor surgical handleability Problems with adhesion
Poly(ether urethane) [154, 174–176]	Similar to polyurethane films, but with improved water permeability to reduce fluid accumulation	
Polyvinyl chloride (PVC) films [7, 177, 178]	Flexible Good tensile strength Reduces pain Semi-permeable to water Permeable to oxygen and carbon dioxide Easy to change Relatively inexpensive	Low water permeability results in rapid fluid accumulation in the wound bed Wounds prone to infection
Poly(2-hydroxymethyl methacrylate) (pHEMA) with polyethylene oxide (PEO) films (Hydron™) [178–180]	Water soluble and vapor permeable Semi-flexible Conforms readily to wound site Transparent Stable	Poor adhesion on large wound sites Low tensile strength Difficult to apply Wounds prone to infection Limited clinical data
Acrylic copolymer with gelatin and glycerol [178]	Good surgical handleability Good water permeability	

(continued)

Table 14.3 Occlusive and semi-occlusive burn dressing biomaterials

Material	Advantages	Disadvantages
Hydrocolloid films (i.e. Duoderm™ – carboxymethyl cellulose, gelatin, pectin, polyisobutylene) [181, 182]	Promotes wound healing and re-epithelialization Easy to change Low rates of infection	Impermeable to oxygen and water vapor Non-adherent Fluid accumulation will result in separation from the wound bed
Spray-on polymers (i.e., acrylic films, polycaprolactone) [178, 183]	For use in first aid applications	Poor adhesion and coverage Limited protection from bacteria and water loss
Fibrin glues and films [38, 184, 185]	Promotes wound healing and re-epithelialization Can be used as a delivery system for keratinocytes and growth factors May protect against infection	Risk of animal to human viral transmission Risk of sensitization Minimal mechanical protection Expensive Storage concerns
Xenogenic collagen [161, 178, 186, 187]	Pliable, soft Permeable to water Stimulates wound healing and re-epithelialization Cost-effective Readily available Transparent	Poor adhesion Wounds prone to infection Low tensile strength Prone to desiccation May promote wound contracture Risk of animal to human viral transmission Risk of sensitization
Foams and sponges (i.e., polyvinyl alcohol (PVA), collagen, silicone, polyurethane) [7, 178, 183, 188]	Absorbent Provides mechanical cushioning Relatively inexpensive	Limited clinical data Fluid accumulation can result in separation from the wound bed Removal of the construct can damage the healing wound bed and cause pain Wounds prone to infection

associated with host immune response [104]. The sheets could be pre-grown, cryopreserved, and stored at the hospital for immediate use. Similar to the autologous sheets, the acceptance of the allogenic cells is dependent on the state of the wound bed. Further, there are also limitations associated with the fragility of the sheets and the lack of a DEJ [1, 10]. Clinically, improved healing has been shown in superficial burns and donor sites, attributed to the cellular secretion of various stimulatory cytokines and growth factors [192]. With time, host cells can replace the donor keratinocytes in the grafted region [193].

Laserskin^{TM}

Laserskin^{TM} is an epidermal substitute developed by Fidia Advanced Biopolymers that is composed of hyaluronan and autologous keratinocytes [194]. Hyaluronan has been shown to promote cellular migration, proliferation, and the formation of new blood vessels [195]. The cells are seeded in a series of small pores that are laser drilled in the hyaluronan gel (Fig. 14.5a). This device has the advantage of being able to be removed from the culture dish without the need for enzymatic digestion [10]. An intact dermal layer may be required to achieve acceptable functional and cosmetic results. Initial clinical investigations in patients with diabetic ulcers, show that the construct is durable and well-accepted, with low rates of infection [161, 196].

(a)   (b)
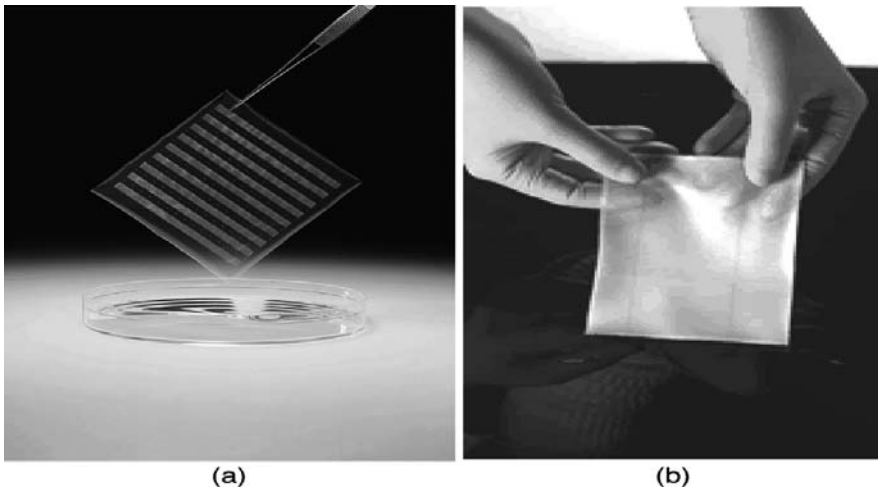
Fig. 14.5 Examples of tissue-engineered skin substitutes. **a**. The epidermal substitute Laserskin^{TM}, developed by Fidia Advanced Biopolymers, is composed of autogolous keratinocytes seeded in pores drilled by a laser into a hyaluronan gel. (Figure reprinted with permission from Fidia Advanced Biopolymers s.r.l.) **b**. The Integra^{TM} composite skin substitute composed of gluteraldehyde cross-linked bovine collagen type I and chondroitin 6-sulfate, covered with silicone. (Noncopyrighted source: U.S. Food and Drug Administration)

14.6.2.2 Dermal Substitutes

The creation of a dermal substitute for the treatment of deep dermal and full thickness burns is mandated by the numerous limitations associated with the use of autologous tissues. Epidermal substitutes are contraindicated in severely burned patients due to the problems associated with skin fragility, blistering, and poor graft integration [132]. In humans, damaged dermal tissue can result in severe scar tissue formation and contracture if left untreated [104]. An engineered dermal substitute has the potential to be used in combination with cultured keratinocytes or extremely thin autografts to reconstruct the injured skin [1, 135].

Dermagraft[TM]

Dermagraft[TM] is a dermal substitute, developed by Advanced Tissue Sciences Incorporated and now distributed by Advanced BioHealing Inc., which is composed of viable neonatal allogenic fibroblasts seeded onto a polyglactin (Vicryl) scaffold. The cells, which are expanded in vitro to confluence, secrete numerous growth factors and ECM components into the mesh [197]. A sufficient number of fibroblasts can be extracted from one foreskin to seed over $20,000\,m^2$ of the construct [198]. Following fibroblastic cultural expansion, the device can be cryopreserved without severely impairing the cellular viability. Trials indicate that the construct promotes wound healing, angiogenesis, and re-epithelialization from the wound periphery [199]. In one specific study, it was well-tolerated as an implant, with no signs of immunological rejection observed in 400 chronic wound patients [198]. The polymeric scaffold will degrade by hydrolysis as regeneration progresses, although the rate of degradation may cause complications. As the device does not include an epidermal component, wound closure is not facilitated. However, the construct can be used to improve functionality and cosmesis in combination with meshed autografts [200, 201]. While the potential exists to seed the dermal substitute with cultured autologous keratinocytes, preliminary trials suggested that cell adhesion may be a problem, with a mean acceptance rate of only 51.2% [202].

Alloderm[TM]

Alloderm[TM] is a dermal substitute developed by LifeCell that is composed of acellularized human cadaveric skin [203]. Extraction of all cellular components while preserving the dermal matrix structure, including the basal lamina of the DEJ, eliminates the immunological complications associated with the use of allogenic skin, but promotes proper tissue organization during regeneration [1, 204]. Studies have indicated that Alloderm[TM] readily adheres to the wound bed, promoting the invasion of cells and blood vessels, and fully integrating with the host system. As it has very limited barrier properties, the scaffold must be grafted with an ultra-thin autograft when it is implanted. Overall, the construct has been reported to reduce scarring and improve skin elasticity and cosmesis [205]. Due to the limited availability of allogenic tissues, acellularized xenogenic skin has also been investigated [161].

14.6.2.3 Composite Substitutes

Composite skin substitutes hold great potential for the treatment of patients with severe burns. With further research, these constructs may be able to replace all of the skin layers, eliminating the need for tissue autografting. One important design note is that it is imperative to reconstruct the basal lamina of the DEJ zone in order to create a strong and well-integrated material [10].

Integra^{TM}

Integra Life Science Corporation markets the most generally accepted composite skin substitute, originally described by Yannas and Burke (Fig. 14.5b) [206, 207]. This device has a dermal component of gluteraldehyde cross-linked bovine collagen type I and chondroitin 6-sulfate, covered with a silicone artificial epidermis [208]. When implanted, the collagen and GAG matrix, with an average pore size of 50 μm, encourages cellular invasion and fibrovascular regeneration to form a normally-organized neodermis within 3–6 weeks [10, 107]. The original scaffold gradually degrades over a 30-day period [161]. The silicone membrane has good surgical handleability and regulates the flux of water at the normal regional rates observed in the skin [1]. When implanted, the average acceptance rate (AAR) is generally over 80%, lower than the 95% AAR for autografts, but comparable to that observed initially in allografts. Moreover, the substitute is well-tolerated by the immune system [135]. Within 3–4 weeks, the silastic layer can be removed and an ultrathin (less than 0.1 mm) autograft can be placed directly on the developing dermis, with an AAR of 90% [202]. The creation of smaller donor site defects expedites and improves healing. Clinical results show that the use of Integra^{TM} promotes wound healing, reduces contracture, and prevents hypertrophic scar formation. Superior cosmetic results are reported to the use of autografting alone, along with good joint functionality and the ability to grow in children [209]. The major disadvantages of the construct are the need for a second surgical procedure, the potential for infection, and expense [10]. In 1996, Integra^{TM} was approved for the treatment of deep dermal and full thickness thermal burns by the FDA [152]. Boyce and Hansbrough have reported a modification to this construct, where the collagen and GAG matrix is seeded with fibroblasts and the silicon layer is replaced by cultured keratinocytes. Experimental results show that the keratinocytes fully differentiate and a complete DEJ forms in vivo [210, 211].

Apligraf^{TM}

Apligraf^{TM}, formerly known as Graftskin^{TM}, is manufactured by Organogenesis Incorporated. To prepare the construct, a gel of bovine collagen type I is seeded with allogenic fibroblasts [10]. Following 6 days in culture, allogenic keratinocytes are then applied and grown at the air-media interface for approximately 14 days. This process results in epidermal differentiation, including the formation of an immature stratum corneum that improves the barrier properties of the device [104]. The

construct, which is FDA approved for the treatment of chronic wounds, has been shown to have good clinical results, including no signs of immunological rejection and the in vivo formation of an intact DEJ [212]. In experimental studies, when used to temporarily cover meshed autograft, Apligraf™ improves wound healing, pigmentation, and reduces scar tissue formation [213]. However, this expensive product is time-consuming to manufacture and has a limited shelf life [104].

Biobrane™

Biobrane™, marketed by Smith and Nephew Inc., is composed of a nylon mesh that is coated in porcine collagen type I and attached to a thin silicone membrane (Fig. 14.6). This non-degradable device is recommended to expedite healing in superficial partial thickness burns and at autologous donor sites [10, 107, 214]. The outer silastic layer regulates the flux of water and prevents against bacterial contamination. As healing progresses, the device slowly separates from the skin surface [10]. It is imperative that the wound site be freshly excised and free of infection prior to the placement of the construct [215]. When used appropriately, the healing period can be reduced by up to 46% [10].

Transcyte™

Transcyte™, formerly known as Dermagraft Transitional Covering™, is marketed by Smith and Nephew Inc. The device has the same basic construction as Biobrane™, but is seeded with neonatal fibroblasts. During a 17-day culturing period, the fibroblasts synthesize and secrete various growth factors and ECM components that accumulate in the mesh. The construct is then frozen and stored, killing
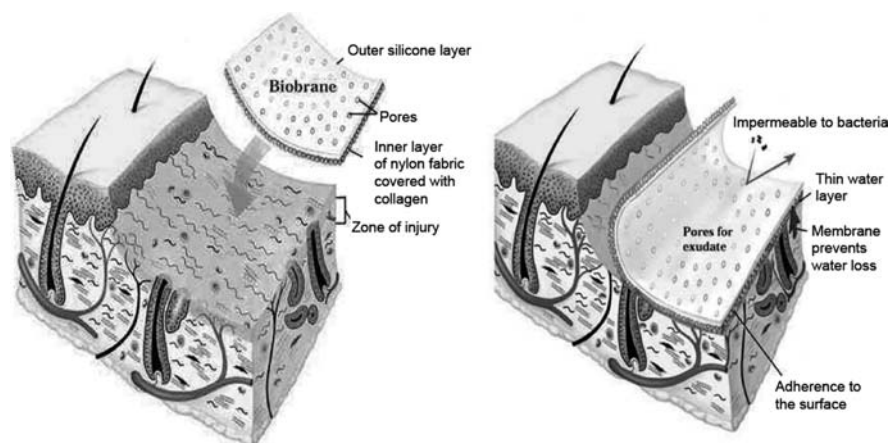
Fig. 14.6 The application of a Biobrane™ graft to a debrided wound bed following a partial thickness burn injury to promote wound healing. (Figure adapted with permission from www.burnsurgery.org)

the cells, but preserving the secreted factors [10]. While this material is extremely expensive, preliminary studies indicate that it has good adhesion, promotes healing and can remain intact for up to 6 weeks [216, 217].

14.6.3 Growth Factor Incorporation

The incorporation of cytokines or growth factors into the skin substitutes could promote regeneration and expedite healing [218–220]. As many interacting mediators are involved in the wound healing process (Table 14.4), research must be conducted in order to determine the optimal combination and concentrations of these factors. Additionally, it is necessary to develop a method for the sustained delivery of the factors to the wound site for the desired period. Localized application may reduce the likelihood of negative side-effects associated with systemic exposure. Binding the mediators to the constructs is a potential design strategy [221]. However, researchers must show that bound stimulatory molecules remain bioactive, but do not over-stimulate the cells, triggering cancerous growth.

14.6.4 Epidermal Stem Cells

As previously discussed, epidermal stem cells are found in the basal epidermal layer and the bulge region of the hair follicle. These undifferentiated cells proliferate rapidly, are capable of differentiation, and have a high capacity for self-renewal [41–43]. The development of practical methods to isolate and culture these cells holds great promise for tissue engineering applications. Using the stem cells, it may be possible to grow autologous epidermal tissue for transplantation in a clinically-acceptable time frame [45]. Potential epidermal stem cell markers include β_1 integrin and keratin 19 [43].

14.7 Future Outlook

While significant progress has been made in the development of burn dressing biomaterials and tissue-engineered constructs, the complete regeneration of fully-functional skin following burn injuries is currently unfeasible. Despite the disadvantages associated with autografting, the use of harvested host tissues remains the best treatment method for permanent wound closure in deep dermal and full thickness burns [1, 8, 104]. Continued research must be conducted to design new constructs or to improve the existing biomaterials in order to promote regeneration rather than repair in severe burns. A more thorough understanding of the events in integumentary organogenesis and wound healing could aid in this process. Further, constructs which more closely mimic the natural complexity of the skin should also be studied. The inclusion of appendage structures in the skin substitutes could greatly improve

Table 14.4 Some of the observed effects of various cytokines and growth factors during wound healing in the integumentary system

	Observed effect
EGF [222–225]	Expedites wound healing following burn injury
	Promotes re-epithelialization
	Stimulates keratinocyte proliferation and migration
	Stimulates dermal fibroblast proliferation
	Promotes collagen synthesis
	Stimulates dermal regeneration
bFGF [226–229]	Expedites wound healing
	Strongly promotes angiogenesis
	Stimulates keratinocyte proliferation
	Stimulates dermal fibroblast proliferation
	Increases the formation of granulation tissue
	Increases the tensile strength of the healed wound
Acidic FGF (aFGF) [221]	Less potent than bFGF
	Stimulates fibroblast proliferation
	May inhibit wound contraction
Platelet-derived growth factor (PDGF) [230]	Strongly promotes angiogenesis
	Chemotactic factor for dermal fibroblasts
	Promotes collagen synthesis and reorganization
	Increases the formation of granulation tissue
TGF-β [220, 229, 231, 232]	Promotes hypertrophic and keloid scar tissue formation
	Stimulates wound contraction
	Promotes re-epithelialization
	Stimulates keratinocyte proliferation, migration and differentiation
	Promotes collagen and GAG synthesis
	Greatly increases the formation of granulation tissue
	Increases the tensile strength of the healed wound
	Functions in T-cell suppression
Keratinocyte growth factor (KGF) [220, 230, 233]	Expedites wound healing
	Promotes re-epithelialization
	Stimulates keratinocyte proliferation, migration, and differentiation
	Protects keratinocytes from oxygen radicals
	Stimulates the expression of VEGF
	Stimulates the expression of activins, (activins promote keratinocyte proliferation and matrix deposition)
Interleukin-1 (IL-1) [221]	Expedites wound healing
	Stimulates keratinocyte proliferation
	Stimulates dermal fibroblast proliferation
	Promotes collagen synthesis
	Promotes immune response
IL-6 [221]	May promote keloid scar tissue formation
	Promotes collagen and GAG synthesis
	Promotes immune response
Vascular endothelial growth factor (VEGF) [221, 229]	Strongly promotes angiogenesis

Table 14.4 (continued)

	Observed effect
Growth hormone (GH) [221, 234]	Expedites wound healing
	Promotes re-epithelialization
	Promotes collagen synthesis
	Increases the formation of granulation tissue
	Increases the tensile strength of the healed wound
	Stimulates the expression of IGF-1
Insulin-like growth factor-1 (IGF-1) [221, 234, 235]	May function in scar tissue formation
	Expedites wound healing
	Promotes re-epithelialization
	Promotes collagen synthesis
	Can have negative side-effects if administered systemically

the device functionality. Incorporating the full epidermal and dermal cellular complement could also be of significant benefit. Seeding tissue-engineered constructs with melanocytes has been shown to improve skin pigmentation and cosmesis [236]. The addition of autologous Langerhans cells could potentially alleviate some of the problems associated with infection leading to tissue necrosis and device failure. As the number of patients who survive with severe burns over large TBSA increases, the need for an economical, effective, off-the-shelf material for permanent wound closure continues to motivate researchers in this field.

References

1. Balasubramani M, Kumar TR, and Babu M. Skin substitutes: a review. Burns, 2001, 27: 534–544.
2. Haake A, Scott GA, and Holbrook KA. Structure and function of the skin: overview of the epidermis and dermis, in Freinkel, RK and Woodley, DT, eds. The Biology of the Skin, The Parthenon Publishing Group Inc.: New York, 2001, pp. 19–45.
3. Pruitt BA and Mason AD. Epidemiological, demographic and outcome characteristics of burn injury, in Herndon, DN, ed. Total Burn Care, W.B. Saunders Company Ltd.: London, 1996, pp. 5–15.
4. Sheridan RL. Burns. Criti Care Med, 2002, 30S: S500–S514.
5. Sjoberg F, et al. Utility of an intervention scoring system in documenting effects of changes in burn treatment. Burns, 2000, 26: 553–559.
6. Nguyen TT, et al. Current treatment of severely burned patients. Ann Surg, 1996, 223: 14–25.
7. Quinn KJ, et al. Principles of burn dressings. Biomaterials, 1985, 6: 369–377.
8. Muller MJ, et al. Modern treatment of a burn wound, in Herndon, DN, ed. Total Burn Care, W.B. Saunders Company Ltd.: London, 1996, pp. 136–147.
9. Berthod F and Rouabhia M. Exhaustive review of clinical alternatives for damaged skin replacement, in Rouabhia, M, ed. Skin Substitute Production by Tissue Engineering, Chapman and Hall: New York, 1997, pp. 23–45.
10. Jones I, Currie L, and Martin R. A guide to biological skin substitutes. Br J Plast Surg, 2002, 55: 185–193.

11. Boyce ST and Warden GD. Principles and practices for treatment of cutaneous wounds with cultured skin substitutes. Am J Surg, 2002, 183: 445–456.
12. Brown AS and Barot LR. Biologic dressings and skin substitutes. Clin Plast Surg, 1986, 13: 69–74.
13. Tortora GT and Anagnostakos NP. Principles of Anatomy and Physiology: 2nd edn. Harper & Row Publishers: New York, 1975, pp. 104–114.
14. Burgeson RE and Christiano AM. The dermal-epidermal junction. Curr Opin Cell Biol, 1997, 9: 651–658.
15. Rouabhia M. Structural and functional complexity of the skin, in Rouabhia, M, ed. Skin Substitute Production by Tissue Engineering, R.G. Landes Company: Austin, 1997, pp. 3–22.
16. Gupta S and Sauder DN. The skin as an immune organ: the role of keratinocyte cytokines. Univ Toronto Med J, 1995, 72: 118–123.
17. Smack DP, Korge BP, and James, WD. Keratin and keratinization. JAm Acad Dermatol, 1994, 30: 85–102.
18. Latkowski JA and Freedberg IM. Epidermal cell kinetics, epidermal differentiation and keratinization, in Freedberg, IM, et al., eds. Fitzpatrick's Dermatology in General Medicine, McGraw-Hill: New York, 1999, pp. 133–143.
19. Steinert PM. The two-chain coiled-coil molecule of native epidermal keratin intermediate filaments is a type I-type II heterodimer. J Biol Chem, 1990, 265: 8766–8774.
20. Moll R, et al. The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. Cell, 1982, 31: 11–24.
21. Sun TT. Classification, expression and possible mechanisms of evolution of mammalian epithelial keratins: a unifying model. Cancer Cell, 1984, 1: 169–176.
22. Valyi-Nagy IT, et al., Undifferentiated keratinocytes control growth, morphology and antigen expression of normal melanocytes through cell-cell contact. Lab Invest, 1993, 69: 152–159.
23. Gordon PR, Mansur CP, and Gilchrest BA. Regulation of human melanocyte growth, dendricity, and melanization by keratinocyte derived factors. J Invest Dermatol, 1989, 92: 565–572.
24. Martini FH, Timmons MJ, and Mackinley MP. Human Anatomy. Prentice-Hall Inc.: New Jersey, 2000, pp. 88–107.
25. Nordlund JJ and Boissy RE. The biology of melanocytes, in Freinkel, RK and Woodley, DT, eds. The Biology of Skin, The Parthenon Publishing Group: New York, 2001, pp. 113–131.
26. Yamamoto O and Bhawan J., Three modes of melanosome transfers in caucasian facial skin: hypothesis based on an ultrastructural study. Pigment Cell Res, 1994, 7: 158–169.
27. Boissy RE and Nordlund JJ., Biology of melanocytes, in Arndt, KA, et al., eds. Cutaneous Medicine and Surgery, W.B. Saunders Company Ltd.: Philadelphia, 1996, pp. 1203–1218.
28. Post PW and Rao DC., Genetic and environmental determinants of skin colour. Am J Phys Anthropol, 1977, 47: 399–402.
29. Tachibana T. The Merkel cell: recent findings and unresolved problems. Arch Histolo Cytol, 1995, 58: 379–396.
30. Fantini F and Johansson O. Neurochemical markers in human cutaneous Merkel cells. Exp Dermatol, 1995, 4: 365–371.
31. Narisawa Y., et al. Biological significance of dermal Merkel cells in the development of cutaneous nerves in human fetal skin. J Histochem Cytochem, 1992, 40: 65–71.
32. Sauder DN and Pastore S., Cytokines in contact dermititis. Am J Contact Dermat, 1993, 4: 215–224.
33. Sauder DN The skin as an immunologic organ, in Soter, NA, ed. Pathophysiology of Dermatologic Diseases, McGraw-Hill Inc.: New York, 1991, pp. 101–110.
34. Bos JD and Kapsenberg ML., The skin immune system. Immunol Today, 1986, 7: 235–240.
35. Barratt-Boys SM, Watkins SC, and Finn OJ., In vivo migration of dendritic cells differentiated in vitro. J Immunol, 1997, 158: 4543–4547.

36. Romani N, Lenz A, and Glassel H., Cultured human Langerhans cells resemble lymphoid dendritic cells in phenotype and function. J Invest Dermatol, 1989, 93: 600–609.

37. Groux H, Fournier N, and Cottrez F., Role of dendritic cells in the generation of regulatory T cells. Semin Immunol, 2004, 16: 99–106.

38. Jones ND, et al. Differential susceptibility of heart, skin and islet allografts to T-cell mediated rejection. J Immunol, 2001, 166: 2824–2830.

39. Erdag G and Sheridan RL. Fibroblasts improve performance of cultured composite skin substitutes on athymic mice. Burns, 2004, 30: 322–328.

40. Hornebeck W. Down-regulation of tissue inhibitor matrix metalloprotease-1 (TIMP-1) in aged human skin contributes to matrix degradation and impaired cell growth and survival. Pathol Biol, 2003, 51: 569–573.

41. Lavker RM and Sun T. Heterogeneity in epidermal basal keratinocytes: morphological and functional correlations. Science, 1982, 15: 1239–1241.

42. Lavker RM and Sun TT. Epidermal stem cells. J Invest Dermatol, 1983, 81: 121S–127S.

43. Janes SM, Lowell S, and Hutter C. Epidermal stem cells. J Pathol, 2002, 197: 479–491.

44. Barrandon Y and Green H. Three clonal types of keratinocyte with different capacities for multiplication. Proc Natl Acad Sci USA, 1987, 84: 2302–2306.

45. Germain L, et al. Skin stem cell identification and culture: a potential tool for rapid epidermal sheet production and grafting, in Rouabhia, M, ed. Skin Substitute Production by Tissue Engineering, R.G. Landes Company: Austin, 2001, pp. 177–210.

46. Brody I. The ultrastructure of the tonofibrils in the keratinization process of normal human epidermis. J Ultrastruct Res, 1960, 4: 264–297.

47. Fuchs E and Cleveland D. A structural scaffolding of intermediate filaments in health and disease. Science, 1998, 279: 514–519.

48. Fukuyama K, Kakimi S, and Epstein WL. Detection of a fibrous component in keratohyalin granules of newborn rat epidermis. J Invest Dermatol, 1980, 74: 174–180.

49. Yaffe MB and Eckert RL. Involucrin, type-I and type-II keratins and filaggrin are covalently cross-linked components of the keratinocyte cornified envelope. Clin Res, 1992, 40: A522–A522.

50. Eckert RL, et al. Involucrin – structure and role in envelope assembly. J Invest Dermatol, 1993, 100: 613–617.

51. Hohl D. Cornified cellular envelope. Dermatologica, 1990, 180: 201–211.

52. McCall CA and Cohen JJ. Programmed cell death in terminally differentiating keratinocytes: role of endogenous endonuclease. J Invest Dermatol, 1991, 97: 111–114.

53. Wrone-Smith T, et al. Keratinocytes from psoriatic plaques are resistant to apoptosis compared with normal skin. Am J Pathol, 1997, 151: 1321–1329.

54. Elias PM. Epidermal lipids, barrier function, and desquamation. J Invest Dermatol, 1983, 80: 44S–49S.

55. Weinstein GD and Boucek RJ. Collagen and elastin of human dermis. JInvest Dermatol, 1960, 35: 227–229.

56. Smith LT, Holbrook KA, and Byers HP. Structure of the dermal matrix during development and in the adult. J Invest Dermatol, 1982, 79: 93S–104S.

57. Bruckner-Tuderman L, Hopfner B, and Hammami-Hauasli N. Biology of anchoring fibrils: lessons from dystrophic epidermolysis bullosa. Matrix Biol, 1999, 18: 43–54.

58. Pulkkinen L and Uitto J. Mutation analysis and molecular genetics of epidermolysis bullosa. Matrix Biol, 1999, 18: 29–42.

59. Braverman IM. The cutaneous microcirculation. J Invest Dermatol Symp, 2000, 5: 3–9.

60. Metze D and Luger T. Nervous system in the skin, in Freinkel, RK and Woodley, DT, eds. The Biology of the Skin, The Parthenon Publishing Group: New York, 2001, pp. 153–176.

61. Skobe M and Detmar M. Structure, function and molecular control of the skin and lymphatic system. J Invest Dermatol Symp, 2000, 5: 14–19.

62. Tzaphildou M. The role of collagen and elastin in aged skin: an image processing approach. Micron, 2004, 35: 873–877.
63. Kielty CM and Shuttleworth CA. Microfibrillar elements of the dermal matrix. Microsc Res Tech, 1997, 38: 413–427.
64. Garrone R, Distribution of minor collagens during skin development. Microsc Res Tech, 1997, 38: 407–412.
65. Keene DR, Marinkovich MP, and Sakai LY. Immunodissection of the connective tissue matrix in human skin. Microsc Res Tech, 1997, 38: 394–406.
66. Pasquali-Ronchetti I and Baccareni-Contri M. Elastic fiber during development and aging. Microsc Res Techn, 1997, 38: 428–435.
67. Fosang AJ and Hardingham TE. Matrix proteoglycans, in Comper, WD, ed. Extracellular Matrix Volume 2, Molecular Components and Interactions, Harwood Academic Publishers: Amsterdam, 1996, pp. 200–218.
68. Iozzo RV. Matrix proteoglycans: from molecular design to cellular function. Ann Rev Biochem, 1998, 67: 609–652.
69. Saliba MJ. Heparin in the treatment of burns: a review. Burns, 2001, 27: 349–358.
70. Savani RC, et al. The role of hyaluronan-receptor interactions in wound repair, in Garg, HG and Longaker, MT, eds. Scarless Wound Healing, Marcel Dekker, Inc.: New York, 2000, pp. 115–142.
71. Trowbridge JM and Gallo RL. Dermatan sulfate: new functions from an old glycosaminoglycan. Glycobiology, 2002, 126: 117R–125R.
72. Kljaer M. Role of extracellular matrix in adaptation of tendon and skeletal muscle to mechanical loading. Physiol Rev, 2004, 84: 649–698.
73. Angulo J, et al. The activation of fibroblast growth factors (FGFs) by glycosaminoglycans: influence of the sulfation pattern on the biological activity of FGF-1. Chem Biochem, 2004, 5: 55–61.
74. Martinez-Hernadez A and Amenta PS. The basement membrane in pathology. Lab Invest, 1983, 48: 656–677.
75. Woodley DT and Chen M. The basement membrane zone, in Freinkel, RK and Woodley, DT, eds. The Biology of the Skin, The Parthenon Publishing Group: New York, 2001, pp. 133–135.
76. Brown TA, Gil SG, and Sybert VP. Defective integrin alpha 6 beta 4 in the skin of patients with junctional epidermolysis bullosa and pyloric atresia. J Invest Dermatol, 1996, 107: 384–391.
77. Larjava H, Expression of β1 integrins in normal human keratinocytes. Am J Med Sci, 1991, 301: 63–68.
78. Graf J, et al. Identification of an amino acid sequence in laminin mediating cell attachment, chemotaxis, and receptor binding. Cell, 1987, 48: 989–996.
79. Bateman JF, Lamande, SR, and Ramshaw, JAM. Collagen superfamily, in Extracellular Matrix Volume 2, Components and Interactions, Comper, WD, ed. Harwood Academic Publishers: Amsterdam, 1996, pp. 22–67.
80. Katz AJ. Emerging approaches to the tissue engineering of fat. Tissue Eng, 1999, 26: 587–603.
81. Widelitz RB, et al. Molecular histology in skin appendage morphogenesis. Microsc Res Tech, 1997, 38: 452–465.
82. Shakespeare P. Burn wound healing and skin substitutes. Burns, 2001, 27: 517–522.
83. Sato K, et al. Biology of the sweat glands and their disorders. I. Normal sweat gland function. J Am Acad Dermatol, 1989, 20: 537–563.
84. Hurley HJ. The eccrine sweat glands: structure and function, in Freinkel, RK and Woodley, DT, eds. The Biology of the Skin, The Parthenon Publishing Group: New York, 2001, pp. 47–76.
85. Speilman AL, et al. Proteinaceous precursors of human axillary odor: isolation of two novel odor-binding proteins. Experientia, 1995, 15: 40–47.

86. Botek AA and Lookingbill DP. The structure and function of sebaceous glands, in Freinkel, RK and Woodley, DT, eds. The Biology of the Skin, The Parthenon Publishing Group: New York, 2001, 87–100.

87. Thody AJ and Shuster S. Control and function of sebaceous glands. Physiolo Rev, 1989, 69: 383–416.

88. Freinkel RK, Hair, in Freinkel, RK and Woodley, DT, eds. The Biology of the Skin, The Parthenon Publishing Group: New York, 2001, pp. 77–86.

89. Spradling A, Drummond-Barbosa D, and Kai, T. Stem cells find their niche. Nature, 2001, 414: 98–104.

90. Akiyama M, et al. Characterization of hair follicle bulge in human fetal skin; the human fetal bulge is a pool of undifferentiated keratinocytes. J Invest Dermatol, 1995, 105: 844–850.

91. Levit EK and Scher RK. Basic science of the nail unit, in Freinkel, RK and Woodley, DT, eds. The Biology of the Skin, The Parthenon Publishing Group: New York, 2001, pp. 101–112.

92. Kitahara T and Ogawa H., Cellular features of differentiation in the nail. Microsc Res Tech, 1997, 38: 436–442.

93. Browder LW, Erickson CA, and Jeffrey, WR. Developmental Biology. Philadelphia, 1991, pp. 260–282.

94. Larsen WJ. Development of the integumentary system, in Human Embryology, 3rd edn. Churchill Livingstone: Philadelphia, 2001, pp. 465–480.

95. Carlson BM. Integumentary, skeletal and muscular systems, in Human Embryology and Developmental Biology, 2nd edn. Mosby Inc.: St. Louis, 1999, pp. 148–183.

96. Dye FJ. Human Life Before Birth. Harwood Academic Publishers: Amsterdam, 2000, pp. 97–104.

97. Toole BP. Hyaluronan in morphogenesis. J Intern Med, 1997, 242: 35–40.

98. Cook D. Classification of burns, in Bosworth, C, ed. Burns Trauma, Bailliere Tindall: London, 1997, pp. 19–34.

99. Monafo W, Wound care, in Herndon, DN, ed. Total Burn Care, W.B. Saunders Company Ltd.: London, 1996, pp. 88–97.

100. Williams WG and Phillips LG. Pathophysiology of the burn wound, in Herndon, DN, ed. Total Burn Care, W.B. Saunders Company Ltd.: London, 1996, pp. 63–70.

101. Clark RAF. Wound repair: lessons for tissue engineering, in Lanza, RP, Langer, R, and Chick, WL, eds. Principles of Tissue Engineering, R.G. Landes Company: Austin, 1997, pp. 737–768.

102. Krisner RS and Eaglstein W. The wound healing process. Wound Healing, 1993, 11: 629–640.

103. Martin P. Wound healing – aiming for perfect skin regeneration. Science, 1997, 276: 75–80.

104. Hardin-Young J and, Parenteau N. Bilayered skin constructs, in Atala, A and Lanza, RP, eds. Methods of Tissue Engineering, Academic Press: San Diego. 2002, pp. 1177–1188.

105. Servold SA. Growth factor impact on wound healing. Clin Podiatr Med Surg, 1991, 8: 937–953.

106. Pankurst S, Wound care, in Burns Trauma, Bosworth, C, ed. Bailliere Tindall: London, 1997, pp. 63–74.

107. Kane JB, Tompkins RG, and Yarmush ML. Burn dressings, in Ratner, BD, et al., eds. Biomaterials Science, Academic Press: San Diego. 1996, pp. 360–370.

108. Tuan TL and Nichter LS. The molecular basis of keloid and hypertrophic scar formation. Mol Med Today, 1998, 4: 19–24.

109. Rockwell WB, Cohen IK, and Ehrlich HP. Keloid and hypertrophic scars: a comprehensive review. Plast Reconstr Surg, 1989, 84: 827–837.

110. Robb EC, et al. A new model for studying the development of human hypertrophic burn scar formation. J Burn Care Rehabil, 1987, 8: 371–375.

111. Hindler F and Traber DL. Pathophysiology of the systemic inflammatory response syndrome, in Herndon, DN, ed. Total Burn Care, W.B. Saunders Company Ltd.: London, 1996, pp. 207–216.
112. Kramer GC and Nguyen TT. Pathophysiology of burn shock and burn edema, Herndon, DN, ed. in Total Burn Care, W.B. Saunders Company Ltd.: London, 1996, pp. 44–52.
113. Erol S, et al. Changes of microbial flora and wound colonization in burned patients. Burns, 2004, 30: 357–361.
114. Kanchanapoom T and Khardori N. Management of infections in patients with severe burns: impact of multiresistant pathogens. J Burns Surg Wound Care, 2002, 1: 1–7.
115. Allgower M, Schoenenberger GA, and Sparkes BG. Burning the largest immune organ. Burns, 1995, 21: S7–S47.
116. Sparkes BG. Immunological responses to thermal injury. Burns, 1997, 23: 106–113.
117. Deveci M, et al. Comparison of lymphocyte populations in cutaneous and electrical burn patients: a clinical study. Burns, 2000, 26: 229–232.
118. Schwacha MG. Macrophages and post-burn immune dysfunction. Burns, 2003, 29: 1–14.
119. Vindenes HA and Bjerknes R. Impaired actin polymerization and depolymerization in neutrophils from patients with thermal injury. Burns, 1997, 23: 131–136.
120. Tyler MPH., et al. Dermal cellular inflammation in burns. an insight into the function of dermal microvascular anatomy. Burns, 2001, 27: 433–438.
121. Munster AM. The immunological response and strategies for intervention, in Herndon, DN, ed. Total Burn Care, W.G. Saunders Company Ltd.: London. 1996, pp. 279–292.
122. Lu S, et al. Effect of necrotic tissue on progressive injury in deep partial thickness burn wounds. Chin Med J, 2002, 115: 323–325.
123. Sheridan RL and Thompkins RG. Etiology and prevention of multisystem organ failure, in Herndon, DN, ed. Total Burn Care, W.G. Saunders Company Ltd.: London, 1996, pp. 302–312.
124. Sheridan RL. What′s new in burns and metabolism. J Am College of Surgeons, 2004, 198: 243–263.
125. Hartford CE. Care of out-patient burns, in Herndon, DN, ed. Total Burn Care, W.G. Saunders Company Ltd.: London, 1996, pp. 71–80.
126. Jones OC. Management of pain in the burns patient, in Bosworth, C, ed. Burns Trauma, Bailliere Tindall: London, 1997, pp. 95–110.
127. Wachtel TL. Initial care of major burns. Postgrad Med, 1989, 85: 178–196.
128. Warden GD. Fluid resuscitation and early management, in Herndon, DN, ed. Total Burn Care, W.G. Saunders Company Ltd.: London, 1996, pp. 53–60.
129. Ho WS, et al. Skin care in burn patients: a team approach. Burns, 2001, 27: 489–491.
130. Zhi-yang F. Local care in severe burn, in Zhi-yang, F et al., ed. Modern Treatment of Severe Burns, Springer-Verlag: Berlin, 1992, pp. 51–63.
131. Ryan TJ. Wound healing and current dermatologic dressings. Clin Dermatol, 1990, 8: 21–29.
132. Boyce ST. Design principles for composition and performance of cultured skin substitutes. Burns, 2001, 27: 523–533.
133. Sheng-de G. Management of full-thickness burns, in Zhi-yang, F et al., eds. Modern Treatment of Severe Burns, Springer-Verlag: Berlin, 1992, pp. 64–79.
134. Tanner JC. The mesh skin graft. Plast Reconst Surg, 1964, 34: 287–292.
135. Thompkins RG and Burke JF. Alternative wound coverings, in Herndon, DN, ed. Total Burn Care, W.B. Saunders Company Ltd.: London, 1996, pp. 164–172.
136. Pruitt BA. The evolutionary development of biologic dressings and skin substitutes. J Burn Care Rehabil, 1997, 18: S2–S5.
137. Hansbrough J. Allograft (Homograft) Skin, in Wound Coverage with Biologic Dressings and Cultured Skin Substitutes, R.G. Landes Company: Austin, 1992, pp. 21–39.
138. Mackie DP. The Euro skin bank: development and application of glycerol-preserved allografts. J Burn Care Rehabil, 1997, 18: S7–S12.

139. Ben-Bassat H, et al. How long can cryopreserved skin be stored to maintain adequate graft performance? Burns, 2001, 27: 425–431.
140. Ramakrishnan KM and Jayaraman V. Management of partial-thickness burns by amniotic membranes: a cost-effective treatment in developing countries. Burns, 1997, 23: S33–S36.
141. Sawhney CP. Amniotic membrane as a biological dressing in the management of burns. Burns, 1989, 15: 339–342.
142. Hansbrough JF. Human amnion, in Wound Coverage with Biologic Dressings and Cultured Skin Substitutes, R.G. Landes Company: Austin, 1992, pp. 10–12.
143. Lin SD, et al. Amnion overlay meshed skin autograft. Burns, 1985, 11: 374–378.
144. Sullivan TP, et al. The pig as a model for human wound healing. Wound Repair Regen, 2001, 9: 66–76.
145. Hansbrough J. Xenograft (Heterograft) Skin, in Wound Coverage with Biologic Dressings and Cultured Skin Substitutes, R.G. Landes Company: Austin, 1992, pp. 13–20.
146. Bach, FH et al. Uncertainty in xenotransplantation: individual benefit versus collective risk. Nat Med, 1998, 4: 141.
147. Kim B, Baez CE, and Atala A. Biomaterials for tissue engineering. World J Urol, 2000, 18: 2–9.
148. Marler JJ, et al. Transplantation of cells in matrices for tissue regeneration. Adv Drug Deliv Rev, 1998, 33: 165–182.
149. Berthod F and Damour O. In vitro reconstructed skin models for wound coverage in deep burns. Br J Dermatol, 1997, 136: 809–816.
150. Gogolewski S and Pennings AJ. An artificial skin based on biodegradable mixtures of polylactides and polyurethanes for full-thickness skin wound covering. Makromol Chem, Rapid Commun, 1983, 4: 675–680.
151. Parenteau N, et al. Biological and physical factors influencing the successful engraftment of a cultured human skin substitute. Biotechnol Bioeng, 1996, 52: 3–14.
152. Sefton MV and Woodhouse KA. Tissue engineering. J Cutan MedSurg, 1998, 3: S1-18–S1-23.
153. Jorgensen PH, Bang C, and Andreassen TT. Mechanical properties of skin graft wounds. Br J Plast Surg, 1993, 46: 565–569.
154. Jonkman MF, et al. Evaporative water loss and epidermis regeneration in partial-thickness wounds dressed with a fluid-retaining versus a clot-inducing wound covering in guinea pigs. Scand J Plast Reconstr Surg, 1989, 23: 29–34.
155. Greenwald SE and Berry, CL. Improving vascular grafts: the importance of mechanical and hemodynamic properties. J Pathol, 2000, 190: 292–299.
156. Peppas NA, et al. Physiochemical foundations and structural design of hydrogels in medicine and biology. Ann Rev Biomed Eng, 2000, 2: 9–29.
157. Chvapil M. Considerations on manufacturing principles of a synthetic burn dressing: a review. J Biomed Mater Res, 1982, 16: 245–263.
158. Chen G, Ushida T, and Tateishi T. Hybrid biomaterials for tissue engineering: a preparative method for PLA or PLGA-collagen hybrid sponges. Adv Mater, 2000, 12: 455–457.
159. Beumer GJ, Blitterswijk CA and Ponec M. Biocompatibility of a biodegradable matrix used as a skin substitute: an in vivo evaluation. J Biomed Mater Res, 1994, 28: 545–552.
160. Wald HL, et al. Cell seeding in porous transplantation devices. Biomaterials, 1993, 14: 270–278.
161. Ruszczak Z. Effect of collagen matrices on dermal wound healing. Adv Drug Deliv Rev, 2003, 55: 1595–1611.
162. Badylak SF. The extracellular matrix as a scaffold for tissue reconstruction. Cell Dev Biol, 2002, 13: 377–383.
163. Alper JC, et al. Moist wound healing under a vapor permeable membrane. J Am Acad Dermatol, 1983, 8: 347–353.
164. Alvarez OM, Mertz PM, and Eaglstein WH. The effect of occlusive dressings on collagen synthesis and re-epithelialization in superficial wounds. J Surg Res, 1983, 35: 142–148.

165. Eldad A, Kadar ST, and Kushnir M. Immediate dressing of the burn wound – will it change its natural history? Burns, 1991, 17: 233–238.
166. Helfman T, Ovington L, and Falanga V. Occlusive dressings and wound healing. Clin Dermatol, 1994, 12: 121–127.
167. Falanga V. Occlusive wound dressings. Arch Dermatol, 1988, 124: 872–876.
168. Queen D, et al. The preclinical evaluation of the water vapour transmission rate through burn wound dressings. Biomaterials, 1987, 8: 367–371.
169. Berardesca E, et al. Effect of occlusive dressings on the stratum corneum water holding capacity. Am J Med Sci, 1992, 304: 25–28.
170. Neal DE, et al. The effects of an adherent polyurethane film and conventional absorbent dressing in patients with small partial thickness burns. Br J Clin Pract, 1981, 35: 254–257.
171. Poulsen TD, et al. Polyurethane film (Opsite∗) vs. impregnated gauze (Jelonet∗) in the treatment of outpatient burns: a prospective, randomized study. Burns, 1991, 17: 59–61.
172. Eldad A and Tuchman I. The use of Omiderm® as an interface for skin grafting. Burns, 1991, 17: 155–158.
173. Staso MA, et al. Experience with Omiderm∗ – a new burn dressing. J Burn Care Rehabil, 1991, 12: 209–210.
174. Jonkman MF, et al. A clot-inducing wound covering with high vapour permeability: enhancing effects on epidermal wound healing in partial-thickness wounds in guinea pigs. Surgery, 1988, 104: 537–545.
175. Jonkman MF, Bruin P, and Pennings AJ. Poly(ether urethane) wound coverings with high vapour permeability compared with conventional tulle gras on split-skin donor sites. Burns, 1989, 15: 211–216.
176. Bruin P, et al. A new porous polyetherurethane wound covering. J Biomed Mater Res, 1990, 24: 217–226.
177. Landrum J and Jones, EB. A new dressing for burns: enclosure in a plasticized polyvinyl chloride sheet. Burns, 1976, 2: 86–89.
178. Davies JWL. Synthetic materials for covering burn wounds: progress towards perfection. Part I. Short term dressing materials. Burns, 1983, 10: 94–103.
179. Dressler DP, Barbee WK, and Sprenger R. The effect of Hydron burn wound dressing on burned rat and rabbit ear wound healing. J Trauma, 1980, 20: 1024–1028.
180. Brown AS. Hydron for burns. Plast Reconstr Surg, 1981, 67: 810–811.
181. Young SR, et al. Comparison of the effects of semi-occlusive polyurethane dressings and hydrocolloid dressings on dermal repair: 1. cellular changes. J Invest Dermatol, 1991, 97: 586–592.
182. Agren MS and Wijesinghe C. Occlusivity and effects of two occlusive dressings on normal human skin. Acta Derm Venereol, 1994, 74: 12–14.
183. Hansbrough J. Synthetic and biosynthetic wound coverings, in Wound Coverage with Biologic Dressings and Cultured Skin Substitutes, R.G. Landes Company: Austin, 1992, pp. 41–59.
184. Currie LJ, Sharpe JR, and Martin R. The use of fibrin glue in skin grafts and tissue-engineered skin replacements: a review. Plast Reconstr Surg, 2001, 108: 1713–1726.
185. DeBlois C, Cote MF, and Doillon CJ. Heparin-fibroblast growth factor-fibrin complex: in vitro and in vivo applications to collagen-based materials. Biomaterials, 1993, 15: 665–672.
186. Cote MF and Doillon CJ. In vitro contraction rate of collagen in sponge-shape matrices. J Biomater Sci Polym Ed, 1992, 3: 301–313.
187. Ramshaw JAM, Werkmeister JA, and Glattauer V. Collagen-based biomaterials. Biotechnol Genetic Eng Rev, 1995, 13: 335–382.
188. Doillon CJ, et al.Porosity and biological properties of polyethylene glycol-conjugated collagen materials. J Biomater Sci Polym Ed, 1994, 6: 715–728.
189. Rheinwald JG and Green H. Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. Cell, 1975, 6: 331–343.

190. Carsin H, et al. Cultured epithelial autografts in extensive burn coverage of severely traumatized patients: a five year single-center experience with 30 patients. Burns, 2000, 26: 379–387.

191. Hafemann B, et al. Treatment of skin defects using suspensions of in vitro cultured keratinocytes. Burns, 1994, 160: 168–172.

192. Blight A, et al. The treatment of donor sites with cultured epithelial grafts. Br J Plast Surg, 1991, 44: 12–14.

193. Gielen V, et al. Progressive replacement of human cultured epithelial autograts by recipient cells as evidenced by HLA class I antigens expression. Dermatologica, 1987, 175: 166–170.

194. Lam PK, et al. Development and evaluation of a new composite Laserskin graft. J Trauma, 1999, 47: 918–922.

195. Lees VC, Fan, TD, and West, DC. Angiogenesis in a delayed revascularization model is accelerated by angiogenic oligosaccharides of hyaluronan. Lab Invest, 1995, 73: 259–266.

196. Lobmann R, et al. Autologous human keratinocytes cultured on membranes composed of benzyl ester of hyaluronic acid for grafting in nonhealing diabetic foot lesions: a pilot study. J Diabetes Complicat, 2003, 17: 199–204.

197. Naughton G, Mansbridge, J, and Gentzkow, G. A metabolically active human dermal replacement for the treatment of diabetic foot ulcers. Artif Organs, 1997, 21: 1203–1210.

198. Naughton GK. Skin and epithelia, in Lanza, RP, Langer, R, and Chick, WL, eds. Principles of Tissue Engineering, R.G. Landes Company: Austin, 1997, pp. 769–782.

199. Gentzkow G, et al. Use of Dermagraft, a cultured human dermis, to treat diabetic foot ulcers. Diabetes Care, 1996, 19: 350–354.

200. Hansbrough JF, Dore C, and Hansbrough WB. Clinical trials of a living dermal tissue replacement beneath meshed, split-thickness skin grafts on excised burn wounds. J Burn Care Rehabil, 1992, 13: 519–529.

201. Hansbrough JF, et al. Evaluation of a biodegradable matrix containing cultured human fibroblasts as a dermal replacement beneath meshed split-thickness skin grafts. Surgery, 1992, 111: 438–446.

202. Kearney JN. Clinical evaluation of skin substitutes. Burns, 2001, 27: 545–551.

203. Wainwright D. Use of an acellular allograft dermal matrix (AlloDerm) in the management of full-thickness burns. Burns, 1995, 21: 243–248.

204. Bannasch H, et al. Skin tissue engineering. Clin Plast Surg, 2003, 30: 573–579.

205. Wainwright D, et al. Clinical evaluation of an acellular allograft dermal matrix in full-thickness burns. J Burn Care Rehabil, 1996, 17: 124–136.

206. Yannas IV, Burke JF, and Orgill DP. Wound tissue can utilize a polymeric template to synthesize a functional extension of skin. Science, 1982, 215: 174–176.

207. Burke JF, et al. Successful use of a physiologically acceptable artificial skin in the treatment of extensive burn injury. Ann Surg, 1981, 194: 413–428.

208. Ellis DL and Yannas IV. Recent advances in tissue synthesis in vivo by use of collagen-glycosaminoglycan copolymers. Biomaterials, 1996, 17: 291–299.

209. Sheridan RL, et al. Artificial skin in massive burns – results to ten years. Eur J Plast Surg, 1994, 17: 91–93.

210. Cooper ML and Hansbrough JF. Use of a composite skin graft composed of cultured human keratinocytes and fibroblasts and a collagen-GAG matrix to cover full-thickness wounds on athymic mice. Surgery, 1991, 109: 198–207.

211. Boyce ST and Hansbrough JF. Biologic attachment, growth and differentiation of cultured human keratinocytes on a graftable collagen and chondroitin-6-sulfate substrate. Surgery, 1988, 103: 421–431.

212. Falanga V, et al. Rapid healing of venous ulcers and lack of clinical rejection with an allogenic cultured human skin equivalent. Arch Dermatol, 1998, 134: 293–300.

213. Guerret S, et al. Long-term remodeling of a bilayered living human skin equivalent (Apligraft®) grafted onto nude mice: immunolocalization of human cells and characterization of extracellular matrix. Wound Repair Regen, 2003. 11: 35–45.

214. Klein RL, Rothman BF, and Marshall R. Biobrane – a useful adjunct in the therapy of outpatient burns. J Pediatr Surg, 1984, 19: 846–847.
215. Purdue GF, et al. Biosynthetic skin substitute versus frozen human cadaver allograft for temporary coverage of excised burn wounds. J Trauma, 1987, 27: 155–157.
216. Purdue GF. Dermagraft-TC pivotal efficacy and safety study. J Burn Care Rehabil, 1997, 18: S13–S14.
217. Hansbrough J. Dermagraft-TC for partial-thickness burns: A clinical evaluation. J Burn Care Rehabil, 1997, 18: S25–S28.
218. Wang HJ, et al. Acceleration of skin graft healing by growth factors. Burns, 1996, 22: 10–14.
219. Ono I, et al. Studies on cytokines related to wound healing in donor site wound fluid. J Dermatol Sci, 1995, 10: 241–245.
220. Smith PD, et al. Efficacy of growth factors in the accelerated closure of interstices in explanted meshed human skin grafts. J Burn Care Rehabil, 2000, 21: 5–9.
221. Hansbrough JF. Biologic mediators in wound healing, in Wound Coverage with Biologic Dressings and Cultured Skin Substitutes, R.G. Landes Company: Austin, 1992, pp. 116–147.
222. Wenczak BA, Lynch JB, and Nanney LB. Epidermal growth factor receptor distribution in burn wounds. Implications for growth factor-mediated repair. J Clin Invest, 1992, 90: 2392–2401.
223. Kiyohara Y, et al. Improvements in wound healing by epidermal growth factor (EGF) ointment. II. Effect of protease inhibitor, nafamostat, on stabilization and efficacy of EGF in burn. J Pharmacobiodynamics, 1991, 14: 47–52.
224. Chvapil M, Gaines JA, and Gilman T. Lanolin and epidermal growth factor in healing of partial-thickness pig wounds. J Burn Care Rehabil, 1988, 9: 279–284.
225. Arturson G. Epidermal growth factor in the healing of corneal wounds, epidermal wounds and partial-thickness scalds. A controlled animal study. Scand J Plast Reconstr Surg, 1984, 18: 33–37.
226. Gibran NS, et al. Basic fibroblast growth factor in the early human burn wound. J Surg Res, 1994, 56: 226–234.
227. Fu X, et al. Recombinant bovine basic fibroblast growth factor accelerates wound healing in patients with burns, donor sites and chronic dermal ulcers. Chin Med J, 2000, 113: 367–371.
228. Fu X, et al. Randomised placebo-controlled trial of use of topical recombinant bovine basic fibroblast growth factor for second-degree burns. Lancet, 1998, 352: 1661–1664.
229. Hakvoort T, et al. Transforming growth factor-beta(1), -beta(2), -beta(3), basic fibroblast growth factor and vascular endothelial growth factor expression in keratinocytes of burn scars. Eur Cytokine Netw, 2000, 11: 233–239.
230. Danilenko DM, et al. Growth factors in porcine full and partial thickness burn repair. Differing targets and effects of keratinocyte growth factor, platelet-derived growth factor-BB, epidermal growth factor, and neu differentiation factor. Am J Pathol, 1995, 147: 1261–1277.
231. Polo M, et al. Cytokine production in patients with hypertrophic burn scars. J Burn Care Rehabil, 1997, 18: 477–482.
232. Smith P, et al. TGF-β2 activates proliferative scar fibroblasts. J Surg Res, 1999, 82: 319–323.
233. Beer HD, et al. Expression and function of keratinocyte growth factor and activin in skin morphogenesis and cutaneous wound repair. J Invest Dermatol Symp, 2000, 5: 34–39.
234. Edmondson SR, et al. Epidermal homeostasis: the role of the growth hormone and insulin-like growth factor systems. Endocr Rev, 2003, 24: 737–764.
235. Wolf SE, et al. Insulin-like growth factor-1/insulin-like growth factor binding protein-3 alters lymphocyte responsiveness following severe burn. J Surg es, 2004, 117: 255–261.
236. Swope VB, Supp AP, and Boyce ST. Regulation of cutaneous pigmentation by titration of human melanocytes in cultured skin substitutes grafted to athymic mice. Wound Repair Regen, 2002, 10: 378–386.

Chapter 15
Natural and Synthetic Polymeric Scaffolds

Diana M. Yoon and John P. Fisher

15.1 Introduction

Clinical therapies have increasingly incorporated polymers as the basis for biomaterials. Polymers are natural or synthetic long chain molecules that are formed by linking repetitive monomer units. In the tissue engineering field, polymeric scaffolds have been widely investigated. A scaffold acts as a temporary support matrix for various cell populations in order to achieve tissue repair or regeneration. Clearly, polymer properties and scaffold design greatly impacts the function of incorporated cells as well as the host tissue. Here we discuss several properties, including scaffold fabrication, assembly, structure, biocompatibility, biodegradability and mechanical properties, which must be considered in order to achieve successful tissue regeneration.

15.2 Natural Polymers for Scaffold Fabrication

There are two main types of natural polymers that have been used in tissue engineering applications: polysaccharides and polypeptides (Table 15.1). Polysaccharides are a chain of sugar units attached by a glycosidic bond, while polypeptides are a chain of amino acids attached by a peptide bond. Since these polymers originate from nature, they are often biocompatible and enzymatically biodegradable. The main advantage for using natural polymers is that they often contain biofunctional molecules that aid in the attachment, proliferation, and differentiation of cells. Principle disadvantages of natural polymers are that their degradation may inhibit cell function and their rate of degradation may not be easily controlled. Since enzymatic activity can vary between hosts, it can be difficult to determine the lifespan of natural polymers when placed within a specific tissue. Additionally, the mechanical properties of natural polymers are often low, although covalent chemistries for crosslinking these polymers have shown to enhance structural stability. Finally, further chemical

Table 15.1 Properties and structural unit of degradable natural polymers used to design tissue engineering scaffolds

Natural Polymer	Structural Unit	Curing Method		Degradation Method	Degradation Products
		Crosslinking	Entanglement		
Agarose			X	Enzymatic agarases	Oligosaccharides
Alginate		X		Alginate lyases	Mannuronic acid and guluronic acid
Chitosan		X		Enzymatic chitosanase	Glucosamines
Hyaluronic Acid			X	Enzymatic hyaluronidase	$\beta(1-4)$ Linked glucuronic acid and glucosamine
Collagen[a]		X		Enzymatic collagenase or lysozyme	Various peptides depending on sequence
Gelatin[a]		E	X	Enzymatic collagenases	Various peptides depending on sequence
Silk[b]			X	Proteolytic proteases	Various peptides depending on sequence

[a]Collagen and gelatin have variable sequences, but are predominately glycine, proline, and hydroxyproline
[b]Silks structure contains repeating glycine residues repeating every second or third amino acid
The variable residues are primarily alanine or serine

modification of natural polymers is sometimes challenging due to their complex structure and varied composition.

15.2.1 Polysaccharides

15.2.1.1 Agarose

Agarose is a polysaccharide polymer with alternating two main monomer units: 1,4-linked 3,6 anhydro-α-L-galactose and 1,3-linked β-D-galactose (Table 15.1) [1]. Agarose is a natural polymer that is predominately obtained by isolation from algae. The presence of hydroxyl groups on the polymer backbone induces polarity within agarose. Therefore, two agarose chains may interact with each other through hydrogen bonding to form a double helix. Additionally, agarose is a water soluble polymer that displays temperature sensitive properties. At high temperatures agarose is water soluble while at low temperatures aqueous agarose solutions form into a gel. Additionally, the properties of agarose, such as strength and permeability, can be adjusted by changing the concentration of the polymer. By increasing the concentration of agarose, the mechanical strength typically increases and the porosity or void space between polymer chains decreases. The degradation of agarose occurs mainly by enzymes into oligosaccharides.

The ability for agarose to form into a hydrogel has led to its use in tissue engineering applications. Agarose gels have been shown to provide a three-dimensional environment that allows for the proliferation of cells and retention of their phenotype [2]. Several different cell types have been encapsulated and have proliferated within agarose gels. For example, chondrocytes, the predominant cell type in cartilage, have been cultured in agarose for 2–6 weeks [3]. During this period of time, chondrocytes maintain their expression of type II collagen and proteoglycans [2, 4]. Another cell type that has been investigated within agarose gels is dorsal root ganglia (DRG), a neural cell. By altering the porosity of the gels and incorporating chitosan into the agarose matrix, an increase in nerve regeneration by chick DRG cells was demonstrated [5, 6].

15.2.1.2 Alginate

Alginate, a natural polysaccharide that is also extracted from algae, has two repeating monomer units, β-D-mannuronate and α-L-guluronate, that are joined by a β(1-4) linkage (Table 15.1) [1, 7, 8]. Additionally, alginate is a polyanion that exhibits polar properties and therefore water solubility. Alginate is readily degraded by naturally occurring enzymes, including alginate lysases. These enzymes primarily degrade alginate at the β(1-4) linkage site [7]. Alginate can be crosslinked to form a hydrogel when it is exposed to divalent ions, such as calcium [9, 10]. The physical and mechanical properties of alginate hydrogels are dependent on the length and proportion of the guluronate block within the chain. The crosslinks of alginate hydrogels can be broken down when the cations are removed or sequestered by

chelating agents. The ease of solubilizing alginate gels makes these materials especially attractive for cell encapsulation studies. Alginate does not contain cellular recognition proteins, limiting cell attachment to the natural polymer. By crosslinking alginate with poly(ethylene glycol)-diamine, the mechanical properties of alginate can be controlled and even improved [1].

Investigations have shown that alginate scaffolds can sustain the growth of various cell types [7, 11–14]. Chondrocytes maintained a round morphology and expressed type II collagen within alginate hydrogels [15]. Other forms of alginate, such as a sponge, was found to support the adhesion of fibroblasts [11]. Hepatocytes, the functional cell type in liver, were also found to proliferate within alginate scaffolds. Additionally, hepatocytes seeded within three-dimensional porous alginate secreted albumin, indicating proper cell function [16]. Other cells demonstrating function within alginate hydrogels include cardiomyocytes, rat marrow cells and Schwann cells [7, 14, 17].

15.2.1.3 Hyaluronic Acid

Hyaluronic acid is a polysaccharide that is also known as hyaluronan or HA (Table 15.1). Hyaluronic acid is a major glycosaminoglycan in the extracellular matrix of tissues that promotes early inflammation critical for wound healing [18]. Structurally, hyaluronic acid is composed of the monomers D-glucuronate and N-acetyl-D-glucosamine that are stabilized with a β(1-3) linkage. In dilute concentrations, the structure of hyaluronic acid is a random coil. At higher concentrations, the coils become entangled, creating a stabilized network [19]. Similarly to most natural polymers, hyaluronic acid displays non-immunogenic and non-adhesive properties. Therefore, it is highly attractive for biomedical applications. Some disadvantages of hyaluronic acid are that it is highly water soluble and rapidly degradable by enzymes, including hyaluronidase. However, when fabricated into a hydrogel, hyaluronic acid is typically more stable [1, 5, 18].

Hyaluronic acid has been used in conjunction with other natural or synthetic polymers to recapitulate the extracellular matrix structure. By creating such an environment, the differentiation, proliferation, and function of many cell types may be enhanced. For instance, hyaluronic acid has shown to facilitate the differentiation of mesenchymal progenitor cells into osteoblasts and chondrocytes [5]. Furthermore, chondrocytes embedded in hyaluronic acid have expressed type II collagen, forming a tissue similar to native cartilage [20]. These results indicate that hyaluronic acid hydrogels may be well suited for osteochondral repair. Alternatively, an ester derivative of hyaluronic acid has been widely investigated for tissue engineering applications [5, 20, 21]. For example, adipose precursor cells proliferated and differentiated in these ester modified hyaluronic acid sponges [50, 21].

15.2.1.4 Chitosan

Chitin is a cellulose-like polymer that contains unbranched chains of N-acetyl-D-glycosamine. This polymer is highly concentrated in, and therefore typically

isolated from, the exoskeletons of crustaceans and insects. Chitosan is a positively charged polymer that is obtained from deacetylated chitin (Table 15.1). Thus, chitosan contains D-glucosamine with randomly dispersed N-acetyl-D-glycosamine groups. These monomers are linked at the β(1-4) site which makes it a polysaccharide [1, 19, 22, 23]. Properties of chitosan are similar to hyaluronic acid due to their common monomer units. Therefore, chitosan closely simulates glycosaminoglycans within the extracellular matrix of tissues [1, 5]. Chitosan is easily degraded by chitosanase enzymes and lysozymes. Furthermore, chitosan is cationic with semicrystalline properties of a high charge density. These attributes allow chitosan to be insoluble above pH 7 and fully soluble below pH 5. At high pH solutions, chitosan can be gelled into strong fibers [23]. Additionally, the cationic nature of chitosan allows it to form polyelectrolytic complexes with anionic polymers. Chitosan hydrogels can also be formed by either ionic bonding or covalent crosslinking, using crosslinking agents such as glutaraldehyde [1].

Chitosan has been established as a well defined matrix and as a porous scaffold. Hepatocytes were found to maintain a rounded morphology and to secrete albumin when cultured within chitosan scaffolds [24]. A different chitosan macrostructure, such as crosslinked gels, have also been investigated to encapsulate hepatocytes. Results indicated that an increased level of urea formation occurred [25]. Furthermore, chitosan has shown promise as an orthopedic scaffold. Bone formation has been demonstrated by rat osteoblasts grown on a sponge form of chitosan [26]. Chitosan scaffolds have also been shown to support the attachment and expression of extracellular matrix components by chondrocytes [27].

15.3 Polypeptides

15.3.1 Collagen

Collagen is a structural protein that is a major component of musculoskeletal tissues (Table 15.1). Therefore, it plays a role in cell function as well as tissue strength. Currently there are at least 19 different types of collagen present within the body, the most abundant of these being type I collagen [28]. Types II and III collagen are also readily found in the extracellular matrix of many tissues. Collagen is composed of three polypeptide chains intertwined with one another to form a triple helix, providing stability within tissues. The three polypeptide chains are similar in sequence, with a repeating sequence of glycine-X-Y, where X and Y are most often proline and/or hydroxyproline. The triple helical structure is formed through hydrogen bonding of the peptide bond present in glycine and an adjacent peptide carbonyl group [29]. Collagen is surface degradable by collagenase enzymes, which can bind to the triple helix. This results in the helix unraveling and increasing the exposure of the polypeptide chains to additional degradation enzymes. Therefore, collagen has a tendency to degrade abruptly, limiting its long term use. Investigations have

shown that crosslinking collagen with glutaraldehyde can decrease its degradation rate [30].

Collagen is a component of natural wound healing, promoting blood coagulation [28]. Additionally, collagen contains cell adhesion domains. These attributes have led collagen to be used as a scaffold for cells, such as fibroblasts, within the epithelium [31–33]. Collagen scaffolds further enable fibroblasts to function appropriately and, in turn, maintain collagen production. In native cartilage, chondrocytes are known to synthesize type II collagen. Therefore, type II collagen has been studied for attaching chondrocytes and aiding in its proliferation and differentiation [34]. Hepatocytes isolated from rats have also been found to attach and secrete albumin while grown on collagen and chitosan composite scaffolds [35]. The growth and proteoglycan synthesis of corneal keratocytes have also been demonstrated on collagen sponges [36].

15.3.1.1 Gelatin

Gelatin is derived by denaturing the triple-helix of collagen into single-stranded molecules (Table 15.1). Gelatin is water soluble and displays temperature sensitive properties. In aqueous solutions, gelatin entangles easily to form coils. At low temperatures, gelatin is more likely to form a helical structure similar to collagen [1, 10]. However, this property is influenced by the presence of covalent crosslinks, gelatin molecular weight and gelatin concentration. Since gelatin is a derivative of collagen, it can be enzymatically broken down by collagenases. Gelatin hydrogels typically have been poor mechanical stability, though there have been investigations to stabilize gelatin hydrogels with chemical crosslinking agents such as glutaraldehyde [19].

Since gelatin closely resembles collagen in structure and function, it has been readily used to support a variety of cells. For instance, gelatin has been used to support cells for orthopedic applications. Rat marrow stromal osteoblasts encapsulated in gelatin microparticles retained their phenotype [37]. Porous gelatin scaffolds supported human adipose derived stem cell attachment and differentiation into a variety of cell lineages. These constructs expressed a chondrogenic phenotype, indicated by the expression of hydroxyproline and glycosaminoglycans (GAG), both of which are present in the extracellular matrix of cartilage [38]. Furthermore, non-orthopedic cell types, such as respiratory epithelial cells and cardiocyocytes, have been shown to attach and maintain round morphologies in gelatin scaffolds [14, 39].

15.3.1.2 Silk

Silks are biopolymers that are fibrous with a protein polymer basis (Table 15.1) [40]. The two primary sources of silks are from spiders and silkworms. While both silks display different properties, they are structurally similar. The primary structure of silks is a repetitive sequence of proteins that are glycine-rich. The secondary structure of silks is a β-sheet, which allows silks to exhibit enormous mechanical strength and is the primary reason for its strong interest as a scaffold. Silks have

been found to be both stronger than steel and very elastic. Silks primarily undergo proteolytic degradation by proteases, such as chymotrypsin [40]. The structure of silk allows for slow degradation and strength maintenance. Since silks and their components are xenogenic tissue components, they may cause an allergic response. Additionally, the host tissue can form a granuloma, where connective tissue and capillaries conglomerate around the silk, leading to further inflammation or blood clotting. The use of silks as a scaffold material for cell attachment and function has been investigated. Results show that a positive response occurs by fibroblasts on fibroin silks. And recently, osteoblasts seeded on silks have displayed bone growth characteristics [41, 42].

15.4 Synthetic Polymers for Scaffold Fabrication
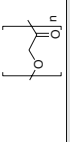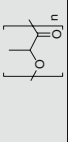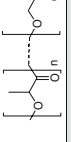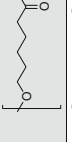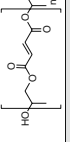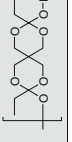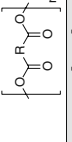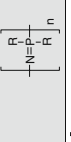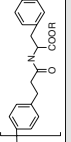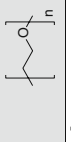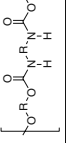
Polymers that are chemically synthesized offer several notable advantages over naturally derived polymers. Most notably, the molecular structure of synthetic polymers can be tailored to suit specific functions and thus exhibit controllable properties. Synthetic polymers are found in either a semicrystalline or an amorphous state. A semicrystalline polymer network contains dense chain regions, which are not uniformly distributed. These regions act as crosslinks and contribute to the mechanical strength of these polymer gels [43]. When semicrystalline polymers are above their melting temperature, they can be easily molded. Amorphous polymers act like rubber or glass, depending on whether the polymer is below or above its glass transition temperature respectively [43]. Their structure can be altered through chemical bonding by copolymerization, or physical mixing combined by blending. Synthetic polymers typically lack important biomolecules that aid in cell attachment. This issue has been addressed by modifying their surface with biomolecules to stimulate cell attachment and growth. Since many synthetic polymers undergo hydrolytic degradation, a scaffold's degradation rate should not vary significantly between hosts. A disadvantage of synthetic polymers is that they may degrade into unfavorable products, such as acids. At high concentrations of these degradation products, local acidity may increase, resulting in adverse responses such as inflammation or fibrous encapsulation [5, 44, 45]. Additionally, the processing technique of synthetic polymers often requires the use of an organic solvent to act as a solubilizer. Residual solvents can lead to adverse effects in the host. The most common synthetic polymers are polyesters. Other types include polyanhydrides, polyphosphazenes, polycarbonates, poly(ethylene glycol) and polyurethane (Table 15.2).

15.4.1 Polyesters

15.4.1.1 Poly(Glycolic Acid)

Poly(glycolic acid) (PGA) is a polyester that belongs to a family of poly(α-hydroxy esters) (Table 15.2). PGA is an aliphatic polymer, which implies that its backbone

Table 15.2 Properties and structural unit of degradable synthetic polymers used to design tissue engineering scaffolds

Synthetic Polymer	Structural Unit	Curing Method		Degradation Method	Degradation Products
		Crosslinking	Entanglement		
Poly(glycolic acid)			X	Ester hydrolysis	Glycolic acid
Poly(L-lactic acid)			X	Ester hydrolysis	Lactic acid
Poly(D,L-lactic acid-co-glycolic acid)			X	Ester hydrolysis	Lactic acid and glycolic acid
Poly(ε-caprolactone)			X	Ester hydrolysis	Caproic acid
Poly(propylene fumarate)		X		Ester hydrolysis	Fumaric acid and propylene glycol
Polyorthoester			X	Ester hydrolysis	Various acids depending upon R group
Polyanhydride			X	Anhydride hydrolysis	Various acids depending upon R group
Polyphosphazene			X	Hydrolysis	Phosphate and ammonia
Polycarbonate(tyrosine derived)			X	Ester and carbonate hydrolysis	Alkyl alcohol and desaminoyrosyl-tyrosine
Poly(ethylene glycol)			X	Non-degradable	Not applicable
Polyurethane		X		Ester, urethane, or urea hydrolysis	Diisocyanate and diols

of carbon atoms are linked linearly or branched without aromatic structures. PGA is formed by ring-opening polymerization of cyclic diesters of glycolide [44]. Since PGA is a linear polymer, it is readily formed into fibers. However, it can be formed into various macrostructures through various casting techniques. PGA is a highly crystalline polymer inherited from the properties of glycolide. Therefore, it has a high melting point (185–225 °C) and low solubility in organic solvents [45, 46]. PGA is also hydrophilic and undergoes bulk degradation, often leading to a sudden loss in mechanical strength. PGA does not undergo enzymatic degradation, but is degraded hydrolytically into glycolic acid, which can be metabolized by the body. This aspect has led to the use of poly(glycolic) acid as a biodegradable material. However, glycolic acid at high concentrations has been found to lower the pH of the surrounding tissue, possibly causing inflammation and tissue damage [44].

The versatility of poly(glycolic acid) in the medical field has led to its use as a scaffold for tissue engineering applications. Bovine chondrocytes seeded on PGA scaffolds in vitro expressed sulfated glycosaminoglycans 25% more than native cartilage [47]. Furthermore, in vivo results using a mouse model have been found to be comparable to the in vitro results. Other studies seeded myofibroblasts and endothelial cells on a PGA fiber matrix. These cells persisted and expressed elastic filaments and collagen, demonstrating the potential use of PGA scaffolds for heart valve engineering [48]. Hepatocytes have also attached onto PGA, which initiated albumin synthesis [49]. Some investigations have coated PGA with other poly(α-hydroxy esters). These PGA coated scaffolds have shown support for muscle and endothelial cell growth [50].

15.4.1.2 Poly(L-lactic acid)

Poly(lactic acid) is a another type of polyester that is classified as a linear poly(α-hydroxy acid). This linear polymer can be formed into three isomers: poly(L-lactic) acid, poly(D-lactic acid) and poly(D,L-lactic acid). Poly(L-lactic acid) (PLLA) is commonly used in tissue engineering applications (Table 15.2). PLLA is formed by ring-opening polymerization of L-lactide [51]. The structure of PLLA is similar to PGA. However, PLLA contains a pendant methyl group, increasing the hydrophobicity of the polymer and lowering the melting temperature to 170–180 °C [46]. Additionally, the methyl group hinders the ester bond cleavage of PLLA, decreasing the degradation rate of the polymer [44]. PLLA typically undergoes bulk, hydrolytic ester-linkage degradation, decomposing into lactic acid. The degradation product of PLLA is naturally produced by the body. It has been noted that the body can excrete lactic acid through the respiratory system in the form of water and carbon dioxide [45, 51]. However, large concentrated quantities of lactic acid can be harmful to the tissue surrounding the implant due to its acidic properties.

PLLA has been found to support the attachment of various cell types onto its surface. However, investigations have shown that less cells adhere to PLLA compared to other polymers, such as PGA [52]. Even with this result, the ability for PLLA to provide support for cell attachment and function has led to its continued use in tissue engineering applications. Human bladder smooth muscle cells have been seeded

onto PLLA scaffolds and presented normal metabolic function and cell growth [13]. Additionally, cartilage formation has been shown to occur when chondrocytes were attached onto PLLA scaffolds by displaying collagen, glycosaminoglycan, and elastic expression [53]. PLLA fibers have shown great promise in allowing cells to differentiate into the appropriate cell lineage. Stem cells differentiated and expressed positive cues for neurite growth for potential regeneration of neurons when they were seeded on PLLA [54].

15.4.1.3 Poly(D,L-lactic acid-co-glycolic acid)

Poly(lactic acid) and poly(glycolic acid) can be copolymerized to form poly(D,L-lactic acid-*co*-glycolic acid) (PLGA) (Table 15.2). The properties of the copolymer are very different from the individual polymers. PGA and PLA are semicrystalline polymers. However, PLGA is amorphous, which is a solid, non-crystalline structure. Therefore, this copolymer can have varying properties depending on the ratio of PLA and PGA. PLGA has been found to degrade slower as well as demonstrate lower mechanical strength compared to the individual crystalline polymers. The presence of ester linkages makes PLGA susceptible to ester hydrolysis and has been shown to primarily undergo bulk degradation. However, it is important to note that the copolymer composition is not linearly related to the mechanical and degradation properties of PLGA. For example, a 50:50 ratio of PGA and PLLA typically degrades faster than either homopolymer [43, 46].

The makeup of PLGA from biodegradable polymers and their adjustable properties has led to their –use as a tissue engineering scaffold. PLGA has been found to support a variety of cell types. Orthopedic cell types seeded onto PLGA scaffolds have been found to express extracellular matrix components [52, 55]. For instance, osteoblastic cells synthesized collagen, fibronectin, and laminin. Additionally, PLGA can facilitate human epithelial fibroblasts to regenerate an extracellular matrix [5]. An epithelial layer was formed when enterocytes, cells derived from the intestine epithelium, were seeded on PLGA [56]. Furthermore, PLGA promoted smooth muscle cell growth as well as the production of collagen and elastin [57]. Similar works showed enhancement of gene expression by neural cells on PLGA scaffolds [58].

15.4.1.4 Poly(ε-caprolactone)

Poly(ε-caprolactone) (PCL) is an aliphatic polyester with one repeating ester group and five methylene groups (Table 15.2). PCL is highly water soluble with semicrystalline properties and has a melting temperature of 58–63 °C [19, 46]. This polyester is formed by ring-opening polymerization of ε-caprolactone. These monomers are linked together by a hydrolytically degradable ester bond. The degradation of PCL occurs by bulk or surface hydrolysis, resulting in a byproduct of caproic acid. At low concentrations, the acidic byproducts do not cause a significant negative reaction in the surrounding tissue. Therefore, these scaffolds may be more effective when the acidic byproducts are not released at a high concentration. PCL degrades

at a relatively slow rate compared to other polyesters. It has been shown to maintain its structure in vivo for up to two years [46]. By copolymerizing PCL with other polymers, the degradation rate can be altered. For instance, to increase the degradation rate, PCL can be copolymerized with collagen, poly(glycolic acid), poly(lactic acid), and poly(ethylene glycol) [59–61]. The ease with which PCL can be copolymerized with a variety of polymers has made it an attractive component of polymeric scaffolds.

PCL has been investigated as a stabilizing polymer on PGA scaffolds to aid the formation of spherical aggregates by human biliary epithelial cells (hBEC). These hBECs proliferated on the synthetic scaffold and expressed phenotypic proteins, indicating cell stability [60]. Further studies of PCL as a homopolymer has demonstrated its support of human osteoblast and dermal fibroblast cell viability [62, 63]. The results with osteoblasts on porous PCL has shown a production of alkaline phosphatase, a marker of bone mineralization [62]. Also, dermal fibroblasts on mechanically stretched PCL films proliferated and maintained their round morphology [63].

15.4.1.5 Poly(propylene fumarate)

Poly(propylene fumarate) (PPF) is a polyester with a repeating unit consisting of two ester groups and one central carbon-carbon double bond (Table 15.2). The presence of an ester bond within PPF allows it to degrade by hydrolysis. The primary degradation byproducts of PPF are fumaric acid and propylene glycol. Recent investigations have shown that the byproducts cause mild and short inflammation, thus indicating biocompatibility [64]. The unsaturated double bonds allow for covalent crosslinking of PPF to occur. A cured form of PPF has been shown to possess significant compressive and tensile strength. Furthermore, this triggerable crosslinking aids in the formation of a synthetic scaffold in situ [65, 66]. Thus PPF can be used as an injectable polymer for tissue engineering applications. Additional research has also been conducted with PPF copolymerized with PEG. PPF is hydrophobic but with the incorporation of PEG, it has increased hydrophilicity. The PPF and PEG copolymer has been shown to support endothelial cell attachment and proliferation [67, 68].

15.4.1.6 Polyorthoester

There are many types of polyorthoesters (POE) that have been synthesized. The main form considered for tissue engineering applications has been created through the reaction of ketene acetals with diols (Table 15.2) [69]. This form has led POE to exhibit hydrophobic properties and undergo surface degradation [5, 70]. Through copolymerization, the properties of POE can be adjusted or modified to exhibit certain attributes. For instance, the degradation rate of the copolymer can be controlled by incorporating short acid groups, such as glycolic or lactic acid [70]. The orthoester linkages within POE have been found to be more hydrolytically sensitive in acids compared to bases [19, 43]. The ability for POE to undergo surface erosion has primarily led to its use in bone reconstruction [5, 71]. Surface degradation

allows a scaffold to sustain mechanical support for the surrounding tissue since the bulk of the material remains structurally intact. In other studies, hepatocytes were grafted onto polyorthoesters and adhered to the surface of the synthetic scaffold [72].

15.4.2 Other Synthetic Polymers

15.4.2.1 Polyanhydride

Polyanhydrides are a class of polymers containing anhydride bonds, which tend to be highly water liable (Table 15.2) [73]. Polyanhydrides are hydrophobic polymers that exist in a crystalline state with a melting temperature of approximately $100\,°C$ [74]. Aliphatic polyanhydrides are synthesized by a dehydration reaction between diacids. The degradation rate for polyanhydrides is fairly quick due to the instability of the anhydride bonds and investigations have shown that these polymers can degrade in a period of days to weeks. However the properties of polyanhydrides can be changed through the diacid building block. One aspect that is easily modifiable is the degradation rate, which can be slowed down by increasing the hydrophobicity of its backbone [74]. Polyanhydrides are generally thought to degrade following a surface degradation mechanism. Additionally, the ability of polyanhydrides to be photocrosslinked has led to its application as an injectable polymer curing in situ [75, 76].

The similarity of polyorthoesters to polyanhydrides has focused their use for primarily orthopaedic tissue engineering applications. However, the modest mechanical properties of entangled polyanhydride networks has limited their application in weight bearing environments [45]. This limitation resulted in further studies involving the formation of crosslinkable polyanhydride networks with incorporated imides [51, 77]. The imide bonds that formed were found to undergo hydrolysis, allowing for degradation of the poly(anhydrides-*co*-imides) copolymer. Scaffolds formed from these networks possess significant mechanical properties, such as a compressive strength of 30–60 MPa, allowing for their use in some weight bearing orthopedic sites [45, 74, 77, 78].

15.4.2.2 Polyphosphazene

Polyphosphazene consists of an inorganic backbone containing nitrogen and phosphorous atom (Table 15.2). Polyphosphazene is hydrophobic and typically undergoes hydrolytic surface degradation into phosphate and ammonium salt byproducts. There are a variety of polyphosphazenes that may be synthesized, due to the numerous types of hydrolytically labile substituents that can be added to the phosphorous atoms [79]. These viable substituents allow the properties of polyphosphazenes to be extremely controllable. Nevertheless, most polyphosphazenes display a slow degradation rate in vivo [79].

A number of different polyphosphazenes have been synthesized for use as tissue engineering scaffolds. Due to its high strength and surface degradation properties, it has been particularly investigated as an orthopedic biomaterial [80]. Osteoblast cells that support skeletal tissue formation have been seeded and proliferated on a three-dimensional polyphosphazene scaffold [81]. Additionally, poly(organo)phosphazenes have been studied as scaffolds for synthetic nerve grafts in peripheral nerve regeneration [82].

15.4.2.3 Polycarbonate

Tyrosine-derived polycarbonate (P(DTR carbonate)) is an amorphous polycarbonate well studied for biomedical applications (Table 15.2). The presence of a pendant group within the linear chain of P(DTR carbonate) allows it to be easily modified by attaching various functional groups [83]. There are three bonds within P(DTR carbonate) that are susceptible to hydrolytic degradation: amide, carbonate, and ester bond [51, 83]. The carbonate bonds have been found to degrade at a faster rate compared to the ester bonds. However, the period of time that P(DTR carbonate) takes to degrade can range from months to years [84]. The ester bond degrades into carboxylic acid and alcohol, while the carbonate bond releases two alcohols and carbon dioxide [83]. The amide has been found to be relatively stable in vitro [83]. Overall, P(DTR carbonate) undergoes bulk degradation. Its mechanical properties are determined by the pendant group, which could create possible crosslinking networks [51, 83].

Since the degradation rate of polycarbonates is relatively slow, it has been primarily used for orthopedic tissue engineering applications. For instance, P(DTR carbonate) elicits a response for bone ingrowth at the bone-polymer implant interface, supporting the use of P(DTR carbonate) as a bone scaffold [84]. Recent studies has shown that osteoblast cells attach onto the surface of P(DTR carbonate). Results indicate that these osteoblasts maintained their phenotype and rounded cell morphology [85].

15.4.2.4 Poly(ethylene glycol)

Poly(ethylene glycol) (PEG) is a linear-chained polymer consisting of a simple oxygen-carbon-carbon repeating unit (Table 15.2). The number of repeating units determines the length and therefore the molecular weight of PEG [86, 87]. PEG is synthesized by an anionic/cationic polymerization [1, 5]. Unmodified PEG is non-degradable. However, PEG may be copolymerized with degradable polymers to allow for material degradation [88]. Due to the oxygen molecule within the backbone, PEG is highly water soluble. When PEG is copolymerized it can contribute significant hydrophilicity to the substituent polymer. The ability of PEG to act as a swelling polymer has primarily led to its function as a hydrogel, and in some instances an injectable hydrogel. Unfortunately, the linear chain form of PEG leads to rapid diffusion and low mechanical stability [5]. PEG networks may be created by

attaching functional groups to the ends of the PEG chain and then initiating covalent crosslinking [5, 29, 87, 89–91]

There are many biomedical applications that incorporate PEG. This is due to the ease in which it can undergo modification either by crosslinking groups for network formation or creating degradable groups for resorbable applications. PEG has been copolymerized with various polymers to create a construct with controlled erosion methods or degradation rates. For example, PEG has been copolymerized with PLGA as well as alginate to form hydrogels. Results have shown that chondrocyte proliferation on copolymerized PLGA-PEG increases with hydrogel degradation rate [92]. Similarly, islets of Langerhans in alginate-PEG hydrogels retained their viability and expressed function through insulin excretion [93].

15.4.2.5 Polyurethane

Recent investigations have developed polyurethane polymers as scaffold materials. Polyurethane is an elastomeric polymer that is typically non-degradable (Table 15.2). Positive attributes, such as flexible mechanical strength and biocompatibility, have led to the synthesis of degradable polyurethanes with non-toxic diisocyanate derivatives [45, 94–96]. Functional groups can be attached to the surface of polyurethanes to aid in cell attachment. Additionally, polyurethane can be modified by incorporating different structural polymers, such as branched or star shaped polymers. Studies have shown that polyurethane scaffolds support cell attachment with chondrocytes, bone marrow stromal cells, and cardiomyocytes [96–98].

15.5 Fabrication Techniques

The fabrication of a biomaterial is an important aspect of polymer technology to consider when constructing a tissue engineering scaffold. There are several conventional techniques readily used to develop porous matrices including fiber bonding, solvent-casting particulate-leaching, gas foaming, freeze drying, phase separation, and melt molding. These techniques tend to be mainly process driven. However, there has been a growing interest in creating design driven techniques such as rapid prototyping or solid freeform fabrication.

15.5.1 Conventional Techniques

15.5.1.1 Fiber Bonding

Fibers are often processed from semicrystalline polymers, such as PGA or PLA. A fiber mesh is a three-dimensional patterned structure created from individual polymer fibers. This construct contains a large surface area with a porous network allowing for greater cell attachment as well as diffusion of nutrients and waste material [46]. However, due to the high porosity of a mesh structure there is often limited

structural stability. Therefore, a fiber bonding technique was developed to resolve this issue [44, 46, 99, 100]. In one case, PLA dissolved in a solvent is cast over a PGA mesh. The construct is heated above the melting temperature of PGA and the solvent is evaporated. By cooling and re-dissolving PLA, the individual fibers within the PGA mesh become connected at the fiber crosspoints. The main disadvantage of this technique is that the solvent being used to dissolve the polymer may be harmful to the host [46].

15.5.1.2 Solvent-Casting Particulate-Leaching

A common technique used for scaffold fabrication is solvent-casting particulate-leaching [101]. This process involves mold casting a solution mixture of dispersed particles and a polymer. A highly interconnected porous scaffold may be generated by leaching out the dispersed particles with water. The void spaces left behind by the particles can form a porous structure. An advantage of this technique is that porosity and pore size can be controlled independently by altering particle concentration and size.

15.5.1.3 Phase Separation

Phase separation is a process that isolates the components of a homogeneous mixture. First, a polymer is dissolved in a solvent [46, 102]. By lowering the temperature of the solution, a phase separation occurs where one phase is polymer rich and one phase is solvent rich. A porous membrane is formed after the solvent is removed [103, 104]. This technique offers an advantage compared to the previous methods mentioned because the moderate conditions allows the incorporation of biomolecules within the structure. Additionally, the structure of the scaffold can be changed by altering polymer material, polymer concentration, phase transitions, and/or solvents. However, the morphology of the resulting scaffold is difficult to predict.

15.5.1.4 Melt Molding

Melt molding is a technique that forms a scaffold by combining a polymer powder and microspheres into a mold [44]. The polymer powder adheres together when pressure is applied to this mold and it is heated above the glass transition temperature of the polymer. By removing the molded polymer and placing it in water, the entrapped microspheres can be removed. This technique results in a porous three-dimensional membrane. Pore size is affected by the microsphere diameter and the porosity is determined by the microsphere concentration. Since this technique does not involve organic solvents, bioactive molecules can be incorporated in the scaffold. Additionally, a defined shape of the construct can be formed, allowing for further variability in its design. A disadvantage of this technique is that high temperatures may be required for semicrystalline polymers [46].

15.5.1.5 Freeze Drying

Freeze drying is another temperature altering method to create a porous structure [105]. Natural or synthetic polymers are dissolved into an extremely cold solvent, such as acetic acid or benzene [99]. By combining this solution with water, an emulsion is created [106]. Rapidly freezing this emulsion produces ice crystals, composed of the solvent and water, which may become aggregated. By removing the ice crystals through a freeze drying technique, a highly connected porous matrix can be formed. Pore size can be controlled by altering the freezing rate where typically, a fast rate creates smaller pores. However, pore structure is difficult to control. In order to address the pore structure concern, freeze drying may be altered by combining it with the particulate-leaching technique [107].

15.5.1.6 Gas Foaming

The gas foaming technique uses gases, either pressurized or generated from reactions, to create pores within the scaffold [107, 108]. The formation of bubbles within the polymer leads to the formation of macropores. The porosity and pore structure are dependent on factors such as gas volume, rate of gas nucleation and gas diffusion. An advantage of the gas foaming technique is that it involves no organic solvents. Improvements of this method have been established by incorporating particulate-leaching [109, 110].

15.5.2 Rapid Prototyping or Solid Freeform Fabrication Techniques

Conventional fabrication techniques have significant limitations in controlling scaffold design parameters, such as pore size, pore shape, pore interconnectivity, and wall thickness. This shortcoming has led researchers to investigate rapid prototyping (RP) or solid freeform fabrication (SFF) techniques for tissue engineering applications [111]. RP or SFF are techniques that create reproducible three-dimensional structures layer by layer. Furthermore, a scaffold's architecture can be precisely designed and assembled using computer aided design methods to drive these fabrication techniques. Additionally, this process is performed at room temperature, possibly allowing the scaffold to be constructed with encapsulated cells and biomolecules without degeneration. The major disadvantage of this technique is that it is only applicable to specific polymers. Specific RP or SFF processes are three-dimensional printing (3-DP), sheet lamination, laser stereolithography, photopolymerization, and fused deposition modeling (FDM) [111–117].

15.6 Properties for Scaffold Design

Scaffold materials for tissue engineering must exhibit certain properties in order to achieve tissue regeneration. A number of scaffold properties must therefore be considered, including assembly, surface properties, macrostructure, biocompatibility, biodegradability and mechanical properties. Each of these scaffold properties will be examined in detail below.

15.6.1 Polymer Assembly

The two main considerations in the assembly of a polymeric tissue engineering scaffold are curing methods and fabrication strategies. Curing methods provide details on how polymer chains are formed into a bulk material. The predominant influence upon curing method is the chemical nature of the polymer, specifically polymer length and polymer functionality [118]. The two main curing methods currently under widespread use in tissue engineering applications are polymer entanglement and crosslinking.

Many polymers will associate with one another in solution [119]. Typically, this occurs with long, linear, saturated polymers. A loosely bound network may be formed when the individual polymer chains become entangled. This type of curing method is often achieved when casting the polymer in a mold. The process involves dissolving a polymer in a solvent and pouring the resulting solution into a heated mold. The solvent is removed by evaporation and the polymer is then formed into the shape of the mold using pressure, heat, or both [120, 121]. An advantage of polymer entanglement is that it is a relatively simple process. However, there may be a lack of mechanical strength of constructs formed from entangled polymer chains.

Another curing method to form a bulk material is through crosslinking of individual polymer chains, by physical or chemical bonds. Physical, non-covalent interactions involve individual polymer chains forming hydrogen or ionic bonds among each other [118]. Polymers can also interact with one another through chemical, covalent interactions. In this situation the polymer must have a reactive site, such as an unsaturated carbon-carbon double bond, to allow for crosslinking. Covalent crosslinking is typically propagated by a free radical that, in turn, is initiated by a signal such as heat, light, chemical accelerant, or time [87]. Crosslinked polymers often have significant mechanical properties. Furthermore, since crosslinked polymers are formed by a response to a signal, they can be utilized as an injectable material that is cured in situ. The major disadvantage of crosslinked systems is that the chemical reactions, as well as reactants and byproducts, may create cytotoxic environments.

The curing of a polymer into a tissue engineering scaffold may occur before implantation, as in the case of prefabricated scaffolds, or during implantation, as in the case of in situ fabricated scaffolds [46]. A prefabricated scaffold is often preferred because the engineered construct is formed outside the body, allowing for

removal of cytotoxic and non-biocompatible components prior to its implantation. Additionally, a cell population can be incorporated into the scaffold after fabrication and then cultured in vitro to create a living, viable and transplantable construct. A significant disadvantage of prefabrication is the construct's rigid boundary which often prevents a seamless fit into the defect site. Gaps between the engineered graft and host tissue may lead to undesirable reactions within the host, including fibrous tissue formations and construct failure. In order to try to alleviate this concern, techniques such as rapid prototyping are under development to create precise fitting constructs.

The drawbacks of prefabricated structures have led to a strategy of in situ fabrication of constructs, which involves curing of a polymeric matrix at the site of the tissue defect [87]. The components that form these structures typically involve injectable liquids. The deformability of in situ fabricated matrices allow for improved tissue integration between the scaffold and the surrounding tissue. Additionally, the use of liquid components limits the need for invasive surgical procedures. A disadvantage of in situ fabrication of scaffolds is that the components required for curing may be harmful to the host tissue. Furthermore, the delivery of a cell population to the defect site may be considerably more complicated with an in situ curing system. Effectively, the use of such a system greatly reduces the possible materials and curing chemistries available to the tissue engineer.

15.6.2 Surface Properties

As most cells utilized in tissue engineering are anchorage dependent, it has been reasoned that the scaffold should facilitate their attachment. Therefore the surface properties of the engineered construct are clearly critical for tissue regeneration. The scaffold surface is the initial and primary site of interaction with surrounding cells and tissue. Most tissue engineering strategies will favor surfaces where cells attach abundantly and easily. Thus scaffolds with a large and accessible surface area are favorable. Many natural polymers have functional proteins that will aid in cell attachment. Synthetic polymers, on the other hand, tend to lack the ligands required to attract cells to the scaffold. This issue can be addressed by chemically modifying the surface of the synthetic polymer matrix with short peptide sequences or long protein chains [122]. Common peptides that function as ligands originate from proteins present in the extracellular matrix, such as fibronectin, vitronectin, and laminin [122]. These biomimetic scaffolds have been widely investigated for tissue engineering applications [122–124].

15.6.3 Macrostructure

A scaffold with an interconnected porous network should be fabricated to allow cells, nutrients, and waste to diffuse throughout the tissue defect site. It is

advantageous for a scaffold to have a high surface area to volume ratio, with a continuous network of small diameter pores. However, a disadvantage of a highly porous network is that it tends to compromise the mechanical integrity of the polymeric scaffold. Clearly, a balance of a scaffold's surface area, void space, and mechanical integrity is a necessary challenge that must be overcome in the construction of viable tissue engineering scaffolds.

Fiber meshes are formed by weaving or knitting individual polymeric fibers into a three-dimensional structure. Meshes are attractive for tissue engineering as they provide a large surface area for cell attachment [46]. Furthermore, a fiber mesh scaffold structure mimics the properties of the extracellular matrix, allowing the diffusion of nutrients to cells and waste products from cells. A disadvantage to this form is the lack of structural stability. However, fiber bonding, which involves bonding at fiber cross points, has been introduced to create a more stable structure [100].

Foam and sponge scaffolds provide a structural template for cells to form into a three-dimensional tissue structure. These scaffolds tend to be constructed outside of the host before implantation. Their porous network also simulates the extracellular matrix architecture allowing cells to interact effectively with their environments. Though foams and sponges are more mechanically stable comparatively to mesh structures, they are still limited due to the open spaces present throughout the scaffold. The porosity as well as the pore structure can be manipulated by several different processing techniques, such as solvent-casting particulate-leaching, melt molding, freezing drying, and gas foaming.

Hydrophilic polymers are often formed into hydrogels by physical polymer entanglements, secondary forces (i.e., van der Waals), or crosslinking [10, 87]. The hydrophilic nature of the constituent polymers within hydrogels allows them to absorb tremendous amounts of water; some hydrogels can absorb water up to a thousand times their own dry weight [10]. The aqueous environment is also an ideal setting for cell encapsulation. The quick diffusion of nutrients, proteins, and waste allows hydrogels to maintain cell growth. Hydrogels are generally easy to inject or mold, and therefore are often incorporated into strategies involving minimally invasive surgical techniques [29]. Some drawbacks to hydrogels are that they lack strong mechanical properties and they are difficult to sterilize [10, 29].

15.6.4 Biocompatibility

One critical property of a scaffold for tissue engineering applications is biocompatibility. We define tissue engineered constructs as biocompatible if they elicit a minimal immune and inflammatory response, reflecting the premise that these responses can not be completely eliminated. This definition further implies that the response should not be prolonged and the degradation byproducts of the construct should not lead to a host response.

In order to create a biocompatible scaffold, it is helpful to understand the inflammation and healing response of the body. The body responds to an implant similarly

to an injury of a tissue. A sequence of events takes place that may be classified into three phases [125–127]. Phase I occurs no matter what material is being implanted. It is initiated when any foreign body is introduced and leads to an acute and chronic inflammation response, which lasts for approximately 1–2 weeks. Phase I is independent on the biodegradability of the polymer. The rate of phase II is determined by the degradation rate of the polymer. A granulation and foreign body response is elicited after the first phase. Usually this step occurs at the interface of the polymer to the tissue, typified by a fibrous capsule formation. A fibrous capsule involves the interaction of fibroblasts and formation of blood vessels on the polymeric scaffold. Lastly, phase III is the initiation of fibrosis, which is enhanced as the scaffold becomes further degraded. The rate at which the scaffold is degraded determines the length of this phase, varying from 1 to 2 weeks to several months.

The time length of each phase is dependent on the scaffold's properties, especially degradation rate. Other strong influences include composition, fabrication, surface properties and macrostructure. Additionally, the site of implantation will determine the local conditioning which drives the tissue response.

15.6.5 Biodegradability

Scaffolds can offer either temporary or permanent support. Most tissue engineering strategies, however, envision scaffold support until the host tissue begins to regenerate or heal. In these cases, the scaffold should be biodegradable. The biodegradation of a polymer is greatly affected by its properties, including components, fabrication and structure. Initial degradation of hydrolytically degradable scaffolds occurs at the water liable linkages of the polymer. This leads to the polymer chains become shorter, which allows for the overall degradation of the construct. The byproducts that form should also be biocompatible and then utilized or removed from the body. Another aspect that is important to consider when constructing a polymeric scaffold is the rate of degradation. The degradation rate may affect the length of a host response as well as the ability for the tissue to regenerate in an appropriate period of time [19, 44]. Polymeric scaffolds that degrade at a rate relatively close to the rate of tissue repair may be preferred. However, the rate of tissue repair in the presence of a degrading construct may be difficult to establish.

Polymeric scaffolds can either undergo bulk, surface, or both types of degradation [43, 128]. Polymeric scaffolds that undergo bulk degradation tend to break down quicker internally compared to the surface. Initially when the scaffold is in contact with water, the surface starts to degrade. The bulk of the structure then undergoes degradation as water penetrates into the internal areas. As time progresses, the bulk degradation rate becomes quicker than the surface degradation. This is most often seen when the degradation product catalyzes degradation, as in the case of acid degradation products. The rise in local pH within the scaffold results in an enhanced degradation rate. Furthermore, once the inner matrix has become completely degraded, the mechanical strength of the overall construct decreases.

However, since the scaffold does maintain its surface, it can still provide a stable platform for cells to attach and function properly.

A polymeric scaffold that primarily undergoes surface degradation can be described similarly to the dissolution of soap. The rate at which the surface degrades is usually constant. Therefore, even though the size of the scaffold becomes smaller, the bulk structure is maintained. These types of degrading scaffolds provide longer mechanical stability for the tissue to regenerate. Surface eroding scaffolds tend to have hydrophobic characteristics, making the constituent polymer difficult to break-down in the aqueous environments that are seen in vivo. Natural polymers that degrade enzymatically undergo surface degradation since most of the functional enzymes are too large to penetrate the bulk of the structure.

15.6.6 Mechanical Properties

The location of a scaffold within a tissue determines the importance of its structural stability. Therefore, the properties of the scaffold should be designed to accommodate the stresses and strains of the surrounding tissue. Scaffold mechanical properties are especially important for weight bearing orthopedic tissues. Since the scaffold should provide temporary support, the rate of degradation should not be shorter than the length of time the tissue needs to regenerate. Additionally, the surrounding tissue should be able to maintain its own mechanical role in stress dissipation. In cases where mechanical forces are thought to be required for cell growth and phenotype maintenance, a scaffold which displays surface eroding properties may be preferred.

15.7 Summary

There are numerous variables to consider when constructing an appropriate polymeric scaffold. In this chapter we have discussed some of the important properties that polymeric scaffolds should exhibit as well as how to create and modify them. Natural and synthetic polymers are widely employed as the basic material in tissue engineering scaffolds. Natural polymers can assimilate into the host tissue and provide positive interaction with cells and biomolecules. Synthetic polymers can easily be adjusted and altered to provide properties necessary for tissue regeneration. Both conventional and novel techniques can be utilized in the fabrication of tissue engineering scaffolds. Conventional techniques, such as fiber bonding and particulate-leaching, are simple bench top techniques that produce viable scaffolds. Rapid prototyping or solid freeform fabrication is a developing technique that allows for greater control over scaffold architecture. Finally, an important aspect to consider when constructing a scaffold for tissue regeneration is design. Overall, the critical design parameters include polymer assembly, surface properties, macrostructure, biocompatibility, biodegradability, and mechanical properties.

References

1. Lee KY and Mooney DJ. Hydrogels for tissue engineering. Chem Rev, 2001a, 101(7): 1869–1879.
2. Vinall RL, Lo SH, and Reddi AH. Regulation of articular chondrocyte phenotype by bone morphogenetic protein 7, interleukin 1, and cellular context is dependent on the cytoskeleton. Exp Cell Res, 2002, 272(1):32–44.
3. Hung CT, Lima EG, Mauck RL, Taki E, LeRoux MA, Lu HH, et al. Anatomically shaped osteochondral constructs for articular cartilage repair. J Biomech, 2003, 36(12):1853–1864.
4. Rahfoth B, Weisser J, Sternkopf F, Aigner T, von der Mark K, and Brauer R. Transplantation of allograft chondrocytes embedded in agarose gel into cartilage defects of rabbits. Osteoarthr Cartil, 1998, 6(1):50–65.
5. Seal BL, Otero TC, and Panitch A. Polymeric biomaterials for tissue and organ regeneration. Mater Sci Eng R Rep, 2001, 34(4–5):147–230.
6. Dillon GP, Yu X, Sridharan A, Ranieri JP, and Bellamkonda RV. The influence of physical structure and charge on neurite extension in a 3D hydrogel scaffold. J Biomater Sci Polym Ed, 1998, 9(10):1049–1069.
7. Wang L, Shelton RM, Cooper PR, Lawson M, Triffitt JT, and Barralet JE. Evaluation of sodium alginate for bone marrow cell tissue engineering. Biomaterials, 2003, 24(20): 3475–3481.
8. Davis TA, Volesky B, and Mucci A. A review of the biochemistry of heavy metal biosorption by brown algae. Water Res, 2003, 37(18):4311–4330.
9. Rowley JA, Madlambayan G, and Mooney DJ. Alginate hydrogels as synthetic extracellular matrix materials. Biomaterials, 1999, 20(1):45–53.
10. Hoffman AS. Hydrogels for biomedical applications. Adv Drug Deliv Rev, 2002, 54(1): 3–12.
11. Shapiro L and Cohen S. Novel alginate sponges for cell culture and transplantation. Biomaterials, 1997, 18(8): 583–590.
12. Guo JF, Jourdian GW, and MacCallum DK. Culture and growth characteristics of chondrocytes encapsulated in alginate beads. Connect Tissue Res, 1989, 19(2–4):277–297.
13. Pariente JL, Kim BS, and Atala A. In vitro biocompatibility evaluation of naturally derived and synthetic biomaterials using normal human bladder smooth muscle cells. J Urol, 2002, 167(4):1867–1871.
14. Shimizu T, Yamato M, Kikuchi A, and Okano T. Cell sheet engineering for myocardial tissue reconstruction. Biomaterials, 2003, 24(13):2309–2316.
15. Liu H, Lee YW, and Dean MF. Re-expression of differentiated proteoglycan phenotype by dedifferentiated human chondrocytes during culture in alginate beads. Biochim Biophys Acta, 1998, 1425(3):505–515.
16. Glicklis R, Shapiro L, Agbaria R, Merchuk JC, and Cohen S. Hepatocyte behavior within three-dimensional porous alginate scaffolds. Biotechnol Bioeng, 2000, 67(3):344–353.
17. Mosahebi A, Simon M, Wiberg M, and Terenghi G. A novel use of alginate hydrogel as Schwann cell matrix. Tissue Eng, 2001, 7(5):525–534.
18. Leach JB, Bivens KA, Patrick CW, and Schmidt CE. Photocrosslinked hyaluronic acid hydrogels: Natural, biodegradable tissue engineering scaffolds. Biotechnol Bioeng, 2003, 82(5):578–589.
19. Hayashi T. Biodegradable polymers for biomedical uses. Prog Polym Sci, 1994, 19(4): 663–702.
20. Aigner J, Tegeler J, Hutzler P, Campoccia D, Pavesio A, Hammer C, et al. Cartilage tissue engineering with novel nonwoven structured biomaterial based on hyaluronic acid benzyl ester. J Biomed Mater Res, 1998, 42(2):172–181.
21. Halbleib M, Skurk T, de Luca C, von Heimburg D, and Hauner H. Tissue engineering of white adipose tissue using hyaluronic acid-based scaffolds. I: in vitro differentiation of human adipocyte precursor cells on scaffolds. Biomaterials, 2003, 24(18):3125–3132.

22. VandeVord PJ, Matthew HWT, DeSilva SP, Mayton L, Wu B, and Wooley PH. Evaluation of the biocompatibility of a chitosan scaffold in mice. J Biomed Mater Res, 2002, 59(3): 585–590.

23. Suh JKF and Matthew HWT. Application of chitosan-based polysaccharide biomaterials in cartilage tissue engineering: a review. Biomaterials, 2000, 21(24):2589–2598.

24. Li J, Pan J, Zhang L, Guo X, and Yu Y. Culture of primary rat hepatocytes within porous chitosan scaffolds. J Biomed Mater Res, 2003, 67A(3):938–943.

25. Kawase M, Michibayashi N, Nakashima Y, Kurikawa N, Yagi K, and Mizoguchi T. Application of glutaraldehyde-crosslinked chitosan as a scaffold for hepatocyte attachment. Biol Pharm Bull, 1997, 20(6):708–710.

26. Seol YJ, Lee JY, Park YJ, Lee YM, Ku Y, Rhyu IC, et al. Chitosan sponges as tissue engineering scaffolds for bone formation. Biotechnol Lett, 2004, 26(13):1037–1041.

27. Nettles DL, Elder SH, and Gilbert JA. Potential use of chitosan as a cell scaffold material for cartilage tissue engineering. Tissue Eng, 2002, 8(6):1009–1016.

28. Lee CH, Singla A, and Lee Y. Biomedical applications of collagen. Int J Pharm, 2001b, 221(1–2):1–22.

29. Drury JL and Mooney DJ. Hydrogels for tissue engineering: scaffold design variables and applications. Biomaterials, 2003, 24(24):4337–4351.

30. Ma L, Gao C, Mao Z, Shen J, Hu X, and Han C. Thermal dehydration treatment and glutaraldehyde cross-linking to increase the biostability of collagen-chitosan porous scaffolds used as dermal equivalent. J Biomater Sci Polym Ed, 2003, 14(8):861–874.

31. Langer R and Vacanti JP. Tissue engineering. Science, 1993, 260(5110):920–926.

32. Falanga V, Margolis D, Alvarez O, Auletta M, Maggiacomo F, Altman M, et al. Rapid healing of venous ulcers and lack of clinical rejection with an allogeneic cultured human skin equivalent. Human Skin Equivalent Investigators Group. Arch Dermatol, 1998, 134(3): 293–300.

33. Noah EM, Chen J, Jiao X, Heschel I, and Pallua N. Impact of sterilization on the porous design and cell behavior in collagen sponges prepared for tissue engineering. Biomaterials, 2002, 23(14):2855–2861.

34. Nehrer S, Breinan HA, Ramappa A, Hsu HP, Minas T, Shortkroff S, et al. Chondrocyte-seeded collagen matrices implanted in a chondral defect in a canine model. Biomaterials, 1998, 19(24):2313–2328.

35. Risbud MV, Karamuk E, Schlosser V, and Mayer J. Hydrogel-coated textile scaffolds as candidate in liver tissue engineering: II. Evaluation of spheroid formation and viability of hepatocytes. J Biomater Sci-Polym Ed, 2003, 14(7):719–731.

36. Orwin EJ and Hubel A. In vitro culture characteristics of corneal epithelial, endothelial, and keratocyte cells in a native collagen matrix. Tissue Eng, 2000, 6(4):307–319.

37. Payne RG, Yaszemski MJ, Yasko AW, and Mikos AG. Development of an injectable, in situ crosslinkable, degradable polymeric carrier for osteogenic cell populations. Part 1. Encapsulation of marrow stromal osteoblasts in surface crosslinked gelatin microparticles. Biomaterials, 2002, 23(22):4359–4371.

38. Awad HA, Wickham MQ, Leddy HA, Gimble JM, and Guilak F. Chondrogenic differentiation of adipose-derived adult stem cells in agarose, alginate, and gelatin scaffolds. Biomaterials, 2004, 25(16):3211–3222.

39. Risbud M, Endres M, Ringe J, Bhonde R, and Sittinger M. Biocompatible hydrogel supports the growth of respiratory epithelial cells: possibilities in tracheal tissue engineering. J Biomed Mater Res, 2001, 56(1):120–127.

40. Altman GH, Diaz F, Jakuba C, Calabro T, Horan RL, Chen J, et al. Silk-based biomaterials. Biomaterials, 2003, 24(3):401–416.

41. Minoura N, Aiba S, Gotoh Y, Tsukada M, and Imai Y. Attachment and growth of cultured fibroblast cells on silk protein matrices. J Biomed Mater Res, 1995, 29(10):1215–1221.

42. Sofia S, McCarthy MB, Gronowicz G, and Kaplan DL. Functionalized silk-based biomaterials for bone formation. J Biomed Mater Res, 2001, 54(1):139–148.

43. Middleton JC and Tipton AJ. Synthetic biodegradable polymers as orthopedic devices. Biomaterials, 2000, 21(23):2335–2346.
44. Thomson RC, Wake MC, Yaszemski MJ, and Mikos AG. Biodegradable polymer scaffolds to regenerate organs. Biopolymers II, 1995,122:245–274.
45. Gunatillake PA and Adhikari R. Biodegradable synthetic polymers for tissue engineering. Eur Cell Mater, 2003,5:1–16; discussion 16.
46. Yang S, Leong KF, Du Z, and Chua CK. The design of scaffolds for use in tissue engineering. Part I. Traditional factors. Tissue Eng, 2001, 7(6):679–689.
47. Freed LE, Marquis JC, Nohria A, Emmanual J, Mikos AG, and Langer R. Neocartilage formation invitro and invivo using cells cultured on synthetic biodegradable polymers. J Biomed Mater Res, 1993, 27(1):11–23.
48. Shinoka T, Ma PX, Shum-Tim D, Breuer CK, Cusick RA, Zund G, et al. Tissue-engineered heart valves. Autologous valve leaflet replacement study in a lamb model. Circulation, 1996, 94(9) Suppl II:164–168.
49. Kaihara S, Kim S, Kim BS, Mooney DJ, Tanaka K, and Vacanti JP. Survival and function of rat hepatocytes cocultured with nonparenchymal cells or sinusoidal endothelial cells on biodegradable polymers under flow conditions. J Pediatr Surg, 2000, 35(9):1287–1290.
50. Mooney DJ, Mazzoni CL, Breuer C, McNamara K, Hern D, Vacanti JP, et al. Stabilized polyglycolic acid fibre-based tubes for tissue engineering. Biomaterials, 1996a, 17(2): 115–124.
51. Agrawal CM and Ray RB. Biodegradable polymeric scaffolds for musculoskeletal tissue engineering. J Biomed Mater Res, 2001, 55(2):141–150.
52. Ishaug-Riley SL, Okun LE, Prado G, Applegate MA, and Ratcliffe A. Human articular chondrocyte adhesion and proliferation on synthetic biodegradable polymer films. Biomaterials, 1999, 20(23–24):2245–56.
53. Park SS, Jin HR, Chi DH, and Taylor RS. Characteristics of tissue-engineered cartilage from human auricular chondrocytes. Biomaterials, 2004, 25(12):2363–2369.
54. Yang F, Murugan R, Ramakrishna S, Wang X, Ma YX, and Wang S. Fabrication of nanostructured porous PLLA scaffold intended for nerve tissue engineering. Biomaterials, 2004, 25(10):1891–1900.
55. El-Amin SF, Lu HH, Khan Y, Burems J, Mitchell J, Tuan RS, et al. Extracellular matrix production by human osteoblasts cultured on biodegradable polymers applicable for tissue engineering. Biomaterials, 2003, 24(7):1213–1221.
56. Mooney DJ, Organ G, Vacanti JP, and Langer R. Design and fabrication of biodegradable polymer devices to engineer tubular tissues. Cell Transplant, 1994, 3(2):203–210.
57. Kim BS, Nikolovski J, Bonadio J, Smiley E, and Mooney DJ. Engineered smooth muscle tissues: regulating cell phenotype with the scaffold. Exp Cell Res, 1999, 251(2): 318–328.
58. Teng YD, Lavik EB, Qu X, Park KI, Ourednik J, Zurakowski D, et al. Functional recovery following traumatic spinal cord injury mediated by a unique polymer scaffold seeded with neural stem cells. Proc Natl Acad Sci U S A, 2002, 99(5):3024–3029.
59. Dai NT, Williamson MR, Khammo N, Adams EF, and Coombes AG. Composite cell support membranes based on collagen and polycaprolactone for tissue engineering of skin. Biomaterials, 2004, 25(18):4263–4271.
60. Barralet JE, Wallace LL, and Strain AJ. Tissue engineering of human biliary epithelial cells on polyglycolic acid/polycaprolactone scaffolds maintains long-term phenotypic stability. Tissue Eng, 2003, 9(5):1037–1045.
61. Park YJ, Lee JY, Chang YS, Jeong JM, Chung JK, Lee MC, et al. Radioisotope carrying polyethylene oxide-polycaprolactone copolymer micelles for targetable bone imaging. Biomaterials, 2002, 23(3):873–879.
62. Ciapetti G, Ambrosio L, Savarino L, Granchi D, Cenni E, Baldini N, et al. Osteoblast growth and function in porous poly epsilon-caprolactone matrices for bone repair: a preliminary study. Biomaterials, 2003, 24(21):3815–3824.

63. Ng KW, Hutmacher DW, Schantz JT, Ng CS, Too HP, Lim TC, et al. Evaluation of ultrathin poly(epsilon-caprolactone) films for tissue-engineered skin. Tissue Eng, 2001, 7(4): 441–455.

64. Fisher JP, Vehof JW, Dean D, van der Waerden JP, Holland TA, Mikos AG, et al. Soft and hard tissue response to photocrosslinked poly(propylene fumarate) scaffolds in a rabbit model. J Biomed Mater Res, 2002, 59(3):547–56.

65. Fisher JP, Holland TA, Dean D, and Mikos AG. Photoinitiated cross-linking of the biodegradable polyester poly(propylene fumarate). Part II. In vitro degradation. Biomacromolecules, 2003, 4(5):1335–1342.

66. Peter SJ, Miller MJ, Yasko AW, Yaszemski MJ, and Mikos AG. Polymer concepts in tissue engineering. J Biomed Mater Res, 1998, 43(4):422–427.

67. Suggs LJ and Mikos AG. Development of poly(propylene fumarate-*co*-ethylene glycol) as an injectable carrier for endothelial cells. Cell Transplant, 1999, 8(4):345–350.

68. Shung AK, Behravesh E, Jo S, and Mikos AG. Crosslinking characteristics of and cell adhesion to an injectable poly(propylene fumarate-*co*-ethylene glycol) hydrogel using a water-soluble crosslinking system. Tissue Eng, 2003, 9(2):243–254.

69. Davis KA and Anseth KS. Controlled release from crosslinked degradable networks. Crit Rev Ther Drug Carrier Syst, 2002, 19(4–5):385–423.

70. Heller J, Barr J, Ng SY, Abdellauoi KS, and Gurny R. Poly(ortho esters): Synthesis, characterization, properties and uses. Adv Drug Deliv Rev, 2002, 54(7):1039–1039.

71. Andriano KP, Tabata Y, Ikada Y, Heller J. In vitro and in vivo comparison of bulk and surface hydrolysis in absorbable polymer scaffolds for tissue engineering. J Biomed Mater Res, 1999, 48(5):602–612.

72. Vacanti JP, Morse MA, Saltzman WM, Domb AJ, Perez-Atayde A, and Langer R. Selective cell transplantation using bioabsorbable artificial polymers as matrices. J Pediatr Surg, 1988, 23(1 Pt 2):3–9.

73. Langer R. Biomaterials in drug delivery and tissue engineering: one laboratory′s experience. Acc Chem Res, 2000, 33(2):94–101.

74. Kumar N, Langer RS, and Domb AJ. Polyanhydrides: an overview. Adv Drug Deliv Rev, 2002, 54(7):889–910.

75. Burkoth AK, Burdick J, and Anseth KS. Surface and bulk modifications to photocrosslinked polyanhydrides to control degradation behavior. J Biomed Mater Res, 2000a, 51(3): 352–359.

76. Burkoth AK and Anseth KS. A review of photocrosslinked polyanhydrides: in situ forming degradable networks. Biomaterials, 2000b, 21(23):2395–2404.

77. Uhrich KE, Gupta A, Thomas TT, Laurencin CT, and Langer R. Synthesis and characterization of degradable poly(anhydride-*co*-imides). Macromolecules, 1995, 28(7): 2184–2193.

78. Muggli DS, Burkoth AK, and Anseth KS. Crosslinked polyanhydrides for use in orthopedic applications: degradation behavior and mechanics. J Biomed Mater Res, 1999, 46(2): 271–278.

79. Qiu LY and Zhu KJ. Novel biodegradable polyphosphazenes containing glycine ethyl ester and benzyl ester of amino acethydroxamic acid as cosubstituents: syntheses, characterization, and degradation properties. J Appl Polym Sci, 2000, 77(13):2987–2995.

80. Laurencin CT, Norman ME, Elgendy HM, el-Amin SF, Allcock HR, Pucher SR, et al. Use of polyphosphazenes for skeletal tissue regeneration. J Biomed Mater Res, 1993, 27(7): 963–973.

81. Laurencin CT, El-Amin SF, Ibim SE, Willoughby DA, Attawia M, Allcock HR, et al. A highly porous 3-dimensional polyphosphazene polymer matrix for skeletal tissue regeneration. J Biomed Mater Res, 1996, 30(2):133–138.

82. Langone F, Lora S, Veronese FM, Caliceti P, Parnigotto PP, Valenti F, et al. Peripheral nerve repair using a poly(organo)phosphazene tubular prosthesis. Biomaterials, 1995, 16(5): 347–353.

83. Tangpasuthadol V, Pendharkar SM, and Kohn J. Hydrolytic degradation of tyrosine-derived polycarbonates, a class of new biomaterials. Part I: Study of model compounds. Biomaterials, 2000, 21(23):2371–2378.

84. Choueka J, Charvet JL, Koval KJ, Alexander H, James KS, Hooper KA, et al. Canine bone response to tyrosine-derived polycarbonates and poly(L-lactic acid). J Biomed Mater Res, 1996, 31(1):35–41.

85. Lee SJ, Choi JS, Park KS, Khang G, Lee YM, and Lee HB. Response of MG63 osteoblast-like cells onto polycarbonate membrane surfaces with different micropore sizes. Biomaterials, 2004a, 25(19):4699–4707.

86. Cai J, Bo S, Cheng R, Jiang L, and Yang Y. Analysis of interfacial phenomena of aqueous solutions of polyethylene oxide and polyethylene glycol flowing in hydrophilic and hydrophobic capillary viscometers. J Colloid Interface Sci, 2004, 276(1):174–181.

87. Gutowska A, Jeong B, Jasionowski M. Injectable gels for tissue engineering. Anat Rec, 2001, 263(4):342–349.

88. Bourke SL and Kohn J. Polymers derived from the amino acid L-tyrosine: polycarbonates, polyarylates and copolymers with poly(ethylene glycol). Adv Drug Deliv Rev, 2003;55(4):447–466.

89. Novikova LN, Novikov LN, and Kellerth JO. Biopolymers and biodegradable smart implants for tissue regeneration after spinal cord injury. Curr Opin Neurol, 2003, 16(6):711–715.

90. Sawhney AS, Pathak CP, and Hubbell JA. Bioerodible hydrogels nased on photopolymerized poly(ethylene glycol)-*co*-poly(alpha-hydroxy acid) diacrylate macromers. Macromolecules, 1993, 26(4):581–587.

91. Sims CD, Butler PE, Casanova R, Lee BT, Randolph MA, Lee WP, et al. Injectable cartilage using polyethylene oxide polymer substrates. Plast Reconstr Surg, 1996, 98(5):843–850.

92. Bryant SJ and Anseth KS. Controlling the spatial distribution of ECM components in degradable PEG hydrogels for tissue engineering cartilage. J Biomed Mater Res, 2003, 64A(1):70–79.

93. Desai NP, Sojomihardjo A, Yao Z, Ron N, and Soon-Shiong P. Interpenetrating polymer networks of alginate and polyethylene glycol for encapsulation of islets of Langerhans. J Microencapsul, 2000, 17(6):677–690.

94. Ganta SR, Piesco NP, Long P, Gassner R, Motta LF, Papworth GD, et al. Vascularization and tissue infiltration of a biodegradable polyurethane matrix. J Biomed Mater Res, 2003, 64A(2):242–248.

95. Gorna K and Gogolewski S. Preparation, degradation, and calcification of biodegradable polyurethane foams for bone graft substitutes. J Biomed Mater Res, 2003, 67A(3):813–827.

96. Zhang J, Doll BA, Beckman EJ, and Hollinger JO. A biodegradable polyurethane-ascorbic acid scaffold for bone tissue engineering. J Biomed Mater Res, 2003, 67A(2):389–400.

97. Grad S, Kupcsik L, Gorna K, Gogolewski S, and Alini M. The use of biodegradable polyurethane scaffolds for cartilage tissue engineering: potential and limitations. Biomaterials, 2003, 24(28):5163–5171.

98. McDevitt TC, Woodhouse KA, Hauschka SD, Murry CE, and Stayton PS. Spatially organized layers of cardiomyocytes on biodegradable polyurethane films for myocardial repair. J Biomed Mater Res, 2003, 66A(3):586–595.

99. Sachlos E and Czernuszka JT. Making tissue engineering scaffolds work. Review: The application of solid freeform fabrication technology to the production of tissue engineering scaffolds. Eur Cell Mater, 2003,5:29–39; discussion 39–40.

100. Mikos AG, Bao Y, Cima LG, Ingber DE, Vacanti JP, and Langer R. Preparation of poly(glycolic acid) bonded fiber structures for cell attachment and transplantation. J Biomed Mater Res, 1993, 27(2):183–189.

101. Mikos AG, Thorsen AJ, Czerwonka LA, Bao Y, Langer R, Winslow DN, et al. Preparation and characterization of poly(L-lactic acid) foams. Polymer, 1994, 35(5):1068–1077.

102. Hua FJ, Kim GE, Lee JD, Son YK, and Lee DS. Macroporous poly(L-lactide) scaffold 1. Preparation of a macroporous scaffold by liquid – liquid phase separation of a PLLA – dioxane – water system. J Biomed Mater Res, 2002, 63(2):161–167.
103. Ma PX and Zhang R. Synthetic nano-scale fibrous extracellular matrix. J Biomed Mater Res, 1999, 46(1):60–72.
104. Nam YS and Park TG. Porous biodegradable polymeric scaffolds prepared by thermally induced phase separation. J Biomed Mater Res, 1999, 47(1):8–17.
105. Hsu YY, Gresser JD, Trantolo DJ, Lyons CM, Gangadharam PR, and Wise DL. Effect of polymer foam morphology and density on kinetics of in vitro controlled release of isoniazid from compressed foam matrices. J Biomed Mater Res, 1997, 35(1):107–116.
106. Whang K, Tsai DC, Nam EK, Aitken M, Sprague SM, Patel PK, et al. Ectopic bone formation via rhBMP-2 delivery from porous bioabsorbable polymer scaffolds. J Biomed Mater Res, 1998, 42(4):491–499.
107. Chen GP, Ushida T, and Tateishi T. Development of biodegradable porous scaffolds for tissue engineering. Mater Sci Eng C-Biomimetic Supramol Syst 2001;17(1–2):63–69.
108. Nam YS, Yoon JJ, and Park TG. A novel fabrication method of macroporous biodegradable polymer scaffolds using gas foaming salt as a porogen additive. J Biomed Mater Res, 2000, 53(1):1–7.
109. Mooney DJ, Baldwin DF, Suh NP, Vacanti JP, and Langer R. Novel approach to fabricate porous sponges of poly(D,L-lactic-co-glycolic acid) without the use of organic solvents. Biomaterials, 1996b, 17(14):1417–1422.
110. Sheridan MH, Shea LD, Peters MC, and Mooney DJ. Bioabsorbable polymer scaffolds for tissue engineering capable of sustained growth factor delivery. J Control Release, 2000, 64(1–3):91–102.
111. Yang S, Leong KF, Du Z, and Chua CK. The design of scaffolds for use in tissue engineering. Part II. Rapid prototyping techniques. Tissue Eng, 2002, 8(1):1–11.
112. Hutmacher DW. Scaffold design and fabrication technologies for engineering tissues – state of the art and future perspectives. J Biomater Sci Polym Ed, 2001, 12(1):107–124.
113. Zein I, Hutmacher DW, Tan KC, and Teoh SH. Fused deposition modeling of novel scaffold architectures for tissue engineering applications. Biomaterials, 2002, 23(4): 1169–1185.
114. Zeltinger J, Landeen LK, Alexander HG, Kidd ID, and Sibanda B. Development and characterization of tissue-engineered aortic valves. Tissue Eng, 2001, 7(1):9–22.
115. Cima LG, Vacanti JP, Vacanti C, Ingber D, Mooney D, and Langer R. Tissue engineering by cell transplantation using degradable polymer substrates. J Biomech Eng, 1991, 113(2): 143–151.
116. Cooke MN, Fisher JP, Dean D, Rimnac C, Mikos AG. Use of stereolithography to manufacture critical-sized 3D biodegradable scaffolds for bone ingrowth. J Biomed Mater Res, 2003, 64B(2):65–69.
117. Lu Y and Chen SC. Micro and nano-fabrication of biodegradable polymers for drug delivery. Adv Drug Deliv Rev, 2004, 56(11):1621–1633.
118. Angelova N and Hunkeler D. Rationalizing the design of polymeric biomaterials. Trends Biotechnol, 1999, 17(10):409–421.
119. Ma PX and Choi JW. Biodegradable polymer scaffolds with well-defined interconnected spherical pore network. Tissue Eng, 2001, 7(1):23–33.
120. Lee SH, Kim BS, Kim SH, Kang SW, and Kim YH. Thermally produced biodegradable scaffolds for cartilage tissue engineering. Macromol Biosci, 2004b, 4(8):802–810.
121. Lu L and Mikos AG. The importance of new processing techniques in tissue engineering. MRS Bull, 1996, 21(11):28–32.
122. Shin H, Jo S, and Mikos AG. Biomimetic materials for tissue engineering. Biomaterials, 2003, 24(24):4353–4364.
123. Holy CE, Fialkov JA, Davies JE, and Shoichet MS. Use of a biomimetic strategy to engineer bone. J Biomed Mater Res A, 2003, 65A(4):447–453.

124. Cunliffe D, Pennadam S, and Alexander C. Synthetic and biological polymers-merging the interface. Eur Polym J, 2004, 40(1):5–25.
125. Shive MS and Anderson JM. Biodegradation and biocompatibility of PLA and PLGA microspheres. Adv Drug Deliv Rev, 1997, 28(1):5–24.
126. Ziats NP, Miller KM, and Anderson JM. In vitro and in vivo interactions of cells with biomaterials. Biomaterials, 1988, 9(1):5–13.
127. Anderson JM and Miller KM. Biomaterial biocompatibility and the macrophage. Biomaterials, 1984, 5(1):5–10.
128. von Burkersroda F, Schedl L, and Gopferich A. Why degradable polymers undergo surface erosion or bulk erosion. Biomaterials, 2002, 23(21):4221–4231.
129. Temenoff JS, Athanasiou KA, LeBaron RG, and Mikos AG. Effect of poly(ethylene glycol) molecular weight on tensile and swelling properties of oligo(poly(ethylene glycol) fumarate) hydrogels for cartilage tissue engineering. J Biomed Mater Res, 2002, 59(3):429–437.

Chapter 16
BioMEMS

Florent Cros

16.1 MEMS General Introduction

In the 1966 motion picture "Fantastic Voyage", a microscopic vessel channels along the arteries of a patient to perform a critical brain surgical operation. The fictitious microscopic submarine is built using a very attractive method of miniaturization: a fully functional macro-scale vessel is "shrunk" to the appropriate dimensions using a contraption which continues to defy laws of physics. Around the same time, breakthrough from the nascent semiconductor industry inspired real scientists and engineers to look for miniaturization techniques and create microscopic objects potentially useful to human-kind. Such objects are now referred to as bioMEMS. BioMEMS are microscopic electro-mechanical systems (MEMS) applied to biomedical applications and life sciences. They are, for example, complete miniature laboratories used by pharmaceutical companies for high throughput screening for drug discovery. They are expected to hasten genomic research. They could provide useful sensory feedback in modern surgical tools and will soon embark on their own fantastic voyage inside the human body.

BioMEMS are historically enabled through discoveries and technological advances in solid-state micro-electronics. Like their solid-state counterpart (integrated circuitry, IC), MEMS and bioMEMS are miniaturized (sub-micrometer to millimeter) systems, often produced using batch fabrication techniques. BioMEMS are however closer to MEMS than pure ICs as their operating principles often relies on moving, rotating or vibrating mechanical and fluidic components that allow physical functions to be performed in conjunction with electrical functions.

The creation and manipulation of miniaturized electromechanical systems has an undeniable scientific and engineering appeal. It offers the possibility to exploit favorably the scaling of laws that govern the operation of such systems and their interaction with the environment, allowing the fabrication of system otherwise useless or cumbersome at larger scale. BioMEMS also offer a venue to couple mechanical, electrical, chemical and fluidic effects, inviting related academic and industrial research to become multi-disciplinary and, for more than three decades, fruitful commercial efforts. As a result, MEMS have become diffused throughout many fields of engineering (RF, optical telecommunications, aerospace, automobile) and first breached the limits of research laboratory in the 1970s, becoming

commercially available as miniaturized low cost pressure sensors or inertial sensors. BioMEMS are slowly reaching commercial maturity as well. Today, MEMS and bioMEMS represent a multibillion dollar market: applications including microfluidic elements already represent 30–40% of the total generated revenue.

This chapter offers a general introduction to bioMEMS. The first part underlines the intrinsic attractiveness and raison d'être of BioMEMS, contrasting with the risks and dangers associated with the use of such miniature systems for biomedical applications. The second part gives a closer look at bioMEMS though the lens of the technologies available to BioMEMS engineers. Common bioMEMS materials and associated micromachining techniques are presented. A third part focuses on a few significant bioMEMS applications.

16.2 BioMEMS General Presentation

16.2.1 What Are They?

16.2.1.1 BioMEMS as Transducers

Like MEMS, BioMEMS can often be classified as transducers. They interact locally with their environment, generally transforming an environmental physical input (such as mechanical stress, pressure, temperature, etc.) into an electrical signal output (frequency, voltage, etc.) and vice versa. The nature and the usefulness of the interaction vary from sensing to actuating.

Whether a system is classified as a sensor or as an actuator, its attractiveness is coupled with the perspective of packing all its components and functions into minimally encumbering packages. It is not hard to envision that such sensing and actuating capabilities can be harnessed to assist and enhance modern health care practice in a variety of ways.

Figure 16.1a offers an artistic three-dimensional view of a bioMEMS used as a sensor: a fully integrated multi-site pressure sensor for wireless arterial flow characterization [1]. The pressure sensitive elements are machined using techniques such as "bulk-micromachining" of silicon and hermetic bonding, notions that will be presented later in the chapter. The complete system is small enough to fit inside a 6 mm diameter stent-like structure and could help in the early detection of restenosis (re-occlusion of the artery) after a coronary stenting procedure.

Figure 16.1b is a scanning electron micrograph of a miniature heart-cell force transducer [2]. This microscopic transducer offers an elegant way for biologists to measure the contractility of isolated ventricular heart cells. The chip includes a cell attachment site at the tip of two microscopic mechanical arms imbedded along with doped silicon strain-gauges and on-chip integrated amplification circuitry. This bioMEMS attempts to transcend a few of the shortcomings of traditional experimental methods, often limited, because of their macroscopic size, in their measurement fidelity or ease in handling a single cell.
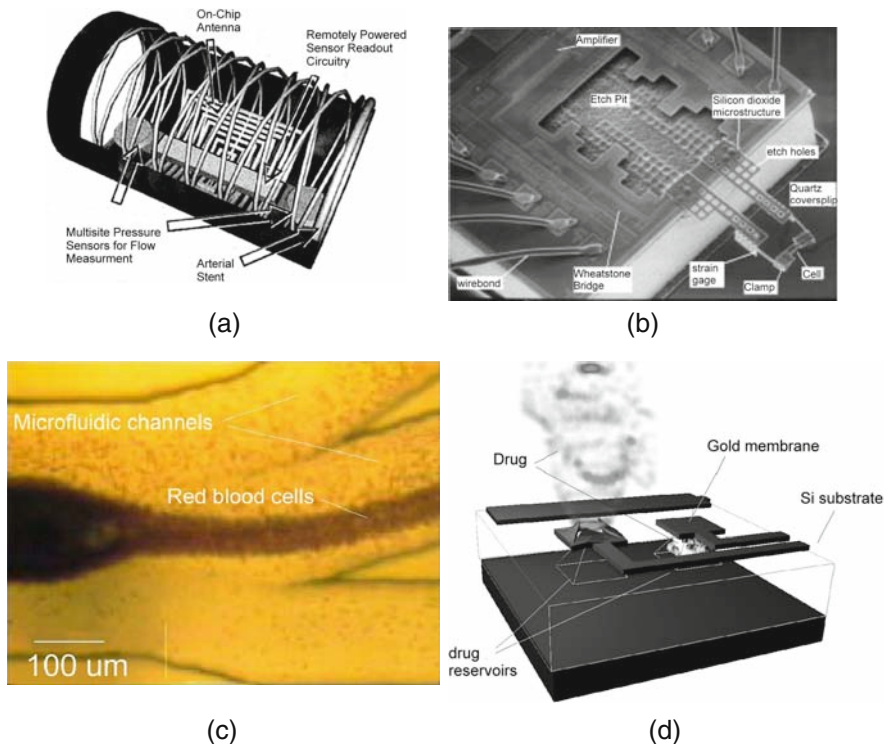
Figure 16.1d is an example of a bioMEMS that can be considered as an actuator, as it is designed to act on its environment as an implantable drug delivery system [3]. This actuator carries loads (tens of nanoliters) of substances like hormones or painkillers stored inside micromachined cavities. Each cavity is sealed using a thin-film gold cap (thin hermetic membrane) electrically connected to an on-board integrated circuitry and a battery (not represented). Applying a voltage to one or more gold caps in the presence of an aqueous environment (e.g., a saline solution) triggers an electrochemical reaction that results in the dissolution of the gold cap, freeing the drug to diffuse out of the cavity.

16.2.1.2 BioMEMS as Building Blocks for Micro Fluidic Elements

BioMEMS do not necessarily have to perform as transducers: they are also used with a class of useful devices that are exempt from performing electrical, mechanical or chemical coupled functions. Coronary stents, arguably represent such a class: stents are scaffold-like metallic elements, small enough to fit inside small

arteries (e.g., arteries that surround the heart). The creation of stents and other useful two- and three-dimensional micro-fluidic structures (e.g., micro-channels, micro-reservoirs) can potentially benefit from materials and fabrication techniques available to BioMEMS engineers, unmatched by conventional techniques. Attempts to duplicate stents [4] and create other complex structures (e.g., micro-needles for painless drug delivery [5]) using micromachining abound in the bioMEMS literature.

16.2.2 Why Building BioMEMS?

The fundamental reasons underlining current and future successes essentially revolved around two intertwined concepts. The first one is that there are many advantages to reduce the size of a system and exploit favorably the scaling of laws and forces that govern its behavior and performances. The second one is more pragmatic in nature: current fabrication technologies do allow the creation of such structures.

16.2.2.1 Favorable Exploitation of Miniaturization

A direct and evident observation is that small systems are highly desirable, especially when considering one of the true pinnacles of future biomedical achievements: systems that can be implanted to perform tasks in-vivo, i.e. inside a human being. Delivery and implantation of in vivo devices may be greatly simplified as miniaturized systems may be delivered using simpler and safer medical interventions such as benign localized surgical procedures, endovascular catheterization or injection through large-gauge hypodermic needles. Once implanted, properly shaped miniaturized devices have the potential to minimize local or systemic reaction in host tissues.

More importantly, the miniaturization of a system is accompanied by the scaling of the fundamental laws governing its behavior and performances. This represents one of the major driving forces analogous to the ones that compel micro-electronic engineers to reduce the size of a transistor: the smaller the transistor, the lesser power is needed to actuate it and the faster it actuates. For electro-mechanical systems, similar improvements are expected: reducing their size affects their ability to be set in motion or to respond rapidly to a given stimulus or an environmental change [7, 8], etc.

The scaling of governing laws and forces can be related in many instances to a change of surface-to-volume ratio. Let's consider a simple object shaped as a parallelepiped. As its typical dimensions (width, length, and thickness) are being scaled down, its volume (cubic function of the typical dimensions) is, in proportion, shrinking faster than the combined surface area (only quadratic function of the typical dimensions) of its faces. Heat related phenomena are greatly influenced as the dimensions of the system are shrinking: in the MEMS scale, cooling, mainly a surface phenomenon, is much more effective than heating, a volume effect.

The advantageous miniaturization of devices such as macroscopic dielectrophoresis separators (DEP) offers a good example [9]. DEPs rely on high electric fields to induce slow migration and separation of genetic materials in a gel. In the MEMS scale, adverse heat effect associated with high electric fields can be handled more efficiently. As a result, miniature DEP systems are able to accommodate higher electrical field and hasten the separation phenomena.

Coupling multi-physics effects is another key feature although not exclusive, but available to BioMEMS. Figure 16.1(c) is a top-view micrograph of a blood cells micro-separator that uses a magnetic field to sort (white blood cells) from paramagnetic (red blood cells) traveling inside a microscopic fluid passage-way [6]. The picture is taken at the junction of three micro-fluidic channels branching off of the main separation channel. The darker region indicates a zone with higher concentration of red blood cells. This system exploits efficiently a characteristic of fluid flow at small scale and couples it with interactions between cells and an external magnetic field. In a microfluidic conduit, fluids tend to flow in a non-turbulent or laminar fashion. Laminar flow is characterized by adjacent layers of fluid moving in parallel at different speeds down the length of the fluidic channel. The lowest velocity is found along the vessel wall. Using an external field such as a magnetic field, it is possible to exert attractive or repulsive forces to either red (paramagnetic) blood cells or white (diamagnetic) blood cells and confine them in various regions of the fluid flow. Different analyte moving at differential speed along a channel results in their physical isolation, purification, and concentration or, in that case, separation.

The latter example helps making a transition to the next section: the feasibility of micro-systems. Field-assisted separators and low Reynolds-number channels (prone to induce laminar flow) can be fabricated using conventional machining techniques. However, using micromachining techniques, a separation channels can be build along with a myriad of other sub-components, "meant to" ease the introduction and the movement of fluids (micro-valves, micro-pumps), enhance the precision of the detection method, minimize the amount of reagents needed and, again, reduce the analysis time.

16.2.2.2 Factual Possibility to Fabricate Such Small Systems

For BioMEMS to serve biomedical functions useful to mankind health, they shall not remain a one-of-a-kind laboratory miniature curiosity. In their introduction to biomedical engineering and materials [10], Enderle et al. notice that the discovery of penicillin is merely as important as the implementation of mass production methods of clinical grade material in quantities large enough to affect the delivery of health care. BioMEMS holds great promises in that respect, as most bioMEMS manufacturing techniques are directly inspired from large-scale production techniques copied, derived or integrated with conventional large-scale low-cost batch fabrication from the microelectronic industry.

Such fabrication techniques and facilities (clean-rooms) offer the possibility to fabricate of tens, hundreds, thousands etc. of interconnected, individually addressable systems with geometrical patterns down to the sub-micrometric range all at

once. When the bouquet of "conventional" fabrication techniques does not suffice, MEMS and bioMEMS engineer often develop additional techniques that then join the general tool-box of so-called micromachining techniques.

In many instances, current microfabrication techniques are mature enough to sustain many of the present needs for a multitude of biomedical and biological applications. Road maps applied to micro-electronics indicate that integrated circuits with sub-components (transistors) displaying minimum features smaller than 50 nm could be mass-produced by 2007. BioMEMS do not necessarily have to comprise sub-components that small. Today, they can fulfil many of the needs of biomedical engineers and biologist. Cellular biology is a good example: most prokaryotic cells have dimensions (diameter) ranging from 1 to 5 μm, eukaryotic cell are generally larger, typically of the order of 10 μm in diameter. Using current fabrication tools, complete functional micro-system can be (and are being) equipped with sub-elements (e.g., microscopic needles, microscopic electrodes) of adequate shape and size well adapted for probing and manipulating cells as well as intracellular materials.

Miniaturized sensors and actuators are already being integrated with circuitry to create "smarter" systems. BioMEMS and their associated wireless or wired I.C. are designed to have frugal energy requirements, accentuating further their appeal for in-vivo applications. Moreover, BioMEMS do not always need on-board sources of energy. Energy coupling schemes based on electromagnetic and ultrasonic phenomena are few of the many efficient ways to energize, actuate and interrogate devices wirelessly. The tool-box of micromachining techniques allows much more: arrayed system of individually addressable detection or reaction sites for fast parallel operations, hermetic packaging techniques allowing tasks to be performed in presence of fluids or gases, etc. This list is not all-inclusive: many other positives points will naturally emerge as materials and micromachining techniques are introduced more in detail later in this chapter.

16.2.3 Risks and Drawback Associated to BioMEMS

In June 2004, Boston Scientific Corp. launched a massive recall of popular drug-coated stents used to treat patients with heart condition, citing at least two patient deaths [11]. Life and death implications of biomedical applications can impact patient's lives as well as justify or threaten the existence of companies who contributed to the commercialization of a given application. For BioMEMS to impact truly on modern healthcare, reliability is of utter importance. MEMS and BioMEMS that comprise mechanical moving, vibrating, rotating parts such as membranes, beams, plates, etc., are already being scrutinized for their susceptibility to wear, fatigue, fracture or being destroyed due to shock, and vibration [12, 13].

The author of [14] exhorts bioMEMS engineers to pay close attention to reliability evaluation, noticing that "reliability is often one of the last areas of consideration for the developers of a new technology." Whether there are intended for in vitro or in vivo application, BioMEMS suffer from various specific ailments.

16.2.3.1 In-Vitro BioMEMS Reliability

A variety of parasitic effects specific to BioMEMS relying on micro-fluidic elements (micro-channels, reaction chambers, etc...) can impede such systems from outputting a correct diagnostic or from performing at all. Scaling laws are not always favorable to miniaturized systems. Fluids forced in microscopic channels and reservoirs flow in a non-turbulent (laminar) fashion. Mixing two fluids becomes more problematic and relies mainly on diffusion. Improper mixing of sample fluids and reagents can prevent the system from outputting the correct response. Clogging and bubble trapping are two other parasitic effects that can yield to critical failure. Purging the system can prove to be either deleterious or simply impractical. Also, because miniaturized in vitro diagnostic systems rely on smaller quantities of reagents, false positive or false negative diagnostic can arise from pipetting errors, carryover (fluid contacting surfaces reused from test to test) and cross-contamination [14].

16.2.3.2 In-Vivo Reliability

The prospect of implanting BioMEMS is accompanied by even more stringent reliability requirements. The device and the function it serves must not cause any short or long term damage for the human host. The microdevice must not be a source of chemical contamination (e.g., out-gassing, corrosion), susceptible to release mechanical debris, etc. Microdevices must survive the implantation procedure (e.g., delivery through an endovascular catheter) and survive the host biological system natural response. Once implanted, micro-devices cannot always be replaced. Consequently, implanted micro-devices must operate according to specification for variable length of time and demonstrate no or an acceptable level of drift.

The importance of reliability is not only restrained to biomedical applications, as MEMS are becoming more diffused throughout other industrial fields of application that involve the safety of the end-user. There is a high degree of confidence that MEMS and bioMEMS devices can meet these requirements, as other MEMS (e.g., inertial sensors) based on the similar materials, manufacturing processes and operating principles have proven to be exceedingly robust in the automotive, military and aerospace industries [15].

16.3 BioMEMS Design, Materials and Fabrication

16.3.1 BioMEMS Design

MEMS and BioMEMS engineering is multidisciplinary in nature. As has been seen in earlier examples, microsystems operation often involves coupled-field phenomena. A valid design effort requires a strong understanding of the underlying multi-physics that rules the operation of the microdevice. Many problems are

simple enough to be modeled first by hand, using simple mechanical, electro-static, magnetic, thermal, and fluidic models, etc. When analytical models cannot be refined further, powerful numerical tools can bring more elaborate answers. Modern coupled-field finite element analysis (FEA) software packages treat most realms of coupled field phenomena applicable to bioMEMS modeling. Accounting for parasitic effects (e.g., residual stress in thin films) can however represent a substantial modeling challenge as such unexpected effects are sometimes larger that the primary design. Moreover, selecting a material with relevant physical properties and identifying the correct fabrication approach can have as much of an impact on the feasibility and the performances of a useful micro-device as refining a design or selecting a mode of operation: there are often more than more way to deposit a given material and, as a result, its properties can vary dramatically.

Considering all aforementioned reasons, modeling and design of MEMS and bioMEMS alone emerges as a subject of great complexity. Fortunately, it is not one of insurmountable difficulties: it is often best tackled by adopting a research and development method that synchronizes conceptualization and design efforts with direct reduction to practice and actual fabrication of prototypes of incremental complexity. Therefore, knowing *how to design* bioMEMS is equally as important as knowing *how to built* them. The reader can elect to discover more of the intricacies of bioMEMS modeling by selecting the relevant dedicated literature [16–18]. The following pages discuss MEMS and bioMEMS from the material and fabrication point of view.

16.3.2 BioMEMS: Importance of Materials and Materials Characterization

On the bioMEMS scale, properties of thin films and micro-machined materials often differ from mechanical properties of their bulk counterpart [19]. Figure 16.2 represents a stress-strain curve for pure nickel and electroplated Ni as found in [20]. The results for electroplated Ni using a generic MEMS fabrication sequence known as LIGA differ from results obtained and reported for other types of pure Ni. Sharpe et al. concludes that "LIGA nickel is stronger that bulk material."

Besides the fact that bulk material properties do not always transfer to material machined or deposited using MEMS technologies, there are other imperative reasons why engineers ought to refine their knowledge about materials and fabrication techniques made available for bioMEMS: repeatability and reliability imperatives to yield thin film with reproducible characteristics (residual stress, density, electrical properties, mechanical properties, etc.), enhancement of particular physical traits through the understanding of the underlying physical mechanisms during deposition, etch, etc., and development of either "new" materials or "new" machining techniques to produce such materials. When no specific information is available, useful data can be extrapolated with adequate level of circumspection from values gathered from bulk materials using macro-scale specimen.
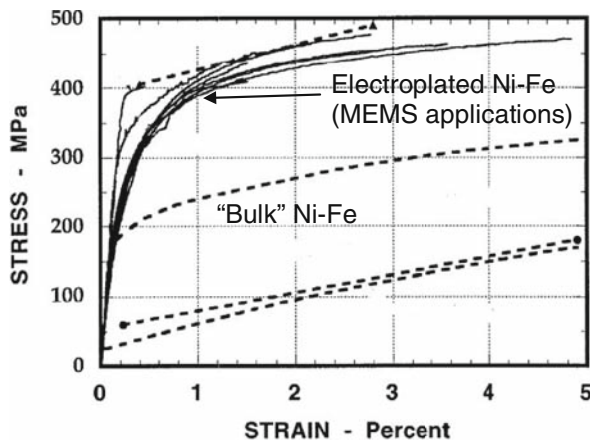
Fig. 16.2 Stress-strain curves for Pure Nickel (© 2005 IEEE) [20] present the results for Ni electroplated using a generic MEMS fabrication sequence known as LIGA (explained in the chapter). Results presented in [20] are in *solid lines*, whereas the *dashed lines* represent other results reported for other types of polycrystalline pure Ni of different grain sizes. Sharpe et al. conclude that "LIGA nickel is stronger that bulk material"

Investigating mechanical properties at relevant MEMS scales is not always an easy task. Specimen are often either two small and two fragile to be handled, gripped or stamped even with the smallest available strain gages: direct methods of extraction such as stress-strain measurements are rarely applicable. Instead, properties can be extracted using indirect methods, often requiring building dedicated MEMS structures when mechanical behavior under local stimuli is recorded and analyzed [18].

The authors of [21] offer one example, among many ways, to investigate material properties on the MEMS scale. The authors use a set of MEMS test structures to characterize 0.7 μm thick, 34 μm long and 1–5 μm wide polysilicon suspended beams. Mechanical forces are locally applied to the microscopic specimen using a well known electrical-to-mechanical force transduction scheme: an electrostatic comb-drive micro-actuator. An electrostatic comb-drive actuator is composed of two electrically conductive electrodes facing one another. The recurring pattern is characteristic of two overlapping comb-like structures: one electrically conductive finger that belongs to the first comb-like electrode is flanked with two similar fingers that belong to the vis-à-vis electrode. An air gap separates (and electrically insulates) the first finger from the two adjacent ones. Local electrostatic forces appear when an electrical potential difference is created (applied) between opposite fingers.

In that case, the global attractive force between two complete sets of fingers (i.e. a comb-like electrode) is the sum of individual forces acting on each set of opposite fingers. In [21], one electrode is suspended above the substrate. Its only mechanical (and electrical) anchoring point is assumed by the small polysilicon beam to be tested. When a voltage difference is applied, forces acting on the electrodes

directly force the specimen to bend in a manner that is proportional to its geometry, the square of the potential difference between the two electrodes and, more importantly, the Young's modulus of its composing material. By monitoring the voltage applied to the electrostatic comb-drive electrodes and by measuring the physical displacement on one electrode compare to the other, the authors of [21] could extract both Young's modulus and, when fracturing was achieved, fracture strength of polysilicon.

Specialized literature [18] lists intrinsic properties of some materials applied to MEMS. The reader can also refer to [18] for a complete review of extraction methods available to MEMS and bioMEMS material scientists.

16.3.3 Material for BioMEMS

16.3.3.1 Silicon

Si is arguably the single best documented material for large scale industrial and research applications in microelectronics related fields.

Its appeal is first attributed to its performances as a semiconductor, its ability to grow a high quality oxide layer (SiO_2) under relevant processing condition, as well as its availability in large volumes of ultra-pure mono-crystalline defect free quality. Single crystal silicon (SCS) is grown from highly purified melt using various methods (Czochralski, Float Zone) producing boules (pear-shaped masses with the atomic structure of a single crystal) subsequently sectioned and polished to create wafers with known crystallographic orientation and virtually devoid of structural defects, point defects or disruptive forms of dislocation, and exhibiting low amounts of chemical impurities (C, O and heavy metals). As a beneficial result, SCS exhibit outstanding mechanical properties as well as a natural predisposition to be physically machined with high precision in reproducible controlled fashion. These two particular traits give SCS an undeniable appeal for MEMS and bioMEMS.

Because SCS is virtually devoid of structural defects, occurrence of fatigue induced cracks and dramatic mechanical failure is greatly minimized. As long as geometrical and/or surface defect occasioned during processing can be minimized [22], SCS is especially suited for the fabrication of sensitive mechanical structures such as high frequency MEMS resonators.

SCS micromachining techniques, anisotropic and isotropic etching, allow for the creation of in-plane grooves and vertical through-holes, essential to the fabrication of micro-fluidic networks of in-plane micro-channels, micro-reservoirs, etc.

Silicon can be porosified very effectively using a variety of micromachining techniques. Many aspects of its final porosity are controllable: spore size (from micrometer, to nanometer to angstroms), density, geometrical orientation, depth, etc. Changes of physical properties (electrical conductivity, electroluminescence or refractive index) upon exposure to fluid or gaseous substances can be monitored and harnessed to a variety of chemical and biological sensing schemes [23]. Porous

silicon is also investigated for its bioactivity, its ability to bond to bones or provide scaffold-like networks for tissue engineering [24, 25].

Si allows direct monolithic integration (on the same die) with relevant CMOS circuitry (e.g., signal conditioning circuitry). Doped Si also exhibits piezoelectric properties that have been exploited since the early 1980s for the creation of commercial pressure sensitive sensors applied to biomedical applications (e.g., General Electric/NovaSensor pressure sensors).

When Si is not grown from a melt, it can be deposited using chemical vapor deposition (CVD) methods. Depending on the condition (pressure, temperature, deposition rate, nature of the substrate), CVD Si can be either mono-crystalline (epitaxial growth) or polycrystalline films. For example, polysilicon is used in the commercially available Analog Devices acceleration sensor [15].

16.3.3.2 Metals

A variety of physical vapor deposition (PVD) systems is available for depositing thin film metals relevant for MEMS fabrication. Systems like electron beam evaporators or plasma assisted sputtering systems yield film thickness ranging typically from a few tens of Angstroms to fractions of a micrometer in thickness. Single metal films or multilayered films can be obtained. The nature of the film, the topology of the substrate as well as the nature of the underlying layers or the requirements of the subsequent processing steps dictate the method of deposition.

Thin metal films enter the bioMEMS fabrication flow at various crucial moments. Ti or Cr are commonly employed to promote adhesion between underlying layers with the new layers or structures being subsequently built atop. Cu, Au and Al are commonly used as electrical conductors. When adequately patterned, such films can be used to produce conducting lines, electrodes, etc. Thin films can be used as a conductive base for metal electro-deposition (Cu), for their magnetic properties (e.g., Ni-Fe), as a metallic diffusion barrier (Ta, Ti, Ti-W), for interconnecting or bonding properties (Au), as a hard mask for subsequent patterning of underlying material, etc. Their electrochemical properties are exploited: Pt, Ag/Ag-Cl, Ir oxide thin films are used as working, reference and counter electrodes in glucose monitoring sensors [26]. Their hydrophilic or hydrophobic properties can also be used for biology applications. Au is often used to provide a site for antibodies or single DNA strands to be immobilized. The authors of [27] use two dimensional arrays of Cr/Au pads to immobilize and study drosophila embryos.

Plating with an adherent coating using electrolysis (also referred to as electroplating or electro-deposition) is a technique complementary to thin film deposition as it often requires first the presence of a PVD thin film metal as a base metallic layer. The PVD thin film is often referred to as the "seed layer" and provides at the surface of the substrate where the deposit is to be grown the necessary electric path to allow the electrochemical reactions to be carried out. The seed layer also provides a surface where the metal to be grown can nucleate, adhere and grow. One major appeal resides in the range of achievable deposit thickness yielding the fabrication

of electroplated structures protruding tens to hundreds of micrometers above the surface of the supporting substrate.

Cu, Au, Ni, Ni-Fe and platinum metals (Pt, Rh, Pd etc.) are a few of the numerous materials commonly plated in a very large number of applications for their electrical, magnetic or mechanical properties [28]. Plated metals may be accompanied with various levels of in-plane and out-of-plane stress. Stress can be controlled through varying the deposition rate (current density), the temperature, the agitation, the chemistry of the bath, etc. The size of the plating cell can also influence the reproducibility of the deposit characteristics.

Electroless deposition offers a useful alternative. Electroless plating does not necessarily require the presence of an underlying seed layer: this growth technique is especially useful when the complete fabrication flow cannot accommodate the presence of a continuous electrical path at the surface of the substrate where the metal is to be grown. In that fashion, adherent metal coating can grow on non-conducting surfaces such as polymers. It tends to produce conformal coatings, uniform uninterrupted films on convex and concave surface directly in contact with the electrolyte. Electroless plating was used for example for the fabrication of complex and useful three-dimensional structures such as micro-needles [5], otherwise nearly impossible to build.

Ti is often used in conventional biomedical application (e.g., for hip replacements) as it fulfils basic functions such as biocompatibility, bioactivity (Ti and Ti alloys have the ability to osseo-integrate), resistance to environmental degradation (corrosion) and/or high resistance to abrasion. Ti brazing is a very useful "conventional" encapsulation technology for pacemaker leads. Unfortunately, Ti is somehow missing from the arsenal of MEMS materials with the exception of PVD thin films: no truly MEMS-friendly machining technique allows us to deposit and accurately pattern large amounts (tens to hundred of micrometers) of Ti. The authors of [29] are attempting to do so, using electrophoresis to deposit thick layers of titanium oxide for MEMS application.

16.3.3.3 Ceramics for MEMS Microfabrication

Many of the common ceramics used in conventional biomedical applications are also available for MEMS and BioMEMS. Conventional approaches such as green body pre-shaping, sintering and final machining (e.g., for heart valve replacement) are not always suitable when considering the fabrication of miniaturized objects. As a result, most metal oxides (alumina, sapphire) and silica were made available in films, tapes or wafer shapes and are being used extensively for many MEMS, BioMEMS applications.

Opaque (optically not transparent) "green tapes" (unfired ceramic tapes) are very useful for the creation of multilayered monolithic package that include essential sub-elements such as micro-channels, passive and active electronics components [30]. They are available in various thicknesses and can be machined using laser ablative techniques prior to final assembly.

Optically transparent ceramic wafers (borosilicate glass AF45, fused silica, Pyrex) are particularly attractive for in vitro applications. They can yield self packaged hermetic devices when wafer-to-wafer or anodic bonding methods are applicable. Transparent glass wafers packaging offer a natural optical window for optical detection to be performed.

Ceramics can also be deposited in thin films (typically less than few micrometers) using CVD, PVD, sol-gel solutions or atomic layer deposition (ADL) methods. PZT (lead zinc titania) and ZnO (zinc oxide) hold a special place for they exhibit strong piezoelectric properties, useful in a myriad of sensing and actuation schemes [31, 32]. Silicon carbide, used in conventional medical application to produce wear resistant coatings was made available to provide similar services for MEMS structures [33].

16.3.3.4 Polymers

Polymers are low cost and relatively simple materials to handle and process. An extended family of polymers, photosensitive or not, either spin-coated, spray coated, electroplated, deposited using CVD systems are available for MEMS micromachining. The majority of polymers are available in liquid form. Some are available in dry films for dry lamination [34]. Photosensitive polymers (photoresist, or PR) represent the single most useful class of polymers as they are a necessary element for photolithography. SU-8 photosensitive epoxy or PMMA (poly(methyl methacrylate)) can be processed to create high aspect ratio complex three dimensional structures (e.g., vertical filters with micro-range opening [35]).

Polymers like parylene [1] or silicone (for implantable applications) can be used as passivation layer or to encapsulate a micro-device. Polymer like poly(dimethylsiloxane) (PDMS) can be used in embossing and molding techniques for inexpensive fabrication and replication of useful passive micro-fluidic elements (channels).

Once patterned, polymer mechanical properties are reliable enough to be used for the creation of mechanical sub-elements such as a deflecting membrane for a micro-pumps or micro-valves. Changes in their properties (shape, volume, etc.) depending on environmental changes can be also harnessed to useful actuating and sensing functions. For example, some hydrogels swell as the pH of a surrounding solution is changed. The authors of [36] demonstrate the possibility to pattern hydrogels on top of compliant micro-cantilevers: as the pH of the solution in which the cantilever beam is submerged changes, so does the volume of the hydrogel, forcing the cantilever beam to bend. Using the proper transduction scheme (e.g., piezo-resistive effects), the deflection of the beam can be measured and proportionally related to a change of pH.

Other polymers are used for their biodegradable properties: [37] uses PCL (poly-caprolactone), PGA (polyglycolide), PLA (poly lactide-*co*-glycolide) and PLLA (poly-L-lactide) to fabricate biodegradable microneedles for percutaneous drug release systems.

Polymers can be used for their semiconducting properties to create biosensors which can sense specific environmental chemical species as they are being absorbed in the polymer [38].

16.3.3.5 Biomaterials and Nanomaterials

DNA, biomolecules, proteins and even living cells can be harnessed to perform useful tasks for bioMEMS and microelectronics applications. Their use and applications are discussed at the end of the section dedicated to bioMEMS fabrication techniques.

16.3.4 Biocompatibility of MEMS Materials

Biocompatibility evaluation of MEMS materials of construction for implantable medical devices is crucial to the development of useful devices, especially when commercialization is at stake.

Kotzar et al. [39] notice that "data on the biocompatibility and sterilizability of MEMS materials, however, are surprisingly limited and data from currently accepted standard tests are almost non-existent". This paper is among the first to report systematic results of MEMS materials biocompatibility and response to sterilization. The methodology uses an internationally recognized test matrix (ISO 10993) to evaluation a small sub-set of useful bioMEMS materials, including Si, SiO3, Si3N4, SiC, polysilicon, thin film Ti and SU-8. Samples are subjected to aqueous and non-aqueous extraction and cyto-toxicity tests. They are examined (using scanning electron microscopy) to detect possible deleterious effects occurring during sterilization procedures (autoclave or gamma irradiation). Samples are implanted in rabbits and tissue response is observed on excised muscle tissues after 1–12 weeks, using conventional microscopy to look at sign of inflammation, presence of gross necrosis, fibrosis, etc. The conclusions show that, when processed using typical MEMS techniques, no indications of problems were uncovered. Their tests, however, are partially complete as (1) they consist of only the cytotoxicity and short term implantation tests and (2) material biocompatibility can be potentially affected by a complete sequence of processing steps needed for the fabrication of a complete functional bioMEMS, including exposure to solvents, dry and wet etching, etc. that could, for example, affect the surface chemistry and solicit different tissue response. Surface modification, either deliberate or occasioned during the various steps of a complete fabrication sequence (wet chemical etching, crosslinking, cleaning, etc.), can have considerable effects on the biocompatibility or the bioactivity of materials.

16.3.5 BioMEMS Fabrication Techniques

A complete fabrication process involves some form of pattern transfer technique used in conjunction with sequences of deposition, annealing, curing, removal,

etching, etc. of materials. Such operations are usually performed at the surface of a medium, offering most of the time a flat polished clean surface to work on (e.g., Si wafer, fused silica wafer, printed wire board, etc.) and are repeated several times to create, layer after layer a structure with the desired geometrical, electrical and mechanical traits.

The following sections describe fabrication tools and techniques used for BioMEMS micromachining: such techniques are not limited to BioMEMS as they are common to many MEMS applications as well.

16.3.5.1 Photolithography

Photolithography is an essential processing step that enters the fabrication of most bioMEMS.

Electron-beam writing, screen-printing or photolithography are examples of techniques developed to transfer a pattern (a geometric shape) from a mask (generally re-usable) to the surface of a medium. Photolithography using ultra violet light source (UV with a wavelength ranging from 193 to 465 nm) in particular is responsible for today's MEMS and IC industrial successes, allowing the transfer of tens, hundreds, thousands, etc. of geometrical patterns down to the sub-micrometric range, all at once.

UV photolithography images patterns in a layer of photosensitive polymer. Photosensitive polymers (often called photo-resists or PR) are usually available in liquid form (such polymers are also available in dry films). Photolithography is mainly a binary pattern transfer: after developing, regions of the PR layer are either completely washed away from the surface of the substrate while others remain until they are not needed any more.

A complete photolithography sequence represented in Fig. 16.3 proceeds as follow: (1) the surface of the substrate is prepared (e.g., cleaning, use of chemical promoting adhesion of subsequent layers deposited atop), (2) a PR coating is applied (e.g., spin casting), and pre-baked (also called soft bake), (3) the patterns on the mask are aligned to any pre-exiting patterns on the substrate, (4) the PR is exposed to UV, (5) the PR is developed, and hard baked. Useful processing is performed (6), using the PR as a masking film. The PR is eventually stripped and the substrate is cleaned (7).

The film thickness, the topography and nature of the substrate, the underlying chemistry of the PR, the position of the mask relative to the substrate, the wavelength of the light source, etc., each have an influence over the in-plane and out-of-plane achievable spatial resolution. UV photolithography offers strict limits in term of achievable in plane minimum feature size (often referred to as critical dimension, or CR), namely around 100 nm. Such CR is achieved in IC applications using ultra-thin PR layers. BioMEMS applications do not systematically ask for such performances. Instead, bioMEMS engineers have access to techniques and materials that enable the creation of structures that protrude tens, hundreds, thousands of micrometers above the surface of the substrate. This is often motivated by the fact that
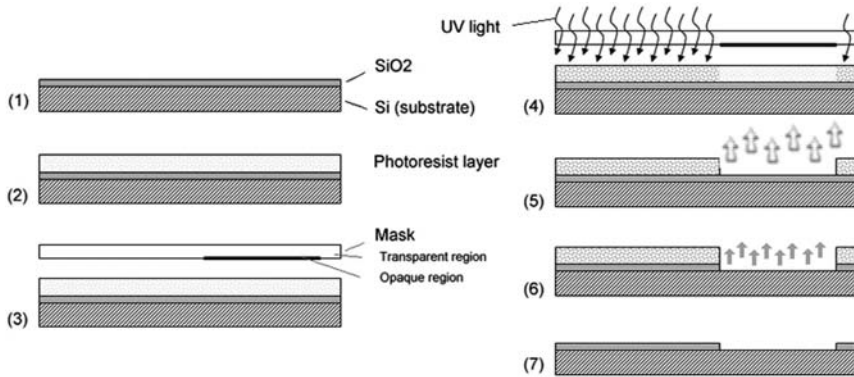
Fig. 16.3 Complete photolithography sequence inspired from the fabrication of a complete C-MOS transistor (first sequence preparing the fabrication of n-wells); (1) the surface (SiO2) of the substrate is prepared for (2) a PR coating to be applied and pre-baked. (3) Patterns on a photolithographic reusable mask are aligned to the substrate, (4) the PR is exposed to UV light that is not blocked by opaque regions on the photolithographic mask, (5) the PR is developed, and hard baked. (6) The SiO2 layer is selectively etched using a solution containing hydrofluoric acid (e.g. "BOE", i.e. Buffered Oxide Etch). (7) The PR is eventually stripped and the substrate is prepared for the next processing steps (cleaned, etc.)

tall extruded structures helps in optimizing the use of on-chip real-estate as well as taking full advantage of certain effects of scaling laws.

Several UV PR, such sensitive photosensitive epoxy SU-8™ or the family of Futurrex™ NR resists, allow the fabrication of such structures. Key features are the total height, the vertically of the side walls (e.g., near vertical using SU-8) and the geometrical aspect ratios of the structures, often expressed as the ratio of the width of a particular feature (e.g., a free standing wall, a trench, etc.) over its height.

Photolithography processes do not always involve ultra-violet light sources. X-ray photolithography uses the X-ray radiation from a synchrotron. PMMA is then a choice polymer. It can be spun and patterned in layers much thicker than UV-sensitive photoresist as X-ray can penetrate thicker layers with less diffraction, yielding structures with very high aspect ratio.

16.3.5.2 Bulk Micro-Machining

Bulk micromachining represents a family of micro-fabrication techniques that use the substrate to etch out structures useful to MEMS and bioMEMS applications. The substrate provides both a mechanical base onto which the complete process is to be performed as well as the primary building material eventually composing the MEMS structures. Etching methods involve either wet or dry processes. Si is a choice material for bulk micro-machining. Wet and dry bulk micromachining is not limited to Si. However, Si bulk micromachining is extremely well characterized and controlled and will be introduced first.

Wet Bulk Micromachining

Wet etching of SCS using an aqueous solution of KOH is a very simple and reliable technique to create structures with predictable geometry. Etching based on KOH is considered to be anisotropic, as the profile of the etched cavity follows certain crystallographic planes. Therefore, the resulting trench side wall verticality, aspect ratio and total depth are functions of the crystallographic orientation of the SCS wafer, the concentration and temperature of the KOH solution. Grooves etched in (100) silicon wafers have 54.9° angles whereas grooves etched in (110) silicon wafer have vertical side wall. Grooves resulting from KOH etching can be tailored to have the shape of channels, reservoirs or needle like structures, as seen Fig.16.4b. A longer etch can yield through-holes (through the thickness of the wafer) openings. When combined with the appropriate mask and complete process, KOH etching can yield the fabrication of free standing structures only attached to the rest of the substrate by small beams.

KOH etching requires the presence of a "hard" mask at the surface of the SCS wafer. The material performing as a hard mask must fulfil at least two requirements. In the presence of a KOH solution, the masking material must not degrade (etch) as fast as Si does. This choice material must also accommodate other forms of etching
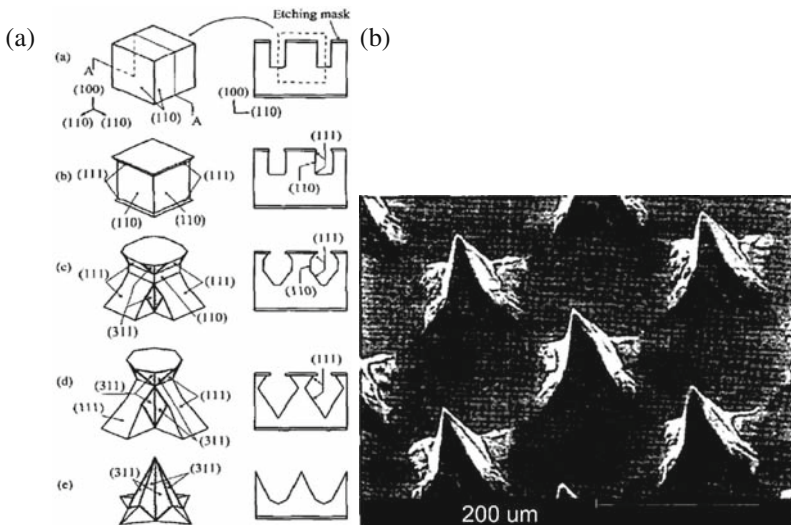
Fig. 16.4 Fabrication of arrayed 3-D microstructures by chemical anisotropic etching as described in [40] (© 2005 IEEE). The pyramidal formations can be explained with a single model of the undercutting phenomena during the anisotropic wet etching. The etching proceeds as follow. Grooves with (110) planes side-walls are cut into a *single crystal* Si wafer using a high precision diamond saw. The anisotropic KOH etch begins, (110) planes etch at fast rate, until (111) planes start to expand. As the etch progresses, (311) planes appear, at the intersection of two (111), (311) planes have a superior local etching rate, (311) plane size increases and finally connects to other peripheral plans of similar crystallographic orientation. Etching is stopped when the final sought-after structures emerge

compatible with the use of photolithography photoresist. High quality silicon nitride grown in a low pressure chemical vapor deposition (LPCVD) chamber is a good choice. An example of a KOH etch sequence can proceed as follow: (1) a <100> SCS wafer is prepared (cleaned) and coated (front and back) with 5000 Å of Si2N3, (2) the front side of the wafer is coated with a PR layer that is subsequently patterned using photolithography, (3) the Si2N3 layer is etched through the openings created in the PR layer, (4) the PR is stripped and the wafer is placed in a KOH solution, and (5) the wafer is removed from the KOH solution when the trench has reached the desired depth. More complex etching sequences rely on differential etch rates between various crystallographic orientation of SCS wafers. The authors of [38] offer a step-by-step description of a single KOH etching sequence, as it progressively etches along various dense planes of a Si wafers and results in the creation of needle-like structures (Fig. 16.4a).

Many additional fabrication techniques pair up with KOH to enhance the controllability of the etch: etch stop barrier using ion-implanted boron doped region of SCS wafers, laser assisted etch, etc. TMAH (tetramethyl ammonium hydroxide-water) is an attractive alternative to KOH.

Isotropic etching techniques for Si exist as well: such techniques can involve acid solutions (e.g., HNO3-HF), require different sets of hard masks (e.g., SiO2 thermal oxide) and result in etch profiles that have rounded semi-circular concave edge profiles. Silicon-dioxide based substrates such as fused silica, Pyrex or borosilicate wafer, are very useful for bioMEMS applications, and have associated techniques (etchant: HF, hard mask: metallic thin film) that also allow the creation of cavities, grooves, channels, etc.

Dry Bulk Micromachining

Dry bulk micromachining techniques usually require more complex apparatus, involving "dry" ablative, reactive or sputtering processes (e.g., laser ablation) to remove material at specific locations on the substrate. Plasma assisted dry etching is one method that has reached a very high level of maturity when used to machine silicon in particular. The directionality (anisotropy), etch rate, selectivity, maximum achievable depth, range of achievable trench aspect ratio, etc. are all related to the etching condition (pressure, power, gas) encountered inside the plasma etching system. Particular methods referred to as deep reactive ion etching (DRIE) involving high density plasma allow the etching of extremely vertical trenches and walls in Si. Figure 16.5 is a scanning electron micrograph of a micro-fluidic valve built using a combination of anisotropic and isotropic dry etching techniques [41]. The fluidic channels are etched in the top layer of a SOI (silicon on insulator) wafer.

Plasma assisted etching also requires the presence of a hard mask. Thick PR or thermal SiO2 can be used as hard masks.

Plasma enhanced dry etching techniques are also available for quartz, fused silica and polymers. Laser ablation and cutting is yet another attractive dry bulk micromachining technique for polymeric, ceramic and metallic substrates. Laser ablation using either UV (Yag), CO_2 or infra-red lasers is a serial processes as one laser beam
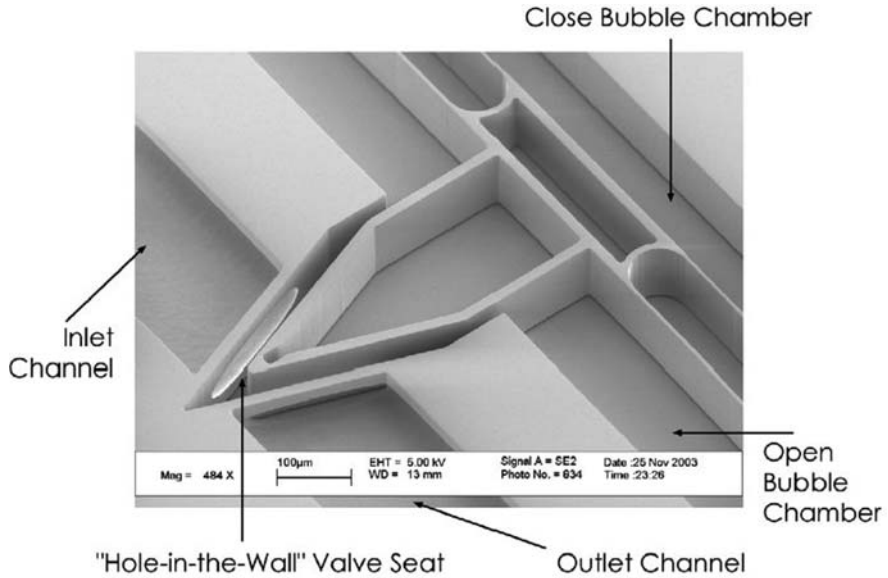
Fig. 16.5 "Hole-in-the-wall" micro gate valve, as depicted in [41] (© 2005 IEEE). The fluidic channels are etched in the top layer of an SOI (silicon on insulator) wafer. The complete fabrication involves both directional as well as isotropic dry etching techniques

cuts one pattern at the time before stepping to the next one. However, powerful well-collimated lasers assisted with computer controlled high precision X-Y-Z tables can accommodate high throughput and high accuracy requirements.

16.3.5.3 Surface Micro-Machining

Surface micromachining represent an ensemble of techniques that use the substrate mainly as a mechanical support to build useful structures after successive deposition and etching of thin film materials.

Thin Films for Surface Micromachining

A wide range of materials, deposition and etching techniques are available, as reviewed in the chapter dedicated to materials. Only relatively thin layers of materials can be processed using thin film evaporation or sputtering method. Thicker deposits can be achieved using micro-molding and electroplating for example.

Wet and Dry Etching

Once a film is deposited at the surface of the wafer, it can subsequently be patterned using a variety of wet and dry etching methods. Adapted hard masks are required. For example, metals and dielectrics can be wet and dry etched directly

through photo resists masks or using another metallic mask. Polymers are routinely dry etched in O_2 plasma using metal (Al, Cu) as hard mask. When looking for the most appropriate wet or dry etching method, selectivity is a key feature to consider. A generic methods known as "sacrificial layer" technique relies on selective etching to create intricate and though-after structures, such as suspended beams or hollow channels, otherwise impossible to fabricate. A sacrificial layer is a structural layer that remains buried throughout the fabrication process under various layers of material until it can be removed, releasing the structures and layers built atop.

Sacrificial layers abide by at least two rules: (1) they must remain unaffected (no cracking, peeling, wrinkling, etc.) during the entirety of the process until it must be removed and (2) selective and efficient removal techniques must exist to remove them without adverse consequences to any remaining structures.

An example of a process involving a photoresist (PR) sacrificial layer as seen in [42] is depicted in Fig. 16.6a. The process depicted here starts as the substrate has already been subjected to previous micromachining steps (not shown). Cavities are etched from the back side of the wafer, resulting in the regions of the wafer where only a thin membrane of silicon is remaining (Fig. 16.6a(1)). Side-by-side metallic structures are fabricated using a fabrication sequence (not shown) involving a seed-layer, the creation of a PR mold, the electroplating of Ni inside the mold and the dissolution of both PR and seed layer (this sequence will be referred to as micromolding and electroplating). The result is shown Fig. 16.6a(2). The wafer is then coated with a thick layer of a PR that is patterned using photolithography
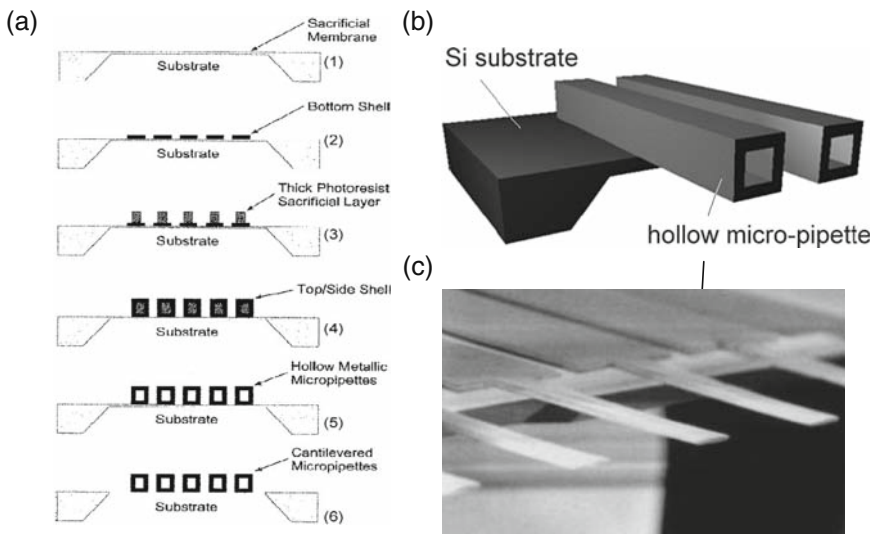
Fig. 16.6 Array on hollow micropipettes fabrication process as found in [42] (© 2005 IEEE). This process relies on a sacrificial layer of thick photoresist and a sacrificial membrane to create hollow cantilever-like structures used in transferring, dispensing or manipulating precise volumes of samples, reagents and buffers

(Fig. 16.6a(3)). The remaining PR patterns, aligned atop the flat metallic structures, will serve as sacrificial structures. The sacrificial PR patterns are hard baked and another complete micromolding and electroplating sequence is performed on top (not shown). This sequence yields the creation of an electroplated Ni structure around the PR sacrificial patterns. The result is shown in Fig. 16.6a(4). The wafer is then soaked overnight in acetone to dissolve completely the PR sacrificial layer (Fig. 16.6a(5)). The process is finally complete when, after a final etching step, the Si membrane is selectively etched under the hollow nickel structures. Figure 16.6b is an artistic representation of the resulting suspended hollow micro-pipettes. Figure 16.6c is a scanning electron micrograph of the completed structures.

More than one sacrificial layer can be implemented during the same complete fabrication process, allowing the creation of complex multilayered structures such as multilayered networks of micro-fluidic channels [43]. Sacrificial layers are generally either metals or PR. Other materials such as Al and Cu are often used as a sacrificial layer in presence of polysilicon, electroplated Ni or Au structures.

16.3.5.4 Embosssing, Injection Molding Techniques

Injection molding and hot embossing are techniques that yield the replication of three dimensional structures first created using either surface or bulk micromachining. These techniques are very appealing as they can yield low cost duplication of a large variety of in-plane micro-fluidic systems.

Injection molding uses polymer in a liquid form. The polymer can be a thermoplastic heated above its glass transition temperature. Once the polymer has cooled down or cross-linked, the structure is demolded. This technique allow the replication of high aspect ratio structures such as thin wall or trenches (e.g., few micrometers wide, few tens of micrometers tall).

Hot embossing uses a layer of thermoplastic polymer deposited at the surface of a supporting substrate. The polymer is brought to the proper temperature so that it can be imprinted. The master key carrying the structures to be replicated is then pressed into the polymer layer, forcing the later to deform. Amorphous and semicrystalline polymers are used in conjunction with hot embossing.

16.3.5.5 Stereophotolithography

Stereo-photolithography uses a computer controlled UV light exposure system to solidify a liquid resin and directly create complex three dimensional structures without the use of any intermediate masking, deposition/etching or conventional photolithography step. Structures are first designed using CAD tools and subsequently built slice by slice in a liquid resin tank. Centimeter size structures with sub-millimeter in-plane and out-of-plane critical dimensions (>5–$10\,\mu m$) can be fabricated.

Stereo-photolithography allows the fabrication of complex structures otherwise cumbersome and/or impossible to fabricate using either bulk or surface micromachining. Structures built using stereophotolithography accommodate reentrant

concave surfaces and semi-enclosed cavities, as long as a path remains for any unprocessed liquid resin to be drained.

Stereo-photolithography is very attractive for inexpensive low resolution microfluidic structures.

16.3.5.6 Bonding, Hermetic Sealing and Packaging

Bonding techniques are crucial to the fabrication and the operation of hermetic bioMEMS as they often offer a simple way to fabricating fully enclosed fluidic networks.

Direct bonding (fusion bonding) is routinely used to bond Si to Si. This process creates a temporary bond between two highly polished ultra clean surfaces brought in contact. The bond is subsequently annealed at elevated temperature (e.g., for Si-Si bond 1200) to reach maximum strength. Wafers including corresponding patterns can be aligned before bonding with very high accuracy (less than few microns). More than two wafers can be stacked, allowing the fabrication of extremely complex structures. Anodic bonding offers a low temperature (400 °C) alternative, requiring instead the presence of a high electric field and a mechanical force to press and bond two wafers together. This technique is well adapted to bonding glass to silicon wafers. The bonding mechanisms require that impurities are present in the glass (such as Pyrex or borosilicate) and that both coefficients of thermal expansion match. Ceramic to ceramic bonding relies on green tape materials that can be bonded, sintered and co-fired together under constant mechanical pressure. Each layer can be effectively patterned individually (using, for example, laser ablation) and eventually stacked together [30]. Bonding can also be applied to metals (e.g., Au), in compression bonding and eutectic bonding.

16.3.5.7 From Top-Down to Bottom-Up

All aforementioned micromachining techniques are sometime referred to as "top-down" manufacturing techniques. The underlying methodology is to fabricate successive generations of ever-smaller systems, from macroscopic down to nanoscopic scale. In contrast, "bottom-up" methods represent an ensemble of techniques to form devices by shaping parts or the entirety of the system atom by atom or molecule by molecule [44].

For example, the affinity of single strands of DNA molecules to assemble with their complementary strands can be exploited to synthesize complex scaffolds. Single strands of DNA attached to micro-particles like gold particles or polystyrene beads will self-assemble under relevant conditions. The linkage between the DNA stands is weak and can be broken by changes in pH or shear forces. Stronger bond strength can be achieved using other affine bio-molecules such as biotin and strep-tavidin [45]. This assembly approach can yield the fabrication of "new" composite materials with useful physical properties. Complex structures could for example be used to trap cells.

Living organism are also being used to synthesize useful materials for micro-electronics and MEMS applications. Fundamental mechanisms of silica production in living organisms like marine sponges or red abalone are being investigated [46].

Entire cells can also be harnessed for useful motoring tasks. Cells of *Escherichia coli* are propelled by their flagella, four to ten long slender filaments that project from random sites on the cell's surface. Despite their appearance and name, flagella do not lash; they rotate. The rotary engine is driven by the cell's proton circulation, which can rotate the flagellum either clock wise or counter-clockwise. The authors of [47] envision the fabrication of a microfluidic pump that is controlled and driven by such biological cellular motors. This pump actuation scheme requires the tethering of harmless strain of *Escherichia coli* cells within a microfluidic channel built using conventional MEMS fabrication techniques. No actual pumping was demonstrated at the time. However, the paper stands at the one potentially fruitful crossing between "top-down" and "bottom-up" techniques.

16.4 BioMEMS Application Review

16.4.1 BioMEMS Classification

BioMEMS applications can be classified in a variety of ways. Various bioMEMS review articles [48–52] list the most famous applications at different points of the last few decades. The authors of [9] categorize bioMEMS as tools for (1) molecular biology and biochemistry (e.g., DNA amplification and separation for genome sequencing [affimetrix]), (2) cell biology (e.g., single cell force sensor [1]), (3) medicine (e.g., pressure sensors [4], drug delivery systems [2], micro-surgical tools), or (4) biosensors[1]. By doing so, the authors of [48] unambiguously demonstrate that bioMEMS are filling key-functions that assist biologist and doctors at all accessible length scales: from molecules to cells, to complex organisms to human beings.

The author of [14] reviews bioMEMS for medical use from the reliability point of view. He classifies them according to three general realms of application. Each realm of application is paired with specific reliability and safety considerations: (1) bioMEMS used in research laboratory that are not intended to perform critical clinical diagnostics (e.g., high through-put screening for drug discovery, genome sequencing), (2) bioMEMS used for clinical diagnostics, which performance could directly impact the way a patient is being treated, and (3) bioMEMS intended for in vivo applications, stigmatized with even higher levels of reliability constraints. This

---

[1]Biosensors are deliberately marginalized. Biosensors usually perform electrochemical-to-electrical conversion tasks for applications such as oxygen, glucose or pH monitoring of fluid samples. The authors of [48] justify this choice by stating that biosensors are "not necessarily microfabricated", i.e., the specifics of their fabrication and operation is often closer to conventional micro-electronics applications than to MEMS. Readers can refer to [54] for a complete description of biosensors.

review gives an interesting guide-line for engineers getting involved with bioMEMS and their eventual commercialization.

Instead of attempting to compile an exhaustive thesaurus of current applications, the author invites the reader to refer to previous sections of this chapter to find other examples, as well as to become acquainted with the aforementioned references. Four important BioMEMS systems that have not been formally discussed yet will how-ever be introduced: (1) BioMEMS for cell culturing, (2) bioMEMS for DNA and protein detection, (3) bioMEMS as prosthesis for interface with the nervous system and (4) bioMEMS for modern surgical tools.

16.4.2 BioMEMS for Cell Culturing

As mentioned in earlier sections of this chapter, bioMEMS help biologist manip-ulate cells or embryos [27]. Micromachining techniques can also be used to create structures for cells to grow and to facilitate their interfacing with electronic circuitry.

Qing et al. [53] presents a two-dimensional array of micro-machined structures designed to trap and provide an adequate site to grow and study neurons. The "neuro-cages" are built in parylene. Each structure has a cylindrical hollow body surrounded at its base by six low-profile tunnel-like structures that allow each neu-ron loaded into the cylindrical compartment to grow neurites and form a neural net-work with adjacent trapped neurons. Figure 16.7a is a scanning electron microscope
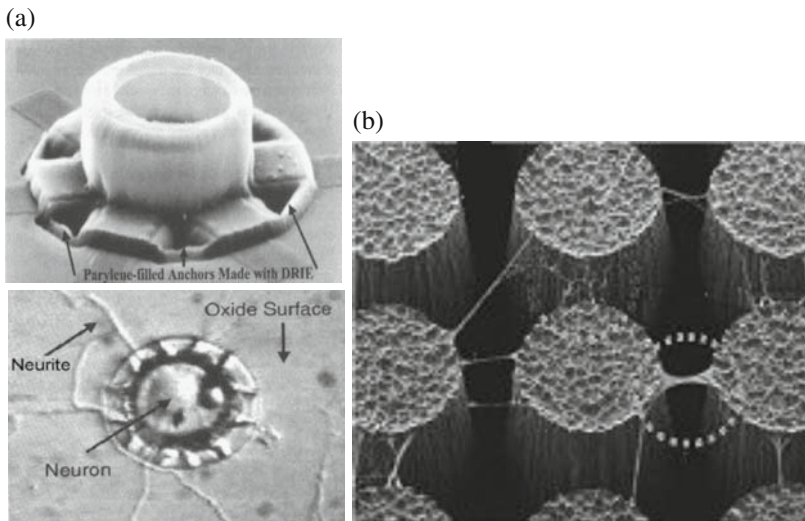
Fig. 16.7a,b a *Top*: scanning electron micrograph of a Parylene-base three dimensional "neuro-cage"; *bottom*: demonstration of neurite outgrowth through the channels (tunnels) of the cage (© 2005 IEEE). **b** SEM pictures of 3-D neuronal cell cultures on SU-8 towers [56] (© 2005 IEEE). Cell bridging and web-like connections of 3-D neuronal networks show the possibility of 3-D cell culturing with these scaffolds

of one "neuron-cage". Figure 16.7b shows a top-view picture of neurite outgrowth through the channels (tunnels) while the neuron cell is being trapped inside the cage. Such structures are intended to impact the study of live neural networks. An unrelated publication [55] demonstrates the fabrication and the use of a CMOS-based monolithic array of micro-electrodes suitable for stimulating and recording the activity of neural networks. Using neural-cages in conjunction with the CMOS-chip design efforts reported in [55] could greatly simplify the immobilization of cells at a specific location (above one electrode) and yield repeated measurements on single neurons.

More recent efforts focus on creating complex three-dimensional structures that mimic the architecture and the functionality of natural micro-vasculatures, i.e., microscopic networks of fluidic channels and electric stimulators capable of sustaining the culture of cells, not in single layers anymore but in multilayered volumes. An attempt to create such structures is presented in [56]. The authors demonstrate the fabrication of a 3-D array of towers using a $300\,\mu$m thick SU-8 photosensitive epoxy layer. Post fabrication, the array is sterilized and soaked in poly-D-lysine, a common protein that promotes cell adhesion. Figure 16.7c is a scanning electron micrograph of the arrays of towers, showing successful neuronal cell growth between two towers after cell delivery and culturing for several days.

16.4.3 BioMEMS for DNA, Proteins and Chemical Analysis

BioMEMS can be applied to fast, multiplexed clinical diagnostics. Such systems can analyze body fluids and samples (e.g., blood, biopsy samples, etc.) and monitor analyte levels indicative of infection, diseases, etc.

The presence of marker proteins (e.g., antigens) in a sample fluid can be detected using the deflection of a microscopic cantilever beam [57]. Figure 16.8b is a scanning electron micrograph of two silicon nitride micro-cantilevers used for multiplexed bio-molecular analysis as shown in [58]. Specificity is attained by first immobilizing various antibodies at the surface of a sensing area, i.e., the upper surface of a cantilever beam (Fig. 16.8a). Upon introduction of a fluid sample, antigens (target protein) present in the fluid sample bind to the specific probing cite (antibody) previously attached to the beam. Antigen to antibody binding occurs only if the specific form of antibody is present at the surface of the beam. The binding induces a rise in the state of mechanical stress at the surface of a cantilever beam, resulting in macroscopic bending of the latter. Minute bending (displacements) can be detected using a variety of detection schemes (optical, capacitive, piezoelectric, etc.). In [58] the bending of the cantilever beam is detected by bouncing a laser beam on the tip of the cantilever and recording the movement of the reflected beam using a CCD image screen (Fig. 16.8c).

For multiplexed analysis, a complete system can, for example, be formed of tens of beams, each one pre-coated with a specific antibody, that may or may not bend, depending on whether specific antigens are present in the fluid sample at the same time.
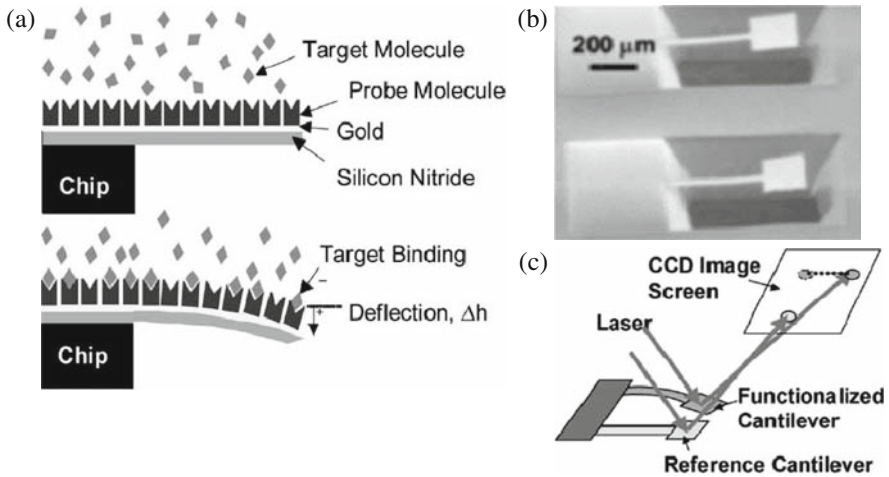
Fig. 16.8 Silicon nitride microcantilevers for multiplexed bio-molecular analysis [58] (© 2005 IEEE). Bio-molecular interaction between target and probe molecules alter intermolecular interactions within a self assembled monolayer on one side of a cantilever beam. The bending of the cantilever beam can be detected by bouncing a laser beam off of the tip of the cantilever beam and recording the movement of the reflected beam using a CCD image screen [58]

The analyte is not always a protein. It can be a clinical chemistry marker such as glucose or oxygen present in a blood sample. The analyte can be in liquid or gaseous form. An example of a system able to perform such analysis is a miniature gas chromatograph. Miniature gas chromatographs (MGC) are portable rapid respiratory and environmental gas analysis systems. As a mixture of gas travels along a micromachined separation column, interaction with the column walls gives a rise to differing effective mobility in the individual gases present in the gas sample. Desirable column length can exceed few meters whereas a typical cross-section is only a few micrometers square. Such requirements take full advantages of both bioMEMS scaling laws and micromachining capabilities. Miniaturized MGCs as envisioned in [59] include not only a suitable separation column, but also on board microfluidic elements such as valves and micro-filtering and chemical sensing systems essential to the introduction and the complete analysis of gases (see Fig. 16.9).

In the case of an MGC, an inert carrier gas is generally used to assist the movement of the gas sample along the column. When handling fluids, other techniques are at hand to move fluids from reservoirs to mixing, reaction (e.g., acoustic, magnetic, d.c. and a.c. electrokinetic phenomena). The authors of [60] present a lab-on-chip that utilizes centrifugal forces acting on fluids confined in passage ways built on a compact disk. Valving, mixing, and detection are demonstrated and have already been yielding commercial realization.

Fig. 16.9 Conceptual
diagram of a MEMS gas
chromatograph that could be
used for the determination of
environmental gas mixtures
[59] (© 2005 IEEE)
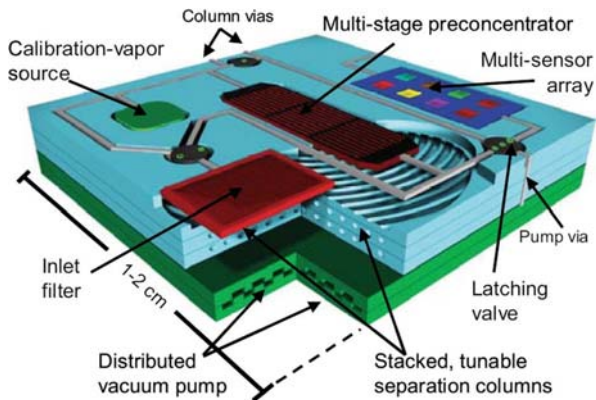
16.4.4 BioMEMS for In-Vivo Applications: Interfacing with the Nervous System

BioMEMS are expected to play an important role in the future development of implanted medical devices, namely stents, pacemakers, drug delivery systems and implantable sensors. We now look at the possibility of using bioMEMS as implantable prostheses to interface with the nervous system.

The ability to monitor and/or stimulate selected groups of cells from the nervous system using the proper electronic-to-tissue interface has been proven to be an effective way to palliate or find a cure to the handicaps of individuals with neurological disorders or disabilities. Work on implantable prostheses for patient suffering of deafness began in the 1960 s. Rehabilitative implants such as visual implants are under scrutiny, relying on either stimulation of the visual cortex or, when adapted, direct stimulation the retina [61]. Miniaturized wired or wireless functional neuromuscular stimulators represent another class of promising implanted prosthesis, expected to help patients with severed or partially injured spinal cords to elicit controllable muscle contractions and to provide sensory feedback. In most case, a suitable electronic-to-tissue interface is needed. A modern electronic-to-tissue interface can include dense arrays of electrodes, on board control electronics and, if needed, wireless links [62]. Current IC technology, coupled with MEMS fabrication and hermetic encapsulation techniques, has already made such interfaces feasible [63, 64], mainly for research purposes, offering irrefutable advantages in the making of probe shanks in particular. In contrast to traditional microelectrode construction methods (e.g., metal electro-erosion), MEMS micro fabrications are well adapted to produce dense three-dimensional arrays of either rigid or flexible electrodes, each one of them carrying more than one stimulating/recording site, able to address a very large number of cells within a block of tissue, and yet small enough to limit cell or microvasculature destruction or tissue displacement upon insertion.
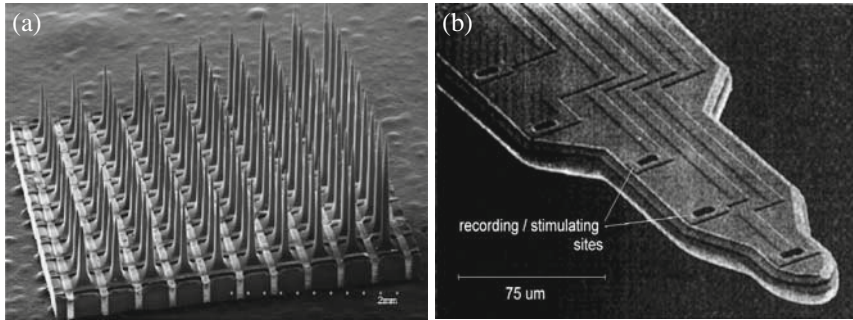
Fig. 16.10a,b a Scanning electron micrograph of the "Utah slanted electrode array" as presented in [64] (© 2005 IEEE). The 10 × 10 array of Si micromachined needles are chronically placed in different fascicles of the sciatic nerve of cats. **b** Scanning electron micrograph of the tip of an in-plane multi-channel microprobe [49] (© 2005 IEEE)

Figure 16.10 presents two types of electrodes. Figure 16.10a is a scanning electron micrograph of the "Utah slanted electrode array" as presented in [64]. The electrodes are built through the thickness of a silicon wafer using bulk micromachining. Each electrode represents a single stimulating-recording site. Figure 16.10b is an example of multi-channel recording microprobe built in plane and realized using bulk micromachining techniques. In that case, one microprobe supports a multitude of recording/stimulating sites.

Several problems related to long term (chronic) implantation of microprobes in peripheral nerve systems are under investigation [64]: breakage of the probe tips during insertion, relative motion between nerves and surrounding muscle tissues, dislodging the implant, unless adapted containment systems are developed, degeneracy of axonal fibers after implantation, stimulation threshold overtime variation, poor ratio of desired signal over noise levels (e.g., noise from EMG activity), etc.

16.4.5 Micro-Surgical Tools

MEMS are particularly suited to improve modern surgical instruments, whether the considered tools are hand tools, teleoperated tools or autonomous surgical robots [65]. As miniaturized sensors, MEMS can provide extra real-time sensory feedback to the surgeon. As miniaturized actuators, MEMS can assist the surgeon in controlling displacements and operations of the end effectors.

Sensors manufactured using MEMS techniques have been proposed to address the need for tactile sensory feedback for surgeons performing minimally invasive surgery (MIS). MEMS sensors strategically embedded inside the end effectors (clippers, scissors, graspers, cautherizers, etc.) can help restore a certain degree of tactile feedback by sensing minute levels of strain and stress at the surface of the tool. The surgeon can gain knowledge of the forces he/she exerts as he/she is cutting, pulling, suturing, etc. MEMS sensors can also be used to help the surgeon identify

the nature of tissues before taking action. Tissue specific density, temperature, etc. can be detected using piezoelectric transducers or miniaturized temperature probes.

As actuators, MEMS can enhance the control of the movements of end effectors or assist in the steering of endoscopes, probes and catheters [28].

Autonomous robots able to propel and act inside the body without the need for tethers or guide wires also benefit from progress in miniaturization techniques, IC integration and hermetic packaging. A fully integrated capsule like NORIKA3 (RF-Systemlab, JAPAN) is an electronic telemetry capsule that can be swallowed for studying the gastro-intestinal track of a patient. The capsule is 9 mm in diameter and 23 mm in length. It incorporates a CCD camera, an LED illumination system and drug delivery modules. Unlike previous capsules, the NORIKA3 is able to move in the gastrointestinal tract by exploiting forces generated by the interaction between an applied external electromagnetic field and local magnetic fields generated by on-board electromagnets (coils).

16.5 Conclusion

BioMEMS rely on a set of material and micromachining techniques that were historically borrowed from other fields of microelectronics, including MEMS. BioMEMS did not long remain in the shadow of their solid-states micro-electronics relatives. To respond to specific needs from biologists or medical doctors, BioMEMS engineers embarked on a long quest for specialized materials, micro-machining tools, adapted design and analysis software, and alternative packaging solutions. Such efforts are receiving inputs from almost every field of modern science. Among prominent *enabling* sciences, material science continues to generate a fertile framework for engineers and scientists to focus their attention on *traditional* materials (e.g., single crystal silicon) as well as *new* materials (e.g., porous silicon, polymers, ceramics, composite materials etc.) that display properties or employ modes of synthesis, deposition, etc. precious to biomedical applications.

As a result, BioMEMS have already made a favorable impact in many areas of modern biology and medicine. This technology offers both relevant conceptual alternatives for conventional macroscopic devices (e.g., di-electrophoresis separators) as well as innovative design spaces for new application (and markets), especially in the fields of medical disposable or implantable sensors.

Many challenges are still ahead, especially when considering long sought implantable miniaturized systems. These challenges are not exclusive to bioMEMS and will be faced best by re-enforcing interactions between all engineering fields common to bioMEMS and conventional biomedical technologies.

References

1. DeHennis A and Wise KD. A fully integrated multi-site pressure sensor for wireless arterial flow characterization. Solid State Sensor, Actuator and Microsystem Workshop, Hilton Head Island, South Caroline, USA, June 2004, pp. 168–171.

2. Gisela L, Palmer RE, Pister KSJ, and Roos KP. Miniature heart cell force transducer system implemented in MEMS technology. IEEE Trans Biomed Eng, 2001, 48(9): 996–1006, September.

3. Scheidt RA, Santini JT Jr, Richards AC, Johnson AM, Rosenberg A, Cima MJ, and Langer R. Microchips as implantable drug delivery devices.. 1st Annual International, Conference on Microtechnologies in Medicine and Biology, 2000, October 12–14, 2000, pp. 483–486.

4. Takahata K and Gianchandani YB. Coronary artery stents microfabricated from planar metal foil: design, fabrication, and mechanical testing. IEEE The 16th Annual International Conference on Micro Electro Mechanical Systems, 2003. MEMS-03 Kyoto, January 19–23, 2003, pp. 462–465.

5. Davis SP, Prausnitz MR, and Allen MG. Fabrication and characterization of laser micromachined hollow microneedles. 12th International Conference on TRANSDUCERS, Solid-State Sensors, Actuators and Microsystems, Vol. 2, June 8–12, 2003, pp. 1435–1438.

6. Han K and Frazier AB. Paramagnetic capture mode magnetophoretic microseparator for blood cells. Solid State Sensor, Actuator and Microsystem Workshop, Hilton Head Island, South Caroline, USA, June 2004, pp. 103–104.

7. Trimmer WSN. Microrobots and micromechanical systems. Sens Actuators, 1989, 19(3): 267–287 (September).

8. Madou MJ. Fundamentals of Microfabrication: The Science of Miniaturization, 2nd edn, CRC Press: Boca Raton.

9. Voldman J, Gray ML, and Schmidt MA. Microfabrication in biology and medicine. Annu Rev Biomed Eng, 1999, 1, 401–425.

10. Enderle JD and Blanchard SM. Introduction to biomedical engineering. Series in Biomedical Engineering, Academic Press, St. Louis, Missouri, 1999, p. 581.

11. http://www.bostonscientific.com

12. Srikar VT and Senturia SD. The reliability of microelectromechanical systems (MEMS) in shock environments., J Microelectromech Syst, 2002, 11(3):pp. 206–214, June.

13. Tanner DM. Reliability of surface micromachined microelectromechanical actuators. Proceedings of the 22nd International Conference on Microelectronics, Vol. 1, May 14–17, 2000, pp. 97–104.

14. Pan JY. Reliability considerations for the BioMEMS designer. Proceedings of the IEEE, Vol. 92 , No. 1, January 2004, pp. 174–184.

15. Judy MW. Evolution of integrated inertial MEMS technology. Solid State Sensor, Actuator and Microsystem Workshop, Hilton Head Island, South Caroline, USA, June 2004, pp. 27–32.

16. Pelesko JA and Bernstein DH. Modeling MEMS and NEMS, CRC Press: Boca Raton, November 26, 2002.

17. Senturia SD. Microsystem design, Kluwer Academic Publisher: Norwell, MA, 2001.

18. Gad-El-Hak M. The MEMS Handbook, CRC Press: Boca Raton, September 27, 2001.

19. Chasiotis I. Mechanics of thin films and microdevices. IEEE Trans Device Mater Reliab, 2004, 4(2), pp. 176–188, June.

20. Sharpe WN Jr, LaVan DA, and Edwards RL. Mechanical properties of LIGA-deposited nickel for MEMS transducers. International Conference on Solid State Sensors and Actuators, 1997. TRANSDUCERS ′97 Chicago, 1997, Vol.: 1, June 16–19, 1997, pp. 607–610.

21. Cacchione F, DeMasi B, Corigliano A, and Ferrera M. Rupture tests on polysilicon films through on-chip electrostatic actuation. Proceedings of the 5th International Conference on Thermal and Mechanical Simulation and Experiments in Microelectronics and Microsystems, 2004. EuroSimE 2004, pp. 347–350.

22. Pourkamali S, Hashimura A, Abdolvand R, Ho GK, Erbil A, and Ayazi F. High-Q single crystal silicon HARPSS capacitive beam resonators with self-aligned sub-100-nm transduction gaps., J Microelectromech Syst, 2003, 12(4):, 487–496, August .

23. De Stefano L, Moretti L, Lamberti A, Longo O, Rocchia M, Rossi AM, Arcari P, and Rendina I. Optical sensors for vapors, liquids, and biological molecules based on porous silicon technology. IEEE Trans Nanotechnol, 2004, 3(1), pp. 49–54, March.

24. Angelescu A, Kleps I, Mihaela M, Simion M, Neghina T, Petrescu S, Moldovan N, Paduraru C, and Raducanu A. Porous silicon matrix for applications in biology. Rev Adv Mater Sci, 2003, 5(5):pp. 440–449.

25. Canham L. Porous silicon as a therapeutic biomaterial. (Sensors & Electron. Div., DERA, Malvern, Worc., UK) Proceedings of the 1st Annual International IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine and Biology, (Cat. No.00EX451), 2000, pp. 109–112.

26. Yang H, Kang SK, Shin D-H, Kim H, and Kim YT. Microfabricated iridium oxide reference electrode for continuous glucose monitoring sensor. 12th International Conference on TRANSDUCERS, Solid-State Sensors, Actuators and Microsystems, , 2003, Vol. No. 1, June 8–12, 2003, pp. 103–106.

27. Bernstein RW, Zhang X, Zappe S, Fish M, Scott M, and Solgaard, O. Characterization of Drosophila embryos immobilized by fluidic microassembly. 12th International Conference on TRANSDUCERS, Solid-State Sensors, Actuators and Microsystems, 2003, Vol. 2, June 8–12, June 2003, pp. 987–990.

28. Haga Y, Esashi M, and Maeda S. Bending, torsional and extending active catheter assembled using electroplating. The 13th Annual International Conference on Micro Electro Mechanical Systems MEMS, January 23–27, 2000, pp. 181–186.

29. Marquordt C and Allen MG. Fabrication of Micromechanical structures of titania and titanium with electrophoretic deposition. 11th International Conference TRANSDUCERS, Solid-State Sensors, Actuators and Microsystems,, 2001, June 8–12, 2003.

30. Fonseca MA, English JM, von Arx M, and Allen MG. Wireless micromachined ceramic pressure sensor for high-temperature applications., J Microelectromech Syst, 2002, 11(4): 337–343, August.

31. Xu F, Wolf RA, Yoshimura T, and Trolier-McKinstry S. Piezoelectric films for MEMS applications. Proceedings of the 11th International Symposium on Electrets, 2002. ISE 11, October 1–3, pp. 386–396.

32. Atkinson GM, Pearson RE, Ounaies Z, Park C, Harrison JS, and Midkiff JA. Piezoelectric polyimide tactile sensors. Proceedings of the 15th Biennial University/Government/Industry Microelectronics Symposium, 2003, June 30—July 2, 2003, pp. 308–311.

33. Rajan N, Mehregany M, Zorman CA, Stefanescu S, and Kicher TP. Fabrication and testing of micromachined silicon carbide and nickel fuel atomizers for gas turbine engines. J Microelectromech Syst, 1999, 8(3): pp. 251–257, September.

34. Conklin JA, Crain MM, Pai RS, Martin M, Pitts K, Roussel TJ, Jackson DJ, Baldwin RP, Keynton RS, Naber JF, and Walsh KM. Alternative fabrication methods for capillary electrophoretic device manufacturing. Proceedings of the Fourteenth Biennial University/Government/Industry Microelectronics Symposium, June 17–20, 2001, pp. 83–85.

35. Yoon Y-K, Park J-H, Cros F, and Allen MG. Integrated vertical screen microfilter system using inclined SU-8 structures. IEEE The 16th Annual International Conference on Micro Electro Mechanical Systems, 2003. MEMS-03 Kyoto, January 19–23, 2003, pp. 227–230.

36. Hilt JZ, Gupta AK, Bashir R, and Peppas NA. A bioMEMS sensor platform based on a cantilever with a precisely patterned environmentally sensitive hydrogel. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society, EMBS/BMES Conference, 2002, Proceedings of the Second Joint Engineering in Medicine and Biology, Vol.: 2, October 23–26, 2002, pp. 1650–1651.

37. Park J-H, Davis S, Yoon Y-K, Prausnitz MR, and Allen MG. Micromachined biodegradable microstructures.. IEEE The 16th Annual International Conference on Micro Electro Mechanical Systems, MEMS-03 Kyoto, January 19–23, 2003, pp. 371–374.

38. Janata J. Electrochemical microsensors. Proceedings of the IEEE, 2003, 91(6): pp. 864–869, June.

39. Kotzar G, Freas M, Abel P, Fleischman A, Roy S, Zorman C, Moran JM, and Melzak J. Evaluation of MEMS materials of construction for implantable medical devices. Biomaterials, 2002, 23:pp. 2737–2750.

40. Shikida M, Odagaki M, Todoroki N, Ando M, Ishihara Y, Ando T, and Sato K. Non-photolithographic pattern transfer for fabricating arrayed 3D microstructures by chemical anisotropic etching. IEEE The 16th Annual International Conference on Micro Electro Mechanical Systems, 2003. MEMS-03 Kyoto, January 19–23, 2003, pp. 562–565.

41. Frank JA and Pisano AP. A low power. Low leakage, bi-stable planar electrolysis micro gate valve. Solid State Sensor, Actuator and Microsystem Workshop, Hilton Head Island, South Caroline, USA, June 2004, pp. 156–159.

42. Papautsky I., Brazzle J, Swerdlow H, Weiss R, and Frazier AB. Micromachined pipette arrays. IEEE Trans Biomed Eng, 2000, 47(6), pp. 812–819, June.

43. Xie J, Shih J, and Tai Y-C. Integrated surface-micromachined mass flow controller. IEEE The 16th Annual International Conference on Micro Electro Mechanical Systems, 2003. MEMS-03 Kyoto, January 19–23, 2003, pp. 20–23.

44. Kawai T. Nanotechnology for the future MEMS. IEEE The 16th Annual International Conference on Micro Electro Mechanical Systems, 2003. MEMS-03 Kyoto, January 19–23, 2003, pp. 1–3.

45. Castelino K, Satyanarayana S, and Sitti M. Manufacturing of two and three-dimensional micro/nanostructures by integrating optical tweezers with chemical assembly. . 3rd IEEE Conference on Nanotechnology, 2003. IEEE-NANO, Vol. 1, August 12–14, 2003, pp. 56–59.

46. Morse DE. Beyond silicon biotechnology: new routes to structure-directed nanofabrication. ASM Conferences Bio-, Micro-, and Nanosystems, July 7–10, 2003, pp. 16.

47. Tung S, Kim J-W, Malshe A, Lee CC, and Pooran R. A cellular motor driven microfluidic system. 12th International Conference on, TRANSDUCERS, Solid-State Sensors, Actuators and Microsystems, Vol. 1, June 8–12, 2003, pp. 678–681.

48. Petersen KE. Silicon as a mechanical material. Proceedings of the IEEE, Vol. 70, No. 5, May 1982, pp. 420–457.

49. Najafi K. Recent progress in micromachining technology and application in implantable biomedical systems. Proceedings of the 16th International Symposium on Micro Machine and Human Science, 1995. MHS ′95, October 4–6, 1995, pp. 11–20.

50. Grayson ACR, Shawgo RS, Johnson AM, Flynn NT, Yawen L, Cima MJ, and Langer R. A BioMEMS review: MEMS technology for physiologically integrated devices. Proceedings of the IEEE, Vol. 92 , No. 1, January 2004, pp. 6–21.

51. Polla DL. BioMEMS applications in medicine. Proceedings of 2001 International Symposium on Micromechatronics and Human Science, 2001. MHS 2001, September 9–12, 2001, pp. 13–15.

52. Kovacs GT. Micromachined Transducers Sourcebook, McGraw-Hill Science/Engineering/Math: Boston, MA, 1st edn, February 1, 1998.

53. He Q, Meng E, Tai Y-C, Rutherglen CM, Erickson J, and Pine J. Parylene neuro-cages for live neural networks study. 12th International Conference on, TRANSDUCERS, Solid-State Sensors, Actuators and Microsystems, Vol. 2, June 8–12, 2003, pp. 995–998.

54. Janata J. Principles of Chemical Sensors. Plenum Press: New York, 1989.

55. Franks W, Heer F, McKay I, Taschini S, Sunier R, Hagleitner C, Hierlemann A, and Baltes H. CMOS monolithic microelectrode array for stimulation and recording of natural neural networks. 12th International Conference on TRANSDUCERS, Solid-State Sensors, Actuators and Microsystems, , Vol. 2, June 8–12, 2003, pp. 963–966.

56. Choi Y, Choi S, Powers R, and Allen MG. Three dimensional tower structures with integrated cross-connects for 3-D culturing of neurons. Solid-State Sensor, Actuator and Microsystem Workshop, Hilton Head Island, South Caroline, USA, June 2000, pp. 286–289.

57. Kumar S, Bajpai RP, and Bharadwaj LM. Lab-on-a-chip based on BioMEMS. Proceedings of International Conference on Intelligent Sensing and Information Processing, 2004 pp. 222–226.

58. Yue M, Lin H, Dedrick DE, Satyanarayana S, Majumdar A, Bedekar AS, Jenkins JW, and Sundaram S. A 2-D microcantilever array for multiplexed biomolecular analysis. J Microelectromechanical Systems, 2004, 13(2): pp. 290–299.

59. Wise KD, Najafi K, Sacks RD, and Zellers ET. A wireless integrated microsystem for environmental monitoring. IEEE International Solid-State Circuits Conference, 2004. Digest of Technical Papers. ISSCC, 2004, Vol. 1, February 15–19, 2004, pp. 434–437.

60. Zoval JV and Madou M. Centrifuge-based fluidic platforms. Proceedings of the IEEE, Vol. 92, No. 1, January 2004, pp. 140–153.

61. Weiland JD, Yanai D, Mahadevappa M, Williamson R, Mech BV, Fujii GY, Little J, Greenberg RJ, de Juan E Jr, and Humayun MS. Electrical stimulation of retina in blind humans. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 3, September 17–21, 2003, pp. 2021–2022.

62. Wise KD, Anderson DJ, Hetke JF, Kipke DR, and Najafi K. Wireless implantable microsystems: high-density electronic interfaces to the nervous system. Proceedings of the IEEE, Vol. 92, No. 1, January 2004, pp. 76–97.

63. Yu H, Olsson RH, Wise KD, and Najafi K. A wireless microsystem for multichannel neural recording microprobes. Solid State Sensor, Actuator and Microsystem Workshop, Hilton Head Island, South Caroline, USA, June 2004, pp. 107–110.

64. Branner A, Stein RB, Fernandez E, Aoyagi Y, and Normann RA. Long-term stimulation and recording with a penetrating microelectrode array in cat sciatic nerve, , IEEE Trans Biomed Eng, 2004, 51(1), , pp. 146–157, January.

65. Rebello KJ. Applications of MEMS in surgery. Proceedings of the IEEE, Vol. 92 , No. 1, January 2004, pp. 43–55.

Chapter 17
Magnetic Particles for Biomedical Applications

Raju V. Ramanujan

17.1 Introduction

This chapter discusses applications of magnetic materials in bioengineering and medicine [1–8]. Magnetism and magnetic materials have been used for many decades in many modern medical applications, and several new applications are being developed in part because of the availability of superior electromagnets, superconducting magnets and permanent magnets [9–12]. Advances in the synthesis and characterization of magnetic particles, especially nanomagnetic particles, have also aided in the use of magnetic biomaterials [6–12]. We begin with an introduction to magnetism and magnetic materials, followed by a discussion of the characterization, synthesis techniques and applications of magnetic biomaterials [8, 9]. Magnetic materials can be applied to cell separation, immunoassay, magnetic resonance imaging (MRI), drug and gene delivery, minimally invasive surgery, radionuclide therapy, hyperthermia and artificial muscle applications [1–5, 7]. Physical properties which make magnetic materials attractive for biomedical applications are, first, that they can be manipulated by an external magnetic field – this feature is useful for separation, immunoassay and drug targeting, and second, hysteresis and other losses occur in alternating magnetic fields – this is useful in hyperthermia applications.

In biology, there has been much interest in the possible use by bees and pigeons of magnetic materials as biological compasses for navigation. Some magnetotactic bacteria are known to respond to a magnetic field, they contain chains of small magnetite particles and they can navigate to the surface or bottom of the pools that they live in using these particles. These particles can be obtained by disruption of the cell wall followed by magnetic separation; the presence of the lipid layer makes these particles biocompatible and they can be readily functionalized for a variety of biomedical applications.

The earliest known biomedical use of naturally occurring magnetic materials involves magnetite (Fe_3O_4) or lodestone which was used by the Indian surgeon Sucruta around 2,600 years ago. He wrote in the book Ayurveda that magnetite can be used to extract an iron arrow tip. Current areas in medicine to which magnetic biomaterials can be applied include molecular and cell biology, cardiology, neurosurgery, oncology and radiology.

In the human body, there is a constant movement of ions within and outside the cells as well as across cellular membranes. This electrical activity is responsible for magnetic fields, called biomagnetic fields, which we can measure using sensitive instruments placed outside the body. The study of such fields, called biomagnetism, is a fascinating area related to magnetism which is not covered in this chapter due to lack of space. Of course, the effect of magnetic fields on humans and animals is also the focus of many studies, examples being the effect of the electromagnetic field produced by power lines and cell phones on humans. Of course, we are all immersed in the earth's magnetic field of about 0.5×10^{-5} T while the magnetic field of a neutron star is of the order of 10^8 T!

Here we focus on the origin of magnetism in materials, the types of magnetic materials used in medicine, and contemporary and future applications of magnetic biomaterials.

17.2 Magnetism and Magnetic Materials

Magnetism is known to all of us from childhood as the phenomenon by which some materials attract or repel other materials from a distance; examples of such materials include iron, lodestone and some steels. Broadly, magnetic forces are generated by moving charged particles, leading to magnetic fields. There are a number of excellent references to magnetism and magnetic materials, and an introduction is provided in Callister, from which the following discussion is derived [10].

Consider a material placed in an external magnetic field. The atoms in this material possess an atomic moment which responds to this external field. It is useful to think of magnetic dipoles existing in magnetic materials; these dipoles can be considered to be small bar magnets with north and south poles. The dipoles possess a magnetic dipole moment which can respond to the external magnetic field. Some field vectors are needed to understand this response: the external magnetic field strength is denoted by H (units A/m), the magnetic induction in the material is denoted by B (units tesla) and the magnetization by M (units A/m). B, H and M are related by

$$B = \mu_0 \,(H + M) \tag{17.1}$$

where μ_0 is the permeability of free space (its magnitude is 1.257×10^{-6} H/m) and M is the magnetic moment m per unit volume of the material. The value of M depends on the type of material and the temperature and can be related to the field H through the volumetric magnetic susceptibility χ by the relation

$$M = \chi H \tag{17.2}$$

17.2.1 Categories of Magnetic Materials

We now discuss the magnetic response of bulk material. In simple cases we can understand in a straightforward fashion this response in terms of the behavior of individual atoms. In other cases, interactions between individual atoms makes the picture more complicated. The magnetic response results in materials being classified as either diamagnetic, paramagnetic or ferromagnetic. Antiferromagnetism and ferrimagnetism fall within the broad category of ferromagnetism.

For most bulk materials, the response to an external magnetic field is weak, e.g., in diamagnetic and paramagnetic materials. Diamagnetism is very weak and not permanent; it persists only as long as the external field is present. It occurs due to a change in the orbital motion of electrons due to the external field, the direction of the induced magnetic moment is opposite to the field. In an inhomogenous field, such materials are attracted towards regions where the field is weak (Fig. 17.1). In paramagnetism, each atom has a permanent dipole moment because of incomplete cancellation of its electron magnetic moments. When a field is applied these atomic dipoles *individually* tend to align with the field, much as a compass needle aligns with the earth's magnetic field.

Diamagnetic and paramagnetic materials exhibit magnetization only in the presence of an external field; the low values of susceptibility χ imply that the magnetic induction in such materials is very weak. Typical values of susceptibility, at room temperature, for diamagnetic copper is -0.96×10^{-5}, for paramagnetic aluminum is 2.07×10^{-5} and paramagnetic manganese sulfate is 3.7×10^{-3} [10].
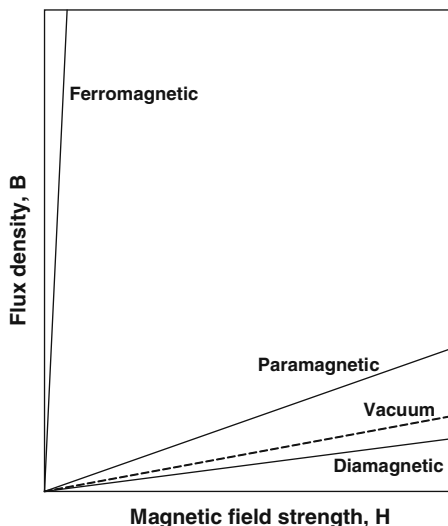
Fig. 17.1 Schematic of the flux density B as a function of H for various materials [10]

Feromagnetism is the most familiar type of magnetism. It occurs, for example, in body centred cubic (b.c.c.) iron, cobalt, nickel, and in many alloy compositions based on Fe, Co and Ni. Ferromagnetic materials, unlike dia- and para- magnetic materials, show permanent magnetic moments even in the absence of an external field. The susceptibility values are very high compared to those of para- and dia-magnetic materials, reaching up to 10^6. The magnetic moments in such materials arise mainly from atomic spin magnetic moments. More importantly, interactions between atoms cause spin magnetic moments to align with one another in a *cooperative* fashion. Thus, large regions in a crystal can have atoms with their spins aligned with one another. When all the magnetic dipoles are aligned the magnetization reaches its saturation value (M_s), e.g., the magnitude of M_s for nickel is 5.1×10^5 A/m.

As mentioned earlier, ferromagnetism results from a cooperative *parallel* alignment of spins. In other materials, e.g., MnO. The magnetic moment coupling between atoms (or ions) results in the spin moments of neighboring atoms being aligned in *opposite* directions. Such materials are antiferromagnetic. In the case of MnO, the moments of adjacent Mn^{2+} ions are antiparallel, thus the material has no net magnetic moment (Fig. 17.2).

Some materials, including the magnetic biomaterial magnetite (Fe_3O_4) mentioned in the introduction, exhibit ferrimagnetic behavior [10]. Hexagonal ferrites and garnets are other ceramic materials that fall in this category. Cubic ferrites, such as magnetite, can be represented as MFe_2O_4, where M is a metal. In the case of Fe_3O_4, Fe ions exist in both the +2 and +3 valence states. The magnetic moments of the two types of Fe ions differs; in this case, there is a net magnetic moment because for the solid as a whole the spin moments are not completely cancelled; although the spin moments of the Fe^{3+} ions cancel one another, the magnetization arises from the parallel alignment of the moments of the Fe^{2+} ions (Fig. 17.3). By adding other ions such as Ni^{2+} and Co^{2+} to Fe_3O_4, ferrites having a range of magnetic properties can be produced. This flexibility can be used to tune the magnetic properties for hyperthermia applications by creating cubic mixed-ferrite material.
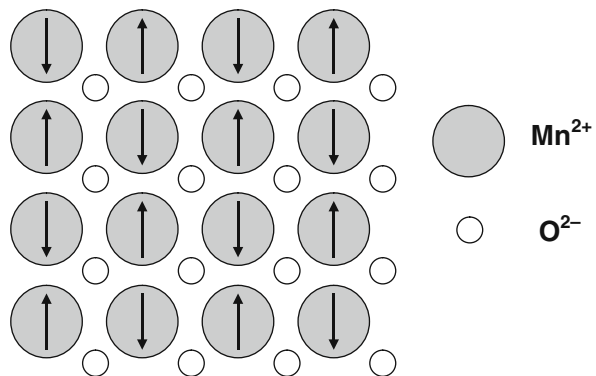
Fig. 17.2 Schematic of antiparallel alignment of spin magnetic moments in antiferromagnetic MnO

Fig. 17.3 Schematic depicting the spin magnetic moments for Fe^{3+} and Fe^{2+} in Fe_3O_4
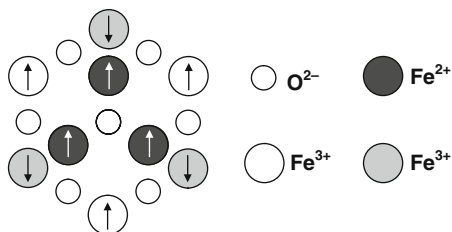
17.2.2 The Influence of Temperature

It can be expected that temperature will play an important role in determining magnetic properties, since entropy effects will be more dominant at high temperatures. The magnetic properties of both ferri- and ferro-magnets depend on the coupling forces between neighboring atoms; at higher temperatures, entropy effects favor a random arrangement of spins, resulting in a reduction in saturation magnetization. The saturation magnetization decreases with increasing temperature; at the Curie temperature T_c it becomes zero and the material becomes paramagnetic above this temperature. The Curie temperature, e.g., of cobalt is $1,120\,°C$ and that of magnetite is $585\,°C$. Thus a given material can change its magnetic behavior depending on the temperature; its use as a magnetic biomaterial will consequently depend on the relative values of the service temperature and the Curie temperature. The ferromagnetic to paramagnetic phase transformation described above has been used to act as an on-off switch in hyperthermia applications; the magnetic material is designed to have a Curie temperature equal to the temperature required for hyperthermia.

17.2.3 Magnetization Processes in Ferromagnetic and Ferrimagnetic Materials

The B-H loop enables us to classify magnetic materials into different categories, the simplest division being that of soft and hard magnetic materials. In a hard magnetic material, the area within the B-H loop is much larger than that in a soft magnetic material, and the coercivity is large, giving rise to a "fat" B-H loop; this implies a greater amount of hysteresis in a hard magnet. On the other hand, soft magnets have skinny B-H loops; as a result, soft magnets are be easily magnetized and demagnetized.

Examples of magnetically soft materials are commercial iron ingot, oriented silicon-iron and Ferroxcube A (48% $MnFe_2O_4$ –52% $ZnFe_2O_4$) with saturation magnetic induction of 2.14, 2.01 and 0.33 T, respectively. Hard magnets are difficult to demagnetize, hence they are referred to as permanent magnets. The energy product, $(BH)_{max}$ is one of the useful parameters to classify hard magnets; it is the area of the largest B-H rectangle we can construct in the second quadrant of the B-H loop. Two types of hard magnets are commercially useful: conventional and high

energy. Examples of conventional hard magnets are Cunife (60% Cu-20% Ni-20% Fe) with a remanence of 0.54 T and an energy product of 12 kJ/m^3 and sintered ferrite 3 (BaO-6 Fe$_2$O$_3$) with a remanence of 0.32 T and an energy product of 20 kJ/m^3. The usual range of energy product of conventional hard magnets is between 2 and 80 kJ/m^3, while that of high energy magnets is greater than 80 kJ/m^3. Examples of the latter are samarium–cobalt magnets (SmCo$_5$) and neodymium–iron–boron (Nd$_2$Fe$_{14}$B) magnets with typical remanence values of 0.92 and 1.16 T respectively and energy product values of 170 and 255 kJ/m^3 respectively [10].

17.2.4 Factors Affecting Magnetic Properties

From a materials standpoint it is useful to distinguish those properties that are structure sensitive from those that are relatively structure-insensitive; susceptibility and coercivity fall in the first category while saturation magnetization is an example of the second. Manipulation of the properties can be performed by altering the composition, crystal structure, stress state and size of the material. Since size plays an important part in many magnetic biomaterials applications, we now consider the effect of particle size on magnetic properties (Fig. 17.4).

In for example drug delivery, gene delivery and hyperthermia, small magnetic particles are used [5]. In large particles (greater than about 1 μm) there are many magnetic domains; this leads to a narrow hysteresis loop. Such particles are useful in immunomagnetic separation of pathogenic microorganisms in microbiology. For smaller particle sizes (less than about 1 μm) it is energetically more favorable for only one domain to exist. The response of such particles to a magnetic field is qualitatively different, resulting in a broader hysteresis loop. If the particle size is reduced further to about 20 nm (the exact size depends on the composition of the material), the material becomes superparamagnetic, which means that the
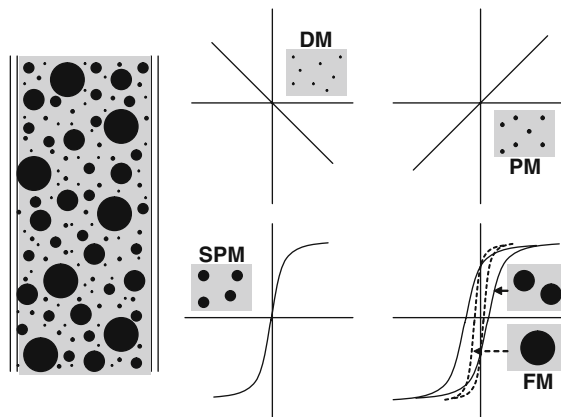
Fig. 17.4 Magnetic properties are affected by the particle size (DM = diamagnetic, PM = paramagnetic, SPM = superparamagnetic, FM = ferromagnetic) [5]

magnetic moment of the particle fluctuates because of the thermal energy ($\sim$kT); at the atomic level the individual atomic moments continue to be ordered relative to each other. Importantly, the remanence is zero, the result is a B-H curve showing no hysteresis; this property is important for reducing the tendency of the particles to agglomerate. The physical basis for the fluctuation of the magnetic moments can be understood as a battle between ΔE, the energy barrier to moment reversal and the thermal energy (kT). In the simplest approximation the energy barrier is the product of the anisotropy energy density K and the volume V. When the particle size is small (small V), the KV term is small and comparable to the thermal energy; this leads to flipping of the magnetic moment. The "blocking" temperature T_B can be regarded as the temperature above which the material becomes superparamagnetic. Superparamagnetic particles are useful as magnetic biomaterials, some are, physiologically well tolerated; an example is dextran-magnetite. Iron oxide coated with dextran is commercially available for MRI, for cell separation and cell labeling applications.

17.3 Physical Principles

We now consider the physical principles involved in the applications of magnetic particles in bioengineering applications. In the case of targeted drug delivery we introduce drug coated magnetic particles into a blood vessel and then apply an external magnetic field. This field attracts and retains the particles at the site of the disease (Fig. 17.5). The blood vessel will exhibit a paramagnetic response to the field from entities such as the hemoglobin. It will also exhibit a diamagnetic response because
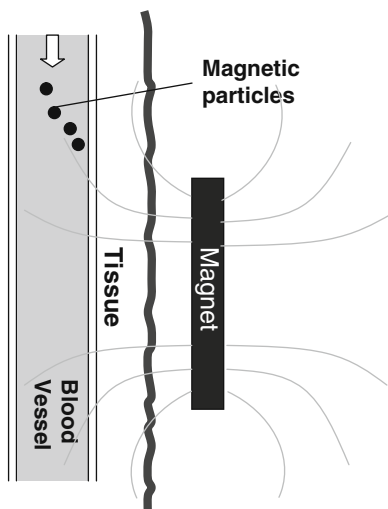
Fig. 17.5 Directed motion of magnetic particles by external magnetic field [5]

of proteins that contain carbon, hydrogen, nitrogen and oxygen atoms. These two responses are much smaller than the response by the magnetic particles. What kind of magnetic field should be applied to target the magnetic particles to the site of the disease? A magnetic field that is uniform gives rise to a torque, but usually we wish to direct the particles in a specific direction, i.e., provide translation motion; this can be accomplished by means of a field gradient. If an appropriate magnetic field gradient is present, a force acts on the particles driving them in a direction which can be chosen to so that the particles are targeted to the site of the disease (Fig. 17.6). Consider a magnetic particle in such a magnetic field gradient; for the case of magnetic nanoparticles suspended in water the magnetic force on the particle has been shown by Pankhurst to be

$$F_m = V_m(\chi_m - \chi_w)\nabla(\frac{1}{2}B.H) \tag{17.3}$$

where V_m is the volume of the particle, χ_m and χ_w are the susceptibility of the particle and water respectively, and the quantity $\frac{1}{2}B.H$ is the magnetostatic field energy density [5]. Assuming that the susceptibility of the particle is greater than that of water, this equation shows that the force is proportional to the differential in the energy density, and recalling the geometrical meaning of the ∇ operator, this magnetic force on the particle acts in the direction of steepest ascent of the energy density field.

   In the case of magnetic hyperthermia applications, a different principle is involved; we wish to raise the temperature to about $43\,°C$ in a localized area in order to destroy cancer cells selectively [3]. This can be done by applying a magnetic field which varies with time; ferro- and ferri-magnetic material will be repeatedly cycled through the B-H loop, resulting in hysteresis and other losses which are then converted to thermal energy and result in an increase in temperature.
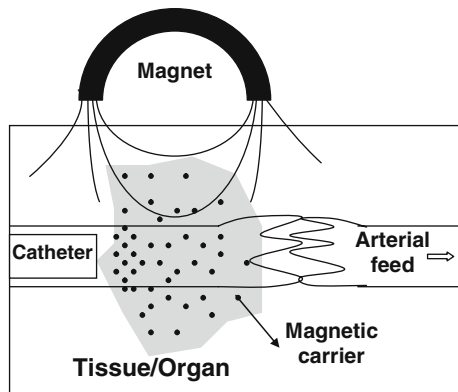
Fig. 17.6 Magnetic carrier for drug targeting and drug delivery [6]

Superparamagnetic materials can also be heated using this technique; the loss mechanisms differ from those observed in ferro- and ferri-magnetic materials.

17.4 Examples and Property Requirements of Magnetic Biomaterials

A common example of a magnetic biomaterial is magnetite. Interestingly, this is the same material used by Sucruta more than 2,000 years ago. Magnetite is found in many biological entities, from bacteria to people. It is an example of cubic ferrites which have an inverse spinel structure. It is ferrimagnetic with a Curie temperature of $578\,°C$ and a saturation magnetization of 4.76×10^5 A/m. Another example of a magnetic biomaterial is maghemite (gamma Fe_2O_3), which is formed when magnetite is oxidized. It has a structure similar to that of magnetite, the difference being that all or most of the iron is trivalent, the saturation magnetization is 4.26×10^5 A/m. Ferritin is a protein that stores iron in humans. It contains typically 4,500 iron atoms in an approximately spherical 12 nm diameter molecule. It has a 12 subunit protein shell containing a ferrihydrite core and an antiferromagnetic core. Gadolinium(III) chelates are commonly used in MRI applications. Iron coated with activated carbon has recently been tried for magnetic drug targeting for the treatment of hepatocellular carcinomas.

The magnetic biomaterial can, in principle, also be Ni, Co, b.c.c. Fe, magnetic alloys of Fe, Co, Ni, Nd–Fe–B or samarium–cobalt materials. In all cases, however, issues of biocompatibility and toxicity limit the choice of materials; however, the use of coatings may make the use of these materials feasible. The hard magnetic materials Nd–Fe–B and samarium–cobalt have the disadvantage that large external fields are required to influence these materials. Materials with high magnetization and high susceptibility are preferred for applications such as drug targeting and magnetic separation. Most of the examples of useful magnetic biomaterials are in powder form, and usually the particles are suitably coated before use. Ideally, the magnetic material should be non-toxic and non-immunogenic. In the case of drug delivery, the particle sizes should be small enough to be injected into the bloodstream and then to pass through the required capillary systems. For in vitro applications, the requirements are more relaxed, larger particle sizes can be used, and biocompatibility and toxicity issues may be less important.

In many cases, it is required to coat the magnetic particles; this is usually done by coating with a biocompatible polymer or with other coatings such as gold, activated carbon or silica (Fig. 17.7). The coating reduces aggregation and prevents the magnetic particle from being exposed directly to the body. In addition, the polymer can be used as a matrix in which drugs, radionuclides or genetic material can be dissolved or as a site for binding of drugs; thus the magnet-coating system can act as "carrier" to deliver useful material to the targeted region. Some examples of common coatings are derivatives of dextran, polyethylene glycol (PEG) and polyethylene oxide (PEO), phospholipids and polyvinyl alcohol [2, 4].
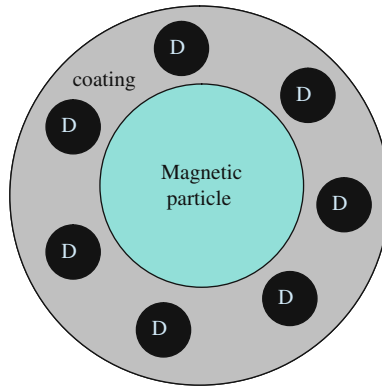
Fig. 17.7 Magnetic particle with biocompatible coating to be used for in vivo applications. D refers to functional groups

17.5 Applications

In this section we discuss present and potential future applications of magnetic biomaterials. We begin with the well studied case of magnetic separation for purification and cell labeling.

17.5.1 Magnetic Separation

Magnetic particles can be used to separate entities from their surroundings so that the surroundings can be purified or to concentrate the entities for further study [1]. This use is based on the difference in the susceptibility between a magnetically labeled entity and the surrounding medium. Examples of the use of this principle are magnetic cell sorting for cellular therapy and immunoassay (which is a process that measures and identifies a specific biological substance such as an antigen). Entities that can be labeled include cells, bacteria and some types of vesicles. The first step is to label the entities with the particles followed by the separation of the labeled entities by magnetic separation (Fig. 17.8). Usually coated particles will be used; the coating will help to bind the particles to the entities such as cells. Specific sites on the cell surfaces can be targeted for attachment by antibodies; this works as a labeling procedure since antibodies bind to their matching antigen. In order to separate out these labeled entities we can use a magnetic field gradient which can attract and "hold" the entities in specific regions, followed by removal of these entities. This method has been applied to the selection of tumor cells from blood as well as to isolate enzymes, DNA and RNA from various sources including body fluids.
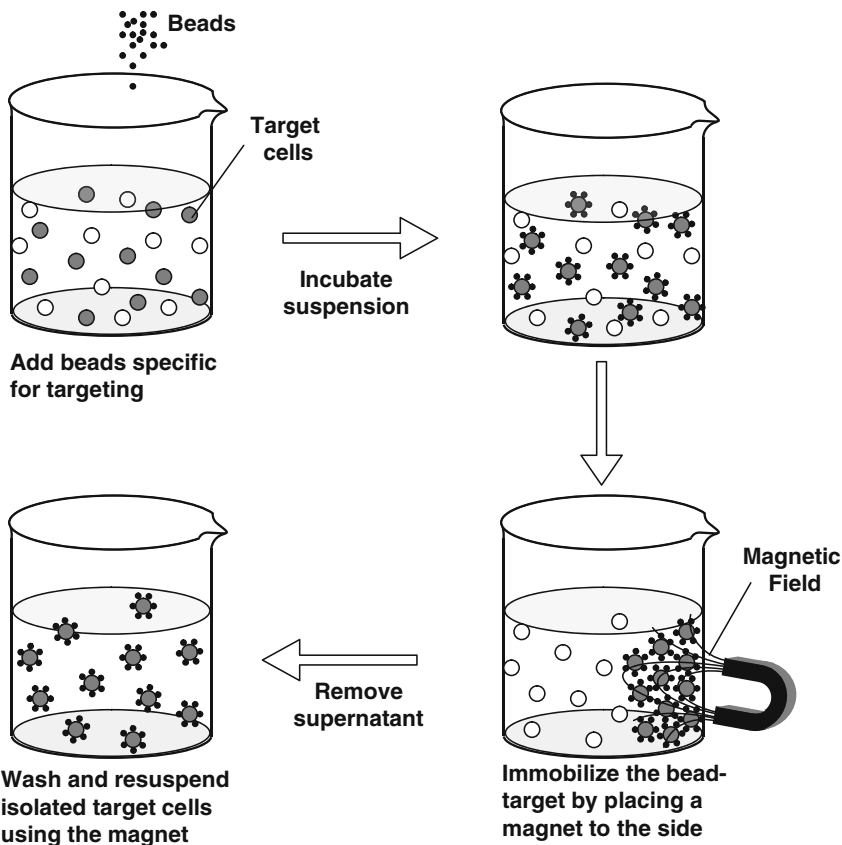
Fig. 17.8 Magnetic separation [4–6]

17.5.2 Drug Delivery

It is well known that most methods of chemotherapy are relatively non-specific. As a result, drugs are "wasted" by being distributed to areas where they are not required, and this can lead to undesirable side effects. What is required is a "magic bullet" which can be used to carry the drugs to the area where they are required. Additionally, it will be advantageous to control the release of the drug once the carrier has reached the target. Much of the work which has been performed so far relates to targeting cancerous tumors. In this case, the magnetic particle first acts as a carrier of the drug which is attached to its outer surface or, more commonly, dissolved in an outer coating. Once the drug coated particles have been introduced into the bloodstream of the patient, a magnetic field gradient, created, e.g., by a strong permanent magnet, is used to "hold" the particles at the targeted region.

Drug release can then take place by enzymatic activity or by means of a specific trigger (controlled release). Physically, the success of this method depends on

the velocity of the particles in the bloodstream, the circulation time as well as the strength of the magnetic field in the region of interest. It is easier to immobilize larger particles in the micron size range; they are less likely to "swept" away by the blood flow. On the other hand, there are many other advantages in having a reduced particle size. It appears that the optimum size may be in the 5–100 nm size range. Typically, a coated magnetite particle is used; noble metals such as gold have been used as a coating in addition to the usual polymer based coatings. Preliminary studies using metallic magnetic particles have also been undertaken. Several studies have reported success in tumor remission (the period during which the symptoms of a disease decrease) in animal models. The difficult case of targeting cytotoxic (cell killing) drugs to brain tumors has also been demonstrated in rats. Human trials using this technique for treating liver tumors have also been performed.

17.5.3 Radionuclide Delivery

Radionuclides (e.g., β–emitters, β is the symbol used to denote an electron) can also be attached to the magnetic particles and this system can then be targeted in the same way as described in the previous section on drug delivery, since the radionuclide does not have to be released in the same way as the drug; one restriction of drug delivery, i.e., control of drug release, is absent.

17.5.4 Gene Delivery

This technique has been called magnetofection, since it involves using magnetic particles to increase the efficiency of gene transfection (introduction of foreign DNA into a host cell) and expression (the process by which a gene's coded information is converted into the structures present and operating in the cell) [7]. A viral vector carrying the gene is coated on the surface of the magnetic particle; by applying a magnetic field; this particle can then be held at the target area. The virus is thus in contact with the tissue for increased duration which increases the transfection efficiency (Fig. 17.9).

17.5.5 Hyperthermia

The idea that a localized rise in temperature (typically about 43 °C) can be used to destroy malignant cells selectively is called hyperthermia; this method of treatment can be effected by magnetic particles. The basic idea is that magnetic material can be heated by an a.c. magnetic field. The mechanisms of heating for ferromagnetic materials include hysteresis losses. In the case of superparamagnetic particles heating can occur by the rotation of the particles themselves or by the rotation of the atomic magnetic moments. Other non-magnetic methods of hyperthermia are also
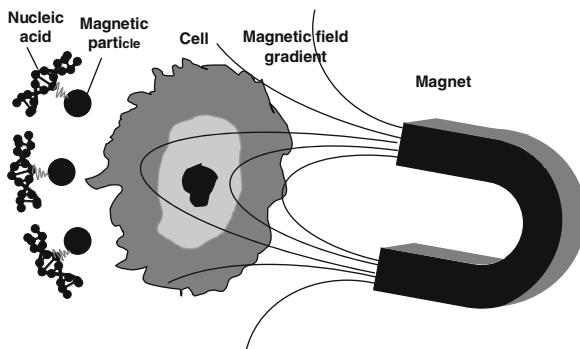
Fig. 17.9 Principle of gene delivery using magnetic particles. The cells to be transfected are positioned between the magnetic field gradient and the nucleic acids which are attached to the particles. The field causes the particles to move towards the target cells, the cell entry is accomplished, e.g., by endocytosis [7]

available. The advantage of using magnetic particles should by now be familiar, i.e., we can target the particles to the targeted region and then heat up the particles by using an external a.c. magnetic field. According to Pankhurst, typically a heat deposition rate of $100 \, mW/cm^3$ is required, and the frequency of the field should be in the kHz range with an amplitude of a few kA/m [5]. Generally iron oxides are used for hyperthermia applications.

17.5.6 Magnetic Resonance Imaging Contrast Agent

So far, we have only considered magnetic properties associated with the electrons in the material. However, protons also have a magnetic moment, and this can be utilized in the powerful imaging technique of magnetic resonance imaging (MRI) [5]. The principle is as follows. We first apply a steady field of about 1 T to a material, causing a very small fraction of protons to line up parallel to the field. The net magnetic moment precesses like a top around the direction of this field (Fig. 17.10). In order to measure the signal produced as a result of this alignment, we now apply a transverse radio frequency magnetic field. The frequency is carefully chosen and its effect is to make the magnetic moment precess in the plane perpendicular to the steady state field. When this second field is switched off, the amplitudes of the magnetic moments relax back to their initial values. This relaxation of the response is measured by pick-up coils. Typically, the relaxation time can be reduced by means of a magnetic particle. Thus if a region is tagged using the magnetic particles, the relaxation time will be lower compared to untagged regions; thus "contrast" is provided and the particle acts as a contrast agent. Usually paramagnetic Gd based materials are used; superparamagnetic iron oxide particles, usually coated with dextran, have also been used for this purpose.

Fig. 17.10 Magnetic
resonance for protons with
magnetic moment **m** in a
field **B**$_z$. (**a**) The moment
precesses around the field.
(**b**) A second field excites
the moment precession into
the plane perpendicular to
B$_z$. (**c, d**) The second field is
removed and the in-plane (**c**)
and longitudinal (**d**) moment
amplitudes relax back to
their initial values [5]
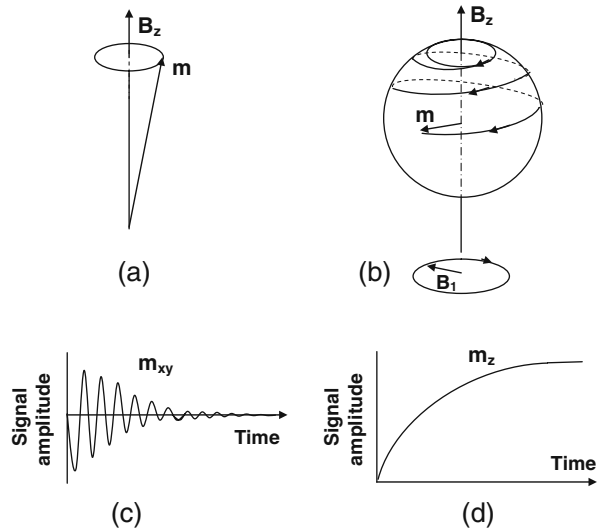
17.5.7 Artificial Muscle

There have been various attempts to synthesize artificial muscles; attempts range
from a robot-like metallic actuator to a more advanced soft actuator. Hydrogels,
which are crosslinked polymer networks swollen with water, can be used to make
soft actuators. However, most gels are relatively homogenous materials which
shrink or swell uniformly, with no dramatic change in shape. Therefore there is
a need to improve the response of gels. Magnetic field sensitive gels in which mag-
netic particles of colloidal size are dispersed and incorporated into the gels have
been developed. These ferrogels combine the magnetic properties of magnetic fillers
and the elastic properties of hydrogel. Shape distortion occurs very quickly and dis-
appears abruptly when the external magnetic field is applied and removed, respec-
tively. When these ferroqels are placed in a magnetic field gradient, forces act on the
magnetic particles and as a result of strong interaction between magnetic particles
and polymer chains, the ferrogel moves as a single unit. Polyvinyl alcohol (PVA)
gel has been used for this application because of its mechanical properties and bio-
compatibility (it has been used, for example, as a synthetic vitreous body to treat
retinal detachment); iron oxide particles are often used as the magnetic material.

17.6 Summary

The phenomenon of magnetism has intrigued and excited scientists for centuries.
Now magnetic biomaterials can fruitfully be applied to a number of biomedical
needs. Applications which are currently being investigated include magnetic cell

separation, immunoassay, hyperthermia, MRI contrast agents and drug, radionuclide and gene delivery. The greatest advantage of such materials is their ability to respond to and be manipulated by external magnetic fields. Magnetism of an atom can be thought to arise from uncompensated spins of its electrons. Magnetism associated with the proton is much weaker but is used for MRI studies. Magnetic biomaterials are usually used in particle form and can be ferro-, para-, ferri-, antiferro- or superpara-magnetic, they are often encapsulated by coatings that can be diamagnetic; this results in a wide range of magnetic responses to an applied field. The most common magnetic biomaterials are the iron oxides magnetite and maghamite, and there is considerable interest in using magnetically superior metals and alloys. Several magnetic properties are a sensitive function of size. The size can be altered by changing process parameters or synthesis techniques, and thus magnetic particles with a wide range of properties can be obtained for a given composition. It has been found that nanosized particles often possess such properties and considerable effort is now underway to create commercially attractive applications using such particles.

Since the field of magnetic biomaterials is a rapidly growing, I have relied on journal, handbook and encylopedia articles the following references are particularly useful.

References

1. Safarik I and Safarikova M. Magnetic nanoparticles and biosciences. Monatshefte fur Chemie, 2002, 133: 737–759.
2. Berry CC and Curtis ASG. Functionalisation of magnetic nanoparticles for applications in biomedicine. J Phys D: Appl Phys, 2003, 36: R198–R206.
3. Bahadur D and Giri J. Biomaterials and magnetism. Sadhana, 2003, 28: 639–656.
4. Shinkai M. Functional magnetic materials for medical applications. J Biosci Bioeng, 2002, 94: 606–613.
5. Pankhurst QA, Connolly J, Jones SK, and Dobson J. Applications of magnetic nanoparticles in biomedicine. J Phys D Appl Phys, 2003, 36: R167–R181.
6. Tartaj P, et al. The preparation of magnetic nanoparticles for applications in biomedicine. J Phys D: Appl Phys, 2003, 36: R182–R197.
7. Plank C, et al. Enhancing and targeting nucleic acid delivery by magnetic force. Expert Opin Biol Theor, 2003, 3: 1.
8. Hafeli U, Schutt W, Teller J, and Zborowski M. Scientific and Clinical Applications of Magnetic Carriers, Plenum Press: New York, 1997.
9. Andra W and Nowak H, eds. Magnetism in Medicine, Wiley-VCH: New York, 1998.
10. Callister WD. Materials Science and Engineering, Wiley: New York, 2003.
11. O'Handley RC. Modern Magnetic Materials, Wiley: New York, 2000.
12. Koch CC, ed. Nanostructured Materials, Noyes Publications: New York, 2002.

Chapter 18
Specialized Fabrication Processes: Rapid Prototyping

C.K. Chua, K.F. Leong, and K.H. Tan

18.1 Introduction

Rapid Prototyping (RP), otherwise known as solid freeform fabrication, originally developed in the late 1980s as a technique for manufacturing engineering, has evolved rapidly over the years to create a niche in the field of biomedical engineering. From the manufacturing of medical devices to customized micro-architectural implants, the manufacturing flexibility and capabilities of RP systems make it one of the most favorable systems for such biomedical applications.

RP technology made its debut in the market during the late 1980s with the introduction of the Stereolithography system (SLA-1) by 3D Systems Inc. (Valencia, CA). As opposed to conventional fabrication processes which involve the subtraction or removal of materials from stock (e.g., conventional machining), casting or molding (e.g., investment casting, injection molding, etc.), all RP techniques are material addition processes which employ the same fabrication concept of layer-by-layer additive building to produce three-dimensional physical parts from wood, plastics, metals and ceramics. RP fabrication starts with the creation of a three-dimensional volumetric computer model of the desired part that can be derived from output data generated by surface digitizers or medical imaging systems (e.g., Computed Tomography (CT) or Magnetic Resonance Imaging (MRI)). The digital model is then mathematically sliced into thin layers having a constant thickness that is user-defined. Using RP fabrication, layers of material representing the cross-sectional profiles of the desired part as obtained from the computer-generated slices are formed by processing solid sheet, liquid or powder material feed stocks. The material layers are automatically and precisely stacked and fused on top of one another to create the desired physical part. Detailed descriptions of the principles and operations of the various RP techniques can be found in existing text [1].

This chapter aims at providing readers with an overview into the biomedical applications of RP. It would be divided into two main sections: (1) Tissue Engineering Scaffolds and (2) Prostheses.

18.2 Biomedical Applications of Rapid Prototyping-Tissue Engineering Scaffolds

Mankind has always been intrigued by the possibility of being able to grow new, fully functional tissues and organs to replace damaged, injured, or worn parts of the human body. It is not until recently that such an aspiration is considered mere science fiction when the emergence of Tissue Engineering (TE) has turned this "aspiration" into medical reality [2–4]. The vast interest generated among clinicians as well as engineers seeking biologic replacements to treat malfunctioned human tissues and organs rapidly and effectively are incited by the immense potential and promise to overcome the severe shortage of donor organs and the limitations of alternative short term therapies including drug, mechanical prosthetics, etc. [5]. The scarcity of donor organs [6, 7] has created huge demands for laboratory produced biologic alternatives, and hence, paved the way for unprecedented growth and advances in TE research and applications.

Scaffold-guided TE is a major TE strategy or approach that is attracting wide interest and research effort from the scientific community. Scaffold-guided TE can facilitate the extensive building and growth of cell masses in three-dimensional forms to achieve successful cell transplantation of either hard (e.g., bone and cartilage) [8, 9] or soft (e.g., liver and blood vessels) [10, 11] tissues and organs. Using this approach, progenitor cells are harvested from the patient or a donor and expanded in population through cell culture systems [12]. Once sufficient quantities of cells are obtained, they are seeded directly onto scaffolds that would allow the cells to attach and proliferate. The cell-seeded scaffolds are subjected to an additional period of in vitro culturing in an environment (e.g., bioreactors [13, 14]) which mimics the biochemical and physical signals that regulate in vivo tissue development, to allow full colonization of the scaffold in order to achieve homogeneous and healthy tissue regeneration.

This section focuses on the application of Rapid Prototyping for creating such TE scaffolds. The main objectives are to compile the information available to date on the characteristics, parameters and applications or potential applications of these new and novel RP fabrication techniques for scaffold creation and to highlight their advantages as well as their limitations. The characteristics of the RP-fabricated scaffolds are assessed and compared. Research to develop CAD strategies to be applied in conjunction with RP to design, customize and control the internal architectures of the scaffolds is also described. It is the authors' aim to provide the TE community the essential information on the critical parameters and factors that are to be considered in choosing the appropriate automated fabrication techniques for specific TE applications.

18.3 Roles and Pre-Requisites for Tissue Engineering Scaffolds

In scaffold-guided TE, the scaffolds serve as temporary surrogates for the native extracellular matrix (ECM) that mechanically supports cellular anchorage,

reorganization of cells in three dimensions and maintaining differentiated pheno-types and tissue specific functions once implanted. Scaffolds for TE applications must be produced from scaffolding materials which exhibit good biocompatibility and as such are limited to materials that are non-mutagenic, non-antigenic, non-carcinogenic, non-toxic and non-teratogenic. To date, a variety of naturally-derived biological substances (e.g., collagens, chitosan, fibrin, etc.) [15, 16] have been used as scaffolding materials with promising results. The need for large quantities of scaf-folding materials to cater for clinical applications as well as assurance of pathogen removal [17, 18] has led to the investigations and inclusions of synthetic biocompat-ible polymers (e.g., poly(glycolic acid), poly(L-lactic acid), etc.), copolymers (e.g., poly(DL-lactic-*co*-glycolic acid)) [19–21] and bioceramics (e.g., hydroxyapatite, tri-calcium phosphate) [22, 23] as scaffolding materials. Besides being formed from a suitable biomaterial, the scaffolds must possess appropriate three-dimensional forms that are spatially and anatomically consistent with the defect and micro-architectural properties [24, 25] that will support tissue regeneration. The scaffold's micro-architecture affects not only cell survival, signaling, growth, propagation and reorganization, but also plays major roles in influencing cell shape modeling and gene expressions that relate to cell growth and the preservation of native pheno-types [26, 27]. In order for a scaffold to be effective in its intended TE application, the scaffold must posses a large surface area to volume ratio to achieve high yields of attached cells [28]. It must also possess a highly porous micro-structure (>90% porosity [29–31]) with interconnected porous networks that will allow the in-growth of cells and provide uninhibited diffusion pathways to facilitate nutrient, gas and waste exchange with cells that have penetrated into the scaffold. In these porous constructions, the pore sizes must be large enough so that the cells can penetrate and grow within the pores without occluding the pore spaces [32, 33]. Besides archi-tectural designs, scaffolds fashioned from degradable biomaterials must be tailored such that sufficient mechanical strengths are preserved in the degrading construct to maintain the pore spaces for the in-growth of cells and to support early mobilization of the implant site.

18.4 Conventional Manual-Based Scaffold Fabrication Techniques

A wide variety of manual-based fabrication methods and techniques are available that can be used to transform the biomaterials into scaffold structures [17, 34–37]. These manual-based techniques include fiber bonding [38], phase separation [39, 40], solvent casting with particulate leaching [41], membrane lamination [42, 43], melt molding [44], gas foaming with high pressure processing [45, 46], hydrocar-bon templating [47], freeze drying [48, 49] and combinations of these techniques (e.g., gas foaming with particulate leaching [50], etc.). As different cell types require different sets of structural requirements, the fabrication techniques mentioned have been tailored to produce scaffolds with their particular and optimized characteristics

for specific TE applications. As such, none of the techniques are sufficiently generic to be suitable for all TE applications. Although most of these techniques have been used in preliminary investigations for engineering a variety of tissues with varying degrees of success, most carry some forms of imperfection that restrict the scope of their applications. The extensive use of porogens and toxic solvents in most of these techniques results in adverse drawbacks due to the risk of inadvertent residual solvents and porogen entrapped in the polymer matrix. Another highly detrimental weakness in manual-based techniques is their high dependency and sensitivity to manual skills, experiences and procedures, often resulting in poor consistency and poor repeatability of end results. Other factors such as high labor involvement, lengthy processing time and poor flexibility to produce complex macro-architectures lead to increased production costs. The processing procedures, advantages and limitations of the various manual-based conventional fabrication techniques can be found in several published reviews [34–37].

18.5 Computer-Controlled Freeform Fabrication Techniques for Tissue Engineering Scaffolds

Although immense research efforts have been placed in developing new biocompatible scaffolding materials and processing techniques for converting such materials into useful scaffold structures, relatively little work had been carried out towards the development of rapid and automated procedures for designing and producing TE scaffolds. Recently, a group of computer-controlled fabrication techniques, namely, rapid prototyping techniques have been identified and recognized to possess significant potentials for TE scaffold fabrication. The manufacturing flexibility and fabrication capabilities of RP systems permit complex irregular three-dimensional scaffold geometries with designed or customized internal micro-architectures to be fabricated. Such capabilities are highly advantageous since the ideal scaffold should replicate the geometry and size of the patient's original anatomy and its internal micro-architecture should support cell proliferation and spreading. The fact that each tissue and organ in the human body has their own unique geometry which varies in size between individuals undermines the applicability of most conventional fabrication techniques which are restricted to the fabrication of scaffolds with highly simplified geometries. Although the application of RP for the production of TE scaffolds is still very much in the laboratories, the vast amounts of interest generated by RP for TE scaffold fabrication is demonstrated by the number of publications generated over the last 5 years [51–71]. Research conducted with existing commercial and non-commercial RP systems has laid down a firm foundation in generating scaffolds with unprecedented quality, accuracy and reproducibility and is fast establishing RP as a fabrication method of choice in TE scaffold production.

To date, only a small number of RP techniques have been exploited for scaffold fabrication. Since most RP systems available commercially are designed to cater mostly for engineering-related industries, work conducted to date had mainly been

focused on adapting and modifying such systems for processing biomaterials and scaffold production. Detailed descriptions of the principles and operations of the various RP techniques can be found in some publications [1]. The following sections describe the major RP techniques that have been extensively researched for fabricating TE scaffolds. The section is divided into three main sub-sections: (1) solid-based techniques; (2) powder-based techniques and (3) liquid-based techniques.

18.5.1 Solid-Based Techniques

Solid-based RP systems are meant to encompass all forms of material in solid state that can be in the form of a wire, roll, laminates and pellets. Most of such systems fabricate parts by means of cutting and joining method or melting and fusing/solidifying method.

18.5.1.1 Fused Deposition Modeling (FDM)

The fused deposition of material technique uses the concept of material extrusion to deposit thin filaments of liquids, pastes, melts, solutions or reactive material to form the cross-sectional profiles of the desired part. The technique was originally introduced by Stratasys Inc. (Eden Prairie, MN), as fused deposition modeling (FDM) and is based on the extrusion of polymer melts. Among the variants of the FDM technique are fused deposition of ceramics (FDC), precise extrusion manufacturing (PEM), low-temperature deposition manufacturing (LDM), rapid prototyping robotic dispensing (RPBOD) and Bioplotter. In these techniques, the scaffolding materials are extruded from a nozzle in thin filament forms known as material roads. The nozzle is attached to a moving mechanism (e.g., robotic arm, x–y table) that translates in the horizontal plane to direct the deposition of the material roads onto a building platform. The material roads are arranged in parallel with a pre-defined spacing (i.e., raster gap setting) to form the individual material layers which are then stacked on top of one another to replicate the third (height) dimension of the scaffold object. Variations in the angle or direction of material deposition (i.e., lay down pattern) for each individual layer can be utilized to create scaffolds with varying honeycomb-like micro-architectures.

The FDM systems available commercially impose many constraints on scaffold fabrication. The nature and operating limits of the FDM systems reduce the choice of scaffolding materials to thermoplastics with low melting points and relatively low melt viscosities. Further to this, there is a need to convert the polymeric scaffolding materials which are usually supplied by the material vendors as pellets, powders or flakes into filament form with good diametrical consistency before they can be processed on the FDM. Currently, the FDM had been successfully adapted by Zein et al. [56] for fabricating poly(ϵ-caprolactone) (PCL) scaffolds with pore sizes (channel widths) and porosity values ranging from 160–700 μm and 48–77% respectively that were subsequently seeded with primary human fibroblast and osteoblast cells [59]. Complete ingrowth and colonization of the scaffolds by cellular tissues were

reported after a 3- to 4-week culture period. Cornejo et al. [62] developed the fused deposition of ceramics (FDC) system which processes ceramic loaded thermoplastic filaments for producing ceramic scaffolds from β-tricalcium phosphate (TCP). The polymeric binder was removed via a burnout process and the ceramic component was sintered at 1250 °C to yield scaffolds with minimum pore widths of 300 μm.

To address the limitations imposed on the choice and physical forms of the scaffolding materials that can be processed by existing FDM systems, different research groups have developed their own material extrusion systems that are dedicated for scaffold fabrication. Xiong et al. [60] developed the Precise Extrusion Manufacturing (PEM) system that is suited for the processing of biopolymers. While the FDM system requires filament feedstock, the PEM system is capable of accepting polymer powder, pellet or granule feedstock. The pressure to extrude the polymer melt is generated by compressed air. Xiong et al. utilized the PEM system for fabricating porous cylindrical poly(L-lactic acid) (PLLA) scaffolds with porosities of around 60.3% and varying channel widths of between 200–500 μm. Further developments achieved by Xiong et al. [72] led to a low-temperature deposition manufacturing (LDM) system that was specially developed for producing composite PLLA/TCP scaffolds for bone tissue engineering applications. The LDM fabrication process is carried out within a sub-zero environment using a slurry feedstock with a composition of 15 wt% PLLA, 15 wt% TCP and 70 wt% dioxane. For their evaluation hollow tubular PLLA/TCP scaffolds with internal and external diameters measuring 5 and 10 mm respectively and channel widths of 400 μm were produced. The average porosity of the scaffolds was around 89.6%. To assess their biocompatibility and efficacy, the LDM-fabricated PLLA/TCP composite scaffolds preloaded with about 150 μL of bovine bone morphogenic protein (bBMP) at 200 g/L concentration were used to repair a 20 mm segmental defect in canine radiuses. Results of radiology and physical examinations after 24 weeks of implantation showed complete regeneration of new bone tissues over the entire scaffold.

Ang et al. [52] developed a rapid prototyping robotic dispensing (RPBOD) system that is capable of processing gel material feedstocks (chitosan and chitosan/HA mixtures). To evaluate the RPBOD system, rectangular chitosan and chitosan/HA composite scaffolds measuring $5 \times 5 \times 3$ mm were fabricated. In vitro cell culture studies were carried out to assess the biocompatibility of the scaffolds. Samples of the scaffolds measuring were seeded with human osteoblast cells and cultured for a 4-week period. The cells exhibited healthy morphology and strong proliferative ability throughout the culture period for all seeded scaffolds therefore ascertaining scaffold cell biocompatibility.

A commercial machine, Bioplotter, developed separately by Freiburg Materials Research Centre, Germany, and marketed by Envision Technologies GmBH, Germany, [70] has similar design concepts and working principles as the RPBOD system. The Bioplotter is capable of processing a wide range of biomaterials including thermoplastic polymers, reactive resins (i.e., silicone resins), structured hydrogels (composed of alginic acid, gelatine, collagen, fibrin or other hydrogels) and thermoreversible hydrogels. In the Bioplotter setup, an 80-μm needle is employed as the extrusion nozzle where liquids, pastes, melts, solutions, hot melts, reactive

oligomers or dispersions which are initially stored in a heated cartridge, are extruded into a temperature controlled liquid dispensing medium. The dispensing medium induces solidification of the deposited material by cooling, heating or through chemical reaction. Also, by using a dispensing solution of similar density as the building material, the buoyancy exerted by the medium on the build can prevent the collapse of complex structures thus eliminating the need for sacrificial support structures which are typical in conventional FDM systems.

The authors have successfully built and studied the micro-architectures of different scaffold constructions produced using high density polyethylene (HDPE) with the FDM. These scaffolds were built with lay down patterns supplied in the system software or from the works of Too et al. [55] and Hutmacher et al. [64]. There are no restrictions to the number of lay down patterns that can be created by changing the direction of material extrusion for each consecutive layer of a FDM structure. However, as demonstrated by Too et al. [55] and Hutmacher et al. [64], the lay down pattern employed has its own set of mechanical and physical characteristics that may generate specific or unique cell or host tissue response. The pore interconnectivity of FDM-fabricated structures is observed from the morphology of a sliced structure taken at a 90° angle to the y-axis (Fig. 18.1).

Although the FDM is capable of fabricating various honeycomb-like architectures, the pore morphologies obtained for FDM-fabricated scaffolds are not consistent along the three dimensions (x-, y- and z-axes) when built using the standard build patterns provided by the system software as illustrated in Fig. 18.2. The pore openings facing the z-direction are determined by the lay down pattern employed during fabrication which is directly controlled by user-defined parameters. Pore openings facing the x- and y-directions (both are similar) are created by the stacking of material layers and material layer. In addition, the pore openings in both the x- and y-directions are partially occluded by the deposited material that bridges consecutive roads. A method to circumvent such limitation to attain consistent pore
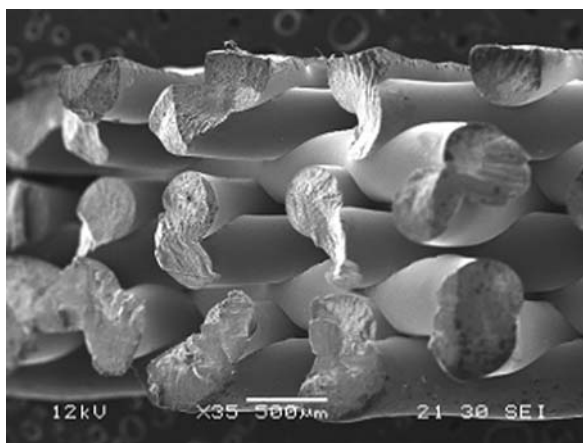
Fig. 18.1 Micrograph showing three-dimensional pore interconnectivity within FDM-fabricated structure

profile of the desired implant structure. Water-based alumina slurry were infiltrated into the mold cavities and allowed to dry before the mold is pyrolyzed. The resulting ceramic structure was then sintered at a higher temperature. Sample scaffolds with porosities between 33% and 50% and various pore sizes ranging from 300 to 750 μm have been fabricated with this method. In order to convert the bioinert alumina implants into bioactive and osteo-conductive devices, the implants were dip-coated with HA [73] and fired to consolidate the coating. In vitro cell culture assessment carried out with human osteosarcoma cells yielded high numbers of attached cells and good uniformity in the distribution of the attached cells throughout the entire structure of the implants.

18.5.1.2 ModelMaker II (MM II)

The machine, developed by Sanders Prototype, Inc. and currently distributed by Solidscape, Inc., employs two ink-jet type printing heads to build models. One head deposits the thermoplastic building materials while the other deposits supporting wax. The liquefied build material cools as it is ejected from the print head and solidifies upon impact on the model. After the layer is completed, a horizontal cutter is employed to flatten the top surface before the process repeats itself till a complete model is obtained.

In recent years, MM II has been employed to fabricate a negative mold to create the TE scaffolds in which biomaterials were cast into the mold to obtain biocompatible scaffolds [74–76]. Limpanuphap et al. [74] obtained a suspension comprised of tricalcium phosphate in diacrylate cross-linking monomers to form a gel and cast into the mold. The scaffolds were then removed by selective dissolution of the mold in acetone to obtain porous scaffold. The scaffold was subsequently sintered in an oven to burn out the polymeric binder. One key disadvantage of this method was the significant shrinkage after the polymer removal and sintering with a reported shrinkage of up to 22%. Taboas et al. [75] combined conventional techniques such as porogen leaching and phase separation with MM II to create porous scaffolds. In an indirect approach, a negative mold was obtained through MM II and thereafter cast with polymer-ceramic composite. The main difference between this approach and the one by Limpanuphap et al. was the usage of computer-aided software to create a global porous scaffold with pre-determined internal pore architecture and the use of salt particles to create local pores by means of porogen leaching technique to obtain porous scaffolds. In a similar manner, the polymeric composite comprising of PLA and HA together with sieved salt particles were cast into the mold. Using a lost mold technique, the mold removal was done via melting or dissolution to obtain the porous scaffolds. Likewise, fabrication through this technique yielded shrinkage of up to 50% in volume. In addition, both works used organic solvents in the fabrication process. Varying internal channels for the molds were fabricated in the research work by Sachlos et al. [76] in similar manner but using different casting materials. The two main disadvantages in these works were the two-step approach in fabricating the scaffolds and the high shrinkage generated by these scaffolds.

Recent researches have explored the potential of MM II further by developing a similar machine that used the concept of dual print heads, if not, multiple print heads, to "print" organs directly circumventing the needs to have a template or scaffold for the cells to grow and proliferate [77–79]. Biomaterials such as collagen and possibly cells can be introduced into the print heads and the direct printing of organs and the seeding of cells can be carried out simultaneously. However, this method needs to be explored further and is only applicable to soft tissues at this stage and hence may not be suitable for anchorage-dependent cells that often rely on three-dimensional scaffolds to guide cell proliferation. However, the author believes that the potential of this method can be extended to such anchorage dependent cell types.

18.5.2 Powder-Based Techniques

Though powder-based systems are strictly considered as solid-based because they are in solid state, it is intentionally created as a category to highlight the fusion techniques employed to fuse them together which often includes binder or laser. This category is meant to mean powder in grain-like form.

18.5.2.1 Three-Dimensional Printing (3D-P)

The 3D-P technique employs ink jet printing technology for processing powder materials. The versatility and simplicity of 3D-P allows the processing of a wide variety of powder materials including polymers, metals and ceramics. During fabrication, a printer head is used to print a liquid binder onto thin layers of powder following the object's profile as generated by the system computer. The subsequent stacking and printing of material layers to the top of previously printed layer recreates the full structure of the desired object. The completed object is embedded inside a cake of unprocessed powders and is extracted simply by brushing away the loose powders.

3D-P is perhaps the most widely investigated RP technique for scaffold fabrication. Kim et al. [69] employed 3D-P with particulate leaching technique to create porous polylactide-coglycolide (PLGA) scaffolds with interconnected 800-μm porous channels and microporosities of 45–150 μm. The scaffolds were subsequently seeded with hepatocytes (HCs). The successful attachment of a large numbers of HCs to the scaffolds' surfaces and in-growth of HCs into the pore spaces of the scaffolds were reported. Zeltinger et al. [61] successfully built poly(L-lactic acid) (L-PLA) scaffolds with varying pore sizes (38–150 μm) and porosities (75–90%) to investigate the influence of scaffold microstructure on the adhesion, proliferation and matrix deposition of canine dermal fibroblasts, vascular smooth muscle cells and microvascular epithelial cells. Cell culture results demonstrated the suitability of scaffolds built with 3D-P in supporting cell proliferation for the different cell types. Park et al. [70] investigated using 3D-P with surface modification methods to create scaffolds with controllable poly(lactide-co-glycolic)/poly(L-lactide) (PLGA/PLLA) anisotrophic microstructures and

regionally selective scaffolds with diverse receptor systems to enable selective cell adhesion. Lam et al. [54] developed a blend of starch-based powder containing corn-starch (50%), dextran (30%) and gelatin (20%) that can be bound by printing distilled water. However, to improve the strength and water resistance of the scaffolds, a secondary infiltration process with copolymer solution consisting of 75% poly(L-lactide) acid and 25% polycaprolactone in dichloromethane is necessary.

One other application of 3D-P in the fabrication of scaffolds was to print a blend of cellulose and starch powders containing 40 and 50 wt% HA loading. A water-based polymeric solution was employed as the binder in producing the scaffolds. The structures obtained yield high degrees of pore interconnectivity within the scaffolds. Control of the scaffolds' microstructural characteristics can be achieved by varying the printing speed, the flow rate, drop position of the liquid binder and the particle size distribution and composition of the powders. As 3D-P fabrication involves the utilization of a powder bed, a scaffold with complex shapes containing overhanging or hollow features can be carried out since the unprocessed powders surrounding the part can provide the necessary support for such features. The simplicity and versatility of 3D-P allow the processing of a wide range of biomaterials including polymers and ceramics. As 3D-P fabrication is conducted at room temperature, the processing of temperature sensitive materials can be conducted.

Limitations encountered with 3D-P-fabricated scaffolds using the standard building styles provided by the system software include the limited pore sizes achievable ($<50\,\mu m$) and their widely distributed values. More consistent pore sizes including larger pores can be generated by mixing porogens (with pre-determined sizes) into the powders prior to scaffold fabrication [61, 69]. Despite the improvements achievable with porogens, their application carried the risk of inadvertent residues remaining within the scaffolds as a result of incomplete leaching. For producing polymeric scaffolds, the use of organic solvents as binders to dissolve the polymer powders in the printed regions will create difficulties in solvent removal and risks of residues. However, the formulation and application of water based binders and specialized powder blends as described previously may lead to the elimination of solvent application. The use of water-based binders in conjunction with room temperature processing may also allow the incorporation of biologically active and pharmaceutical agents into the scaffolds during fabrication [54, 61]. The mechanical properties and accuracy of 3D-P-fabricated scaffolds are other considerations that need to be further addressed [71].

18.5.2.2 Selective Laser Sintering (SLS)

The SLS technique employs a carbon dioxide, CO_2, laser beam to sinter selectively thin layers of powdered polymeric materials. The powders which are pre-heated to a temperature close to melting point will encounter a rise in temperature beyond melting point when exposed to the laser spot, causing them to melt and fuse together into a solid mass. Subsequent layers of powders that are sintered directly on top of previously formed layers will cause the fusion of adjacent layers to form the height of the build object.

Lee and Barlow [80, 81] developed various forms of poly(methyl methacrylate-*co-n*-butyl methacrylate) (PMMA) coated calcium phosphate powders for processing on a commercial SLS system for producing ceramic scaffolds for hard tissue regeneration. The SLS-fabricated parts were fired in a furnace to remove the polymeric binder and subsequently to sinter the ceramic structures. To improve the density and strengths of the ceramic structures, infiltration of the structures with a phosphoric acid based inorganic cement was carried out before and after the sintering cycle. The final ceramic scaffolds were reported to possess pore sizes ranging from 60 to 200 μm and an overall porosity of about 30%.

Biocompatibility assessment conducted by implanting SLS-fabricated oral implants on the dog model revealed excellent integration of the implant with the host bone after only 4 weeks of implantation.

In the authors' current research [82, 83], an investigation is being conducted to identify and assess the suitability of different biopolymers for processing via the SLS technique. So far, preliminary sintering experiments conducted on a commercialized SLS system, Sinterstation 2500 (3D Systems, Valencia, CA, USA), have ascertained the feasibility of processing poly-L-lactide acid (PLLA), poly-caprolactone (PCL) and polyetheretherketone (PEEK) powder stocks with varying degrees of success. Figure 18.4 illustrates the microstructures of SLS-fabricated scaffold specimens produced from the sintering of PEEK powders and a powder blend of PEEK/HA. Figure 18.5 presents the microstructures obtained from the sintering of PLLA and PCL powders. Due to the use of powder stocks and the relatively low compaction forces exerted on the powder bed during the layering of the powders, high porosity and pore interconnectivity were obtained for all the specimens. The porosity of the specimens can be changed by varying the SLS process parameters such as laser power (or exposure density), scan speed and part bed temperature. Similar to the 3D-P technique, the SLS is highly capable of producing scaffolds with irregular three-dimensional shapes including structures containing channels and overhanging features due to support provided by the powder bed.

(a)   (b)
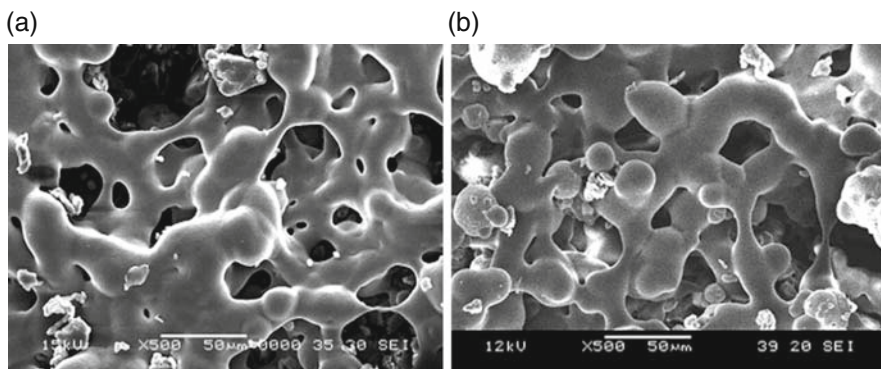
Fig. 18.4 Microstructure of SLS-fabricated scaffolds, **(a)** PEEK and **(b)** PEEK with 10 wt% HA

(a)   (b)
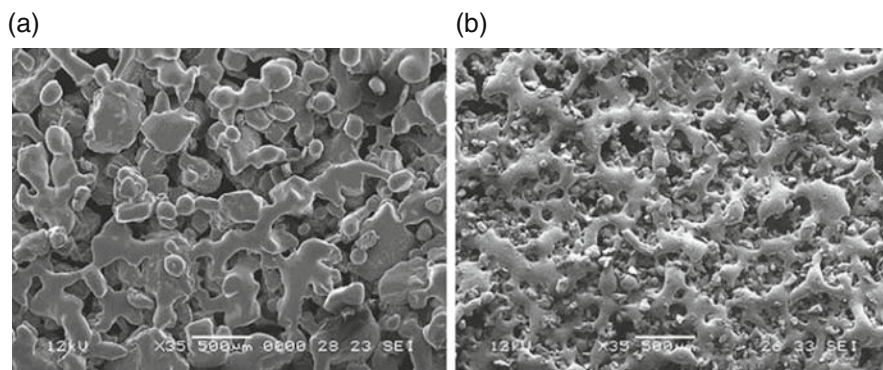
Fig. 18.5 SLS-fabricated scaffolds obtained from the sintering of **(a)** PLLA and **(b)** PCL powders

Other advantages of SLS for polymer processing include the fact that it is solvent free and does not require any secondary binder system, hence, minimizing any risk of material contamination.

Limitations encountered with the SLS technique include the fact that the pores created in SLS-fabricated scaffolds are dependent on the particle size of the powder stock used and the compaction pressure exerted onto the powder bed while depositing the powder layers. As with the scaffolds fabricated by 3D-P, scaffolds fabricated by SLS possess pores with sizes that are distributed over a wide range of values. Since the SLS technique involves high processing temperatures, the technique is limited to the processing of thermally stable polymers.

18.5.3 Liquid-Based Techniques

Liquid-based RP systems have the initial form of its material in liquid form. Through a process known as curing, the liquid is converted into solid state hence the physical part is obtained.

18.5.3.1 Stereolithography Apparatus (SLA)

The SLA process, a liquid based rapid prototyping system, builds parts in a vat of photo curable liquid resin that solidifies upon exposure to laser radiation usually in the UV range [1]. The process begins with the vat filled with the photo curable liquid resin and the elevator table set just below the surface of the liquid resin. The computer-controlled optical scanning system then directs the focused laser beam so that it solidifies the two dimensional cross section corresponding to the slice data on the surface of the resin to a depth greater than one layer thickness with the elevator descending with each layered scanned. The elevator table then descends enough to cover the solid part that is just fabricated with another layer of resin on the surface.

The laser proceeds to draw the next layer and the process is repeated until the whole part is fabricated.

Research on SLA using ceramic materials [84–87] shows the feasibility of using ceramic in SLA systems. Brady et al. [84] and Hinczewski et al. [85] in their research used a ceramic suspension containing alumina powder, UV curable monomer, diluent, photoinitiator and dispersant in the process to produce three-dimensional ceramic parts. Upon polymerization, the photopolymer was subjected to heating to remove the organic phase. Subsequent sintering (by means of heating) was carried out to obtain the final dense ceramic piece. The key requirement in these researches is that the suspension must be UV curable that is required for all stereolithography (SL) processes. Other concerns include the ability to obtain a resin with a low viscosity (1,200 cps or less [88]) and a curing depth sufficient for typical layer thickness of 100 μm [1].

In the need to seek for alternative materials to be used on SLA, various research teams looked into the possibility of using HA [88–90] and polymers such as poly (propylene fumarate), PPF [91]. Chu et al. [88, 89] in their research explored the possibility of using an aqueous HA suspension in acrylamides to fabricate custom made ceramic implants by SL. By using different volume percentage of HA powder and different weight percentage of dispersant, different varying viscosities of the resin were obtained. A higher volume percentage of HA powder would result in a higher viscosity making it impossible to be used for SL. Parts fabricated using ceramic suspensions tended to be non-porous which are not desirable for tissue scaffolds because a porous structure is required for cell proliferation.

Hollister et al. [90] combined an image-based approach (based on Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) data) with SLA to fabricate custom designed TE scaffolds for application in craniofacial surgery. Computer software was used to read the CT data of the area of defect that was to be reconstructed and the data was regenerated to form the external shape. With the shape of the defect obtained, internal pore architecture based on the defect was generated taking into consideration crucial factors such as tissue infiltration and mechanical behavior and was combined via Boolean combination to obtain the image of the scaffold. The image of the scaffold was exported to a format readable by SLA (in this case .stl) and was ready to be fabricated using the two techniques proposed by the team. In the first technique, casting used as HA slurry was poured into the molds that were fabricated by SL and cured. The second technique used by the team is similar to the one carried out by Chu who is also a member of this research team. Both techniques are capable of fabricating scaffolds that allow control over the external shape as well as internal pore architecture.

The presence of double bonds in PPF allows the polymer to be cross-linked into a solid using a thermal initiator making it a suitable material for use in SL for the work carried out by Fisher et al. [91]. Soft and hard tissue response to the scaffolds was investigated upon in a rabbit model and results showed that such scaffolds elicited a mild tissue response in both soft and hard tissue making PFF a viable material for use in TE.

Due to the need to obtain resins that are UV curable, the application of SL on TE is limited. Furthermore, the process involves various other sub-components (such as diluent and dispersant) making it a tedious and complex process if HA powders are to be introduced. This inevitably prolongs the fabrication time because parts fabricated by SL process needs further curing in the UV chamber. The usage of organic solvent also hinders its application for TE scaffolds. Like FDM, SL process would include the need to allocate supports in parts with overhang that need to be removed during the post processing stage.

18.5.3.2 Rapid Freeze Prototyping (RFP)

Rapid Freeze Prototyping (RFP) [92–95], a liquid based RP technique, employs water as its building material. Like most RP techniques, RFP fabricates parts by depositing water droplets layer by layer till the final shape is obtained. The building material is pumped from a reservoir of water to the nozzle and is deposited onto the previously solidified ice surface. The low temperature environment and the previously solidified ice surface allow the deposited water droplets to solidify rapidly and hence stick to the previous layer. In order to ensure that the flow of water is smooth, the nozzle and pipes are maintained at a temperature just above the freezing point of water. Similar to FDM, RFP also employs a support material which in this case is a solvent with a melting point that is lower than water hence allowing the easy removal of supports at a sub zero environment.

As RFP is a relatively new RP technique developed in the late 1990s in a joint effort between Tsinghua University in China and University of Missouri in Rolla (USA), publications in the field of TE with respect to RFP techniques remain limited. Feng [95] in her thesis examined the possibility of building TE scaffolds on RFP by using composite materials comprising of poly (L-lactic) acid (PLLA) and nano-HA-collagen as the materials for scaffolds. In her approach, water was used as supports for creating the scaffolds. Water droplets were deposited onto a flat surface in an alternate direction. Upon forming the first layer of the ice pattern, the composite materials which were dissolved in an organic solvent were extruded into the gaps between each ice pattern and were subsequently milled to form a flat surface before another layer of ice pattern was formed in the opposite direction. Using the same techniques, the process was repeated till a three-dimensional scaffold was obtained. Thereafter, the ice pattern was melted leaving behind a porous three-dimensional structure comprising of PLLA and nano-HA-collagen structure. The thesis reported pore sizes ranging from 500 to 1000 μm which are adequate to promote cell proliferation of bone tissues. Furthermore, pore sizes can be varied by changing the width between the ice lines making RFP advantageous as compared to conventional techniques of building scaffolds whereby consistent pore sizes have always been a concern.

The potential of using RFP in TE is greatly enhanced by the abundant availability of water and the non-toxic effect of water makes it a suitable material in building scaffolds. The low temperature environment where the fabrication process is carried out minimizes degradation or decomposition of the materials in use and

the bioactivity of the materials are not affected. In addition, RFP does not require additional support materials as water itself is used as the supports. One of the disadvantages of this technique lies in the necessity to operate in a low temperature environment making it costly to maintain such a facility. Furthermore, the set-up is difficult to operate as care has to be taken in ensuring the right ratio of the in-nozzle material flow rate to XY scanning speed [92]. If the ratio is too large, excess materials are extruded resulting in parts being distorted and out of proportion. On the other hand, a low ratio would result in inconsistent extrusion of the materials resulting in incomplete forming of the fabricated parts. Like most current RP techniques in use, RFP requires the usage of organic solvent to dissolve PLLA and nano-HA-collagen to obtain a slurry that is used to filled up the gaps within the ice patterns. The usage of organic solvent remains a concern as the work carried out did not further investigate the effect of the organic solvent on tissues. As the structure obtained was highly porous, the lack of mechanical strength limited the scaffolds to low load bearing tissues.

While each RP system has its advantages and disadvantages, most RP systems that are currently in used for TE scaffolds required the usage of organic solvent which is undesirable. Table 18.1 presents the advantages and disadvantages of each technique.

18.6 Development of CAD Strategies and Solutions for Automated Scaffolds Fabrication

Research on RP techniques for scaffold fabrication has ascertained their potential in overcoming many of the limitations encountered using manual-based fabrication techniques. Among the main advantages derivable with RP is the ability to produce spatially and anatomically accurate scaffolds with complex irregular three-dimensional geometries. However, currently available RP systems are optimized for producing highly dense prototypes for engineering applications and as such are limited when it comes to the fabrication of porous structures with consistent pore morphologies and characteristics that are essential for TE scaffolds. Therefore, to enhance better RP scaffold fabrication, new building strategies would have to be developed. One solution is to design the internal micro-architecture of the scaffolds using CAD modeling before committing the scaffold designs to RP fabrication [63, 96, 97]. However, due to the minute and intricate details of possible scaffold micro-architectures, preparation of the computer model of the scaffold using traditional CAD techniques for RP fabrication can be extremely prohibitive even at the hands of experienced CAD operators. The task of having to design, assemble and integrate individual pore geometries to form a consolidated three-dimensional architecture can be highly tedious and time-consuming making the entire process impractical for supporting customized production. The need to shape the assembled micro-architectural design to conform to the anatomic shape of the organ or tissue adds to the difficulty of CAD modeling.

Table 18.1. Advantages and disadvantages of RP techniques

Process	Materials	Advantages	Disadvantages
FDM	Thermoplastic polymers, ceramics	100% interconnected porous structure; no trapped material problem	Requires support structures; high heat effect on raw materials; usage of organic solvents; unable to build complex structures
MM II	Thermoplastic polymers, biodegradable polymers, collagen	Wide range of material choice, able to fabricate porous structure	High shrinkage, dual step fabrication process
SLA	Polymers, ceramics	100% interconnected porous structure; no trapped material problem; capable of producing very fine features	Requires support structure; limited material choice
RFP	Water	Able to vary pore sizes	Needs organic solvent; limited material choice; needs to operate in a low temperature environment
3DP	Polymers, ceramics, starch, metals	100% interconnected porous structure; no heating of powders necessary; wide material choice	Trapped materials in the interior may be difficult to remove; lack of mechanical strength as no chemical bonding is involved
SLS	Polymers, ceramics, metals	100% interconnected porous structure; wide material choice; able to build intricate complex structures	Trapped materials in the interior may be difficult to remove

In order to overcome the complications of the CAD modeling process, several automated image-based approached for scaffold assembly had been reported. In the work carried out by Hollister et al. [63] which is focused on the fabrication of scaffolds for bone tissue engineering in craniofacial reconstructive surgery, a digital model possessing the external shape of the defect is first reconstructed based on surface profile data acquired at the region of interest on the patient using CT or MRI scans. An internal micro-architecture model is then created using voxel density distribution approach. Both the defect model and the micro-architectural models are then combined to create the final scaffold model which is then fabricated on an RP system.

The authors have developed an automated image-based technique [96–98] that can be used to construct and assemble scaffold micro-architectures from a set of selected polyhedral shapes to yield open cellular solids that can be applied for TE applications. Figure 18.6 presents the process chain for the automated process.

Fig. 18.6 Process chain for
automated approach to
scaffold fabrication
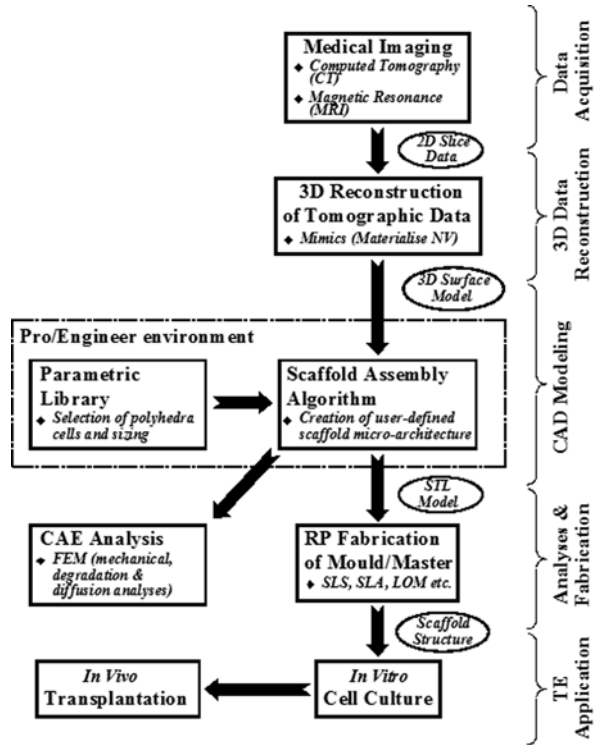
To allow a wide range of scaffold micro-architectures to be generated, a parametric library of open polyhedral unit shapes was first established. With the library, the user is given the option to select the shape of the polyhedral or combinations of polyhedras to be used and to size the selected polyhedras to yield the correct pore sizes. Control is also provided for the user to determine and build model with desired surface area to volume ratio, porosity and pore sizes of the scaffold that can be assembled from the selected polyhedra.

Once an appropriate polyhedra unit shape has been selected and sized, a scaffold assembly algorithm is developed to assemble and link multiple copies of the selected and modified polyhedra shape into a three-dimensional scaffold structure that conforms to the external surface profiles of the defect. Figure 18.7 presents the CAD images of several assembled structures that were generated using the scaffold assembly algorithm and photographs of their physical counterparts that were fabricated using the SLS as compared to a millimeter scale rule. The assemblies shown in Fig. 18.7a,b were obtained by combining two different polyhedra forms using an alternating pattern arrangement while the assemblies shown in Fig. 18.7c,d were assembled from a single polyhedra form. High macro-porosity values ranging from 83% to 93% are achieved with the different cell configurations presented in Fig. 18.7. This is done without encountering any adverse effects on the structural

Fig. 18.7 CAD generated
scaffold assemblies and their
RP fabricated counterparts.**(a)**
Octahedron-tetrahedron. **(b)**
Octahedron-truncated cube.
(c) Square prism. **(d)** Square
pyramid
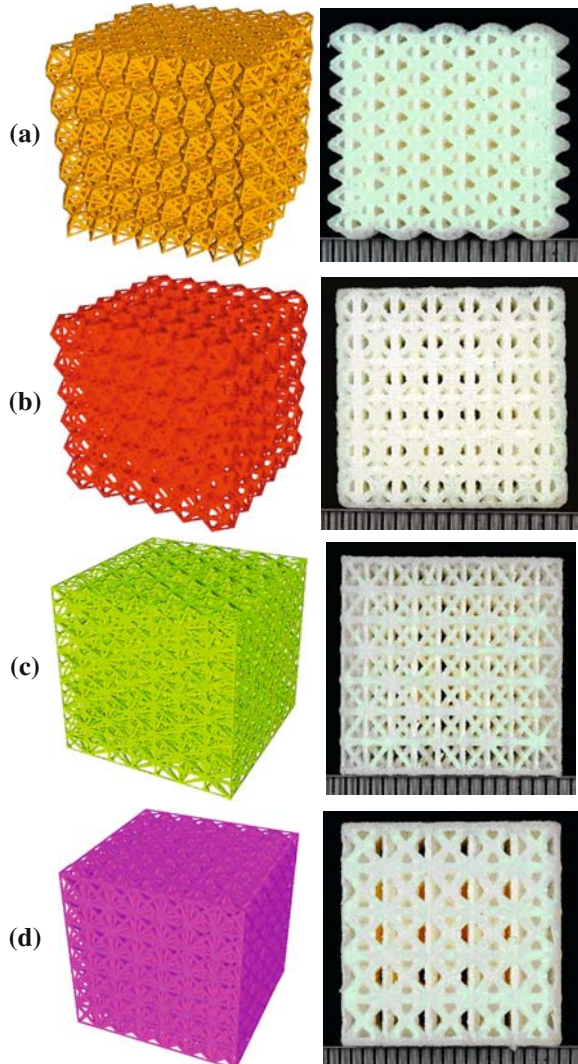
integrity of the RP fabricated structures. Figure 18.8 presents a scaffold structure
that was assembled to conform to the external and internal shape of the bone seg-
ment shown in the same figure.

Besides the benefits of RP fabrication that have been described earlier, the appli-
cation of CAD strategies for designing scaffold micro-architectures will enhance
the capabilities of the RP techniques. With such strategies, a wide variety of inter-
nal architectural designs to facilitate experiments that can be used to characterize
the different micro-architectures can be derived. The results of such experiments

(a)   (b)
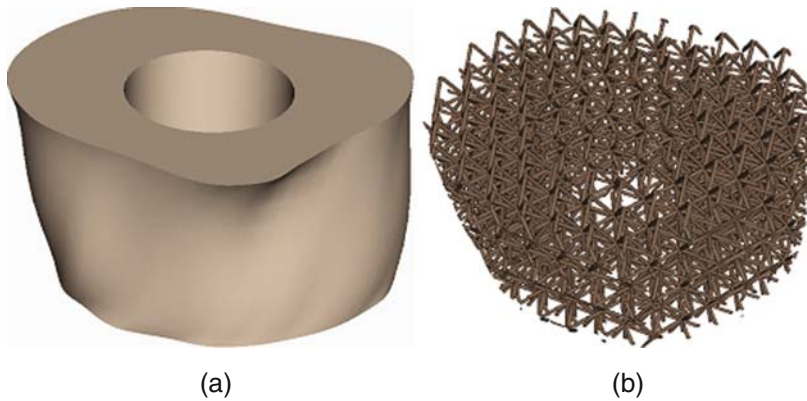
Fig. 18.8 (**a**) Surface model for bone segment reconstructed from CT data.(**b**) CAD-generated scaffold structure for the bone segment

are necessary inputs for determining the optimal parameters to be used in the RP process and cell culture. The wide variety of internal micro-architectures will also facilitate in vitro and in vivo experiments to determine the suitability and to optimize the designs of the various micro-architectures for different TE applications.

18.7 Prostheses

Prostheses are man made substitutes fabricated for body parts that have been removed either for traumatic or medical reasons. The main aim of prosthesis is to help a person appear as if the body part has not been removed, hence restoring the confidence of the person in their social lives. While there are numerous prostheses such as hip, artificial legs, breasts or facial, this section will only explore the applications of RP techniques on facial prostheses.

Facial deformities due to acquired (e.g., surgical excision of diseased tissues or trauma) or congenital (e.g., microtia) causes have always been a concern for plastic surgeons. Patients with defects in the facial region may often be subjected to ridicule or self-isolation and very often prostheses would be required to provide psychological benefit in the rehabilitation of the patient. With the advancements in modern surgical procedures, instruments and equipment, plastic surgery may improve a defect in the facial region, but in some cases, it may not be possible to create the lost anatomy with sufficient detail. In addition, patients who do not have the option of surgery due to age, medical condition and preference of the patients, or in cases whereby surgery would not yield satisfactory results, facial prostheses seem to provide the best esthetical solution.

In this section, an integrated manufacturing system for producing extraoral facial prostheses utilizing laser surface scanning/digitizing, computer-aided design (CAD) (e.g., three-dimensional modeling) and computer-aided manufacturing (CAM)

(e.g., rapid prototyping, RP) technologies is introduced [99–102]. Two case studies involving the making of different facial prostheses (forehead and auricular prostheses) using the integrated manufacturing system are briefly described.

18.7.1 Integrated Approach to Prostheses Production

The integrated approach encompasses a series of steps (Fig. 18.9) that include (1) data acquisition, (2) CAD remodeling, (3) fabrication of prosthesis via RP and (4) trial fitting and casting of actual prosthesis.

18.7.1.1 Data Acquisition

The first step in the integrated approach is to obtain topographic data of both the defective region and a healthy "donor" organ upon which the design of the prosthesis would be based. In order to capture anatomic details and profiles, surface scanning is carried out by means of a laser surface scanners or digitizers. Surface scanning is being used more frequently now in the medical field and its usefulness is evident in studies related to data capture of surface morphology and quantitative assessment of surface changes. Data obtained from such surface scanners are three-dimensional point cloud data sets stored in various formats (e.g. IGES, DXF, ACIS, etc.). Such data are read by specialized CAD software so as to reconstruct the surfaces of the
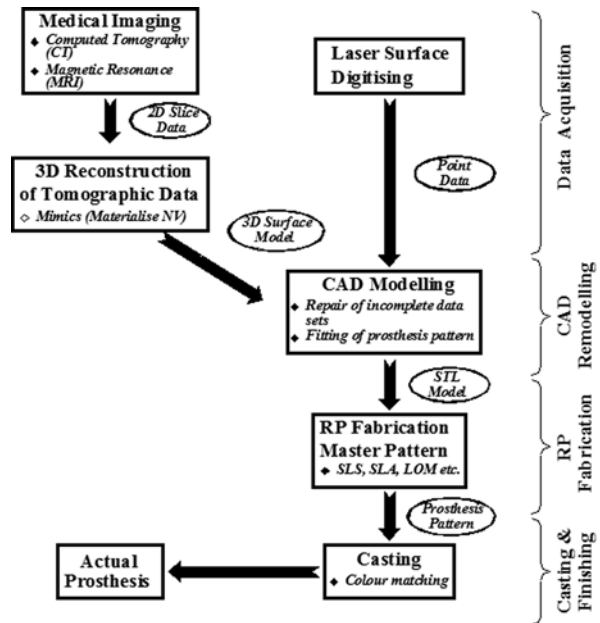
Fig. 18.9 Process chains of the new integrated manufacturing system

scanned object from the output digitized data to yield a computer model of the object that can then be exported to the relevant RP systems. One key advantage of such scanners is that the digitizers utilize low power lasers that produce no harmful radiations. Moreover, scanning times are in the order of 0.6–7 s per scan which eliminates the need for the patient to remain perfectly still for a lengthy period of time. However, one of the limitations with most surface scanners is the need for special darkroom environment or special rooms with the scanning components forming permanent fixtures in these environments.

18.7.1.2 CAD Remodeling

CAD software is often utilized to edit and manipulate the point cloud data generated by a digitizer. The ability of such software to convert CAD models to STL data format is important since the STL format is the de facto standard for data interfacing between CAD and RP systems. Although most CAD software are capable of generating STL data, not all are capable of reading-in, editing or manipulating STL data. As such, the authors have used Surfacer Version 9 (Imageware Inc., USA) for the purpose of their research. The point cloud data is triangulated or polygonized to reconstruct the surface profiles of the site of deformity or the facial organ. Any missing data patches or blind spots on the polygonized point cloud model can be regenerated using a number of CAD editing and data manipulation methods. The computer models generated by both techniques are converted into STL data for fabrication in an RP system.

18.7.1.3 Fabrication of Prosthesis Via RP

Most RP systems generally adopt the additive fabrication technique whereby the part to be fabricated is built layer-by-layer. The individual layer is represented and determined by its corresponding slice file that stores the geometric detailed data of the model to be built. For the purpose of discussion in this section, Selective Laser Sintering (SLS) would be described.

18.7.1.4 Casting of Actual Prosthesis

Casting of the prostheses starts from the master model of the prototype. Vacuum casting is often preferred in prototype making because low cost components can be obtained. Furthermore, slight adjustments to their shapes which can be achieved by removing materials or adding wax where desired, are conducted to improve the comfort of the prostheses. The RP-replicas can also serve as patterns to produce the negative molds necessary to cast the actual prostheses.

18.8 Case Studies

The authors have focused on producing facial prostheses using computerized imaging techniques, CAD and RP on facial organs (such as ear) and deformity (forehead) in their research [101, 102].

18.8.1 Case Study 1: Prosthetic Ear

In this case study, the patient suffers from unilateral microtia involving the total loss of the right external ear. As such, the profile of the deformity as well as the patient's contralateral external normal ear were scanned using the laser digitizer. For the purpose of comparison between the healthy ear and the prosthetic ear that is to be created, the anterolateral and posterior surfaces of both sides of the normal ear were scanned. Figure 18.10a shows the point cloud data of the healthy ear as acquired by the digitizer and Fig. 18.10b illustrates the application of the three-dimensional curve generation technique in Surfacer Version 9.0 environment to restore missing areas that were not captured during scanning. Figure 18.10c shows the fully patched
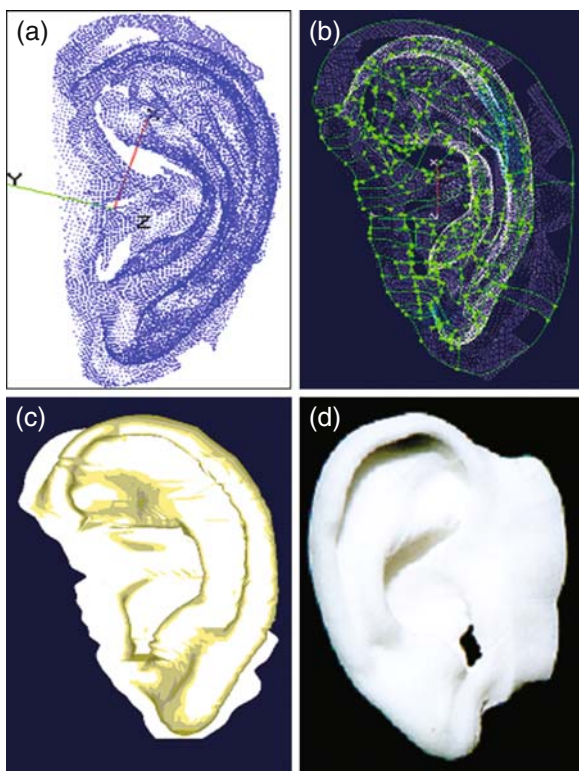
Fig. 18.10 a Point cloud data of the healthy ear. **b** Surface regeneration using three-dimensional curve generating technique. **c** Completed CAD model (before mirror-imaging). **d** SLS-fabricated pattern

and completed computer model of the ear. Upon the completion for the patching of the point cloud data in the Surfacer Version 9.0 environment, the completed model was mirror-imaged to obtain a contralateral version, which provided for the prosthetic ear pattern to be fabricated. A point to note is that, in order to create an accurate prosthetic ear with good surface fitting attributes, the positioning and fitting of the prosthesis was conducted in the CAD environment. Factors taken into consideration in creating a natural looking prosthesis include the inclination of the ear, the size and the degree of erection. Moreover, to ensure perfect fitting between the under-surface of the prosthesis and the face of the patient, the surface profile of the site of deformity was extracted and used to create the under-surface of the prosthetic ear model. When a satisfactory model of the prosthetic ear was reached, it was converted into the STL format and fabricated using the SLS system.

Figure 18.10d presents the prosthetic ear pattern built from polyamide (Duraform^{TM}, 3D Systems Inc., USA) material. It was observed that the undercuts around the helix and concha regions of the ear were reproduced with high degrees of realism using the surface patching technique. The duration of the entire process from data acquisition to completion of the pattern was approximately 5.5 h including the 3-h SLS fabrication time.

18.8.2 Case Study 2: Prosthetic Forehead

For this case study, the patient has a congenital defect in the form of a depression on her forehead. The objective was to create a forehead prosthesis that can be worn by the patient to cover the depression. Conventional procedures for such deformity include facial implants or local flap [103, 104]. The first technique, facial implants, requires the raising of skin flap down to the skull bone. Thereafter, a pocket is created and the implant, usually made of silastic or methylmethacrylate, is embedded into the pocket. The implant is anchored in place by placing sutures around the enveloping muscles and tissues. After ensuring that the implant is secured, the skin flap is returned to its original position. In the second technique, areas such as the supraorbital and supratrochlear vessel-based island flap are used to cover the unilateral forehead defects. Detailed pre-operative dimensioning is needed to determine the size and shape of the forehead skin to be replaced by portion of the forehead island flap. Full thickness of skin is incised around the inverted pattern of the defect, outlining the skin portion of the island flap [104]. Such conventional techniques are often dependable on the condition of the patient as well as the availability of donor site for such operations. Furthermore, such surgical procedures often require up to 3 h. Compared to the conventional procedures, RP fabricated prosthetic forehead is non-invasive hence reducing the duration of the surgical procedures and recovery time for the patient.

This particular prosthesis presents a challenge to the CAD operator as the profile of the forehead prosthesis had to be empirically recreated based on anthropometrical measurements and visual assessments made at the site of deformity. A full-face

scan was conducted on the patient using a hand-held laser scanner to obtain the full profile of the depression. The scanned data was then imported into Surfacer software environment where it was triangulated to generate the surface model of the patient's face. Figure 18.11a shows the surface model generated using Surfacer with the congenital forehead depression visible on the patient's left.

The facial profile on the unaffected side of the patient's forehead was extracted and mirror-imaged to serve as a template for the CAD operator to capture and recreate the profile of the missing forehead. Using the CAD tools in Surfacer, the forehead prosthesis was created based on the boundaries as defined in Fig. 18.11b. Surface lofting was then used to generate a series of external surfaces as well as the under surfaces of the forehead prosthesis based on the set of defined curves which can be further adjusted to match the profile of the forehead depression. The realism
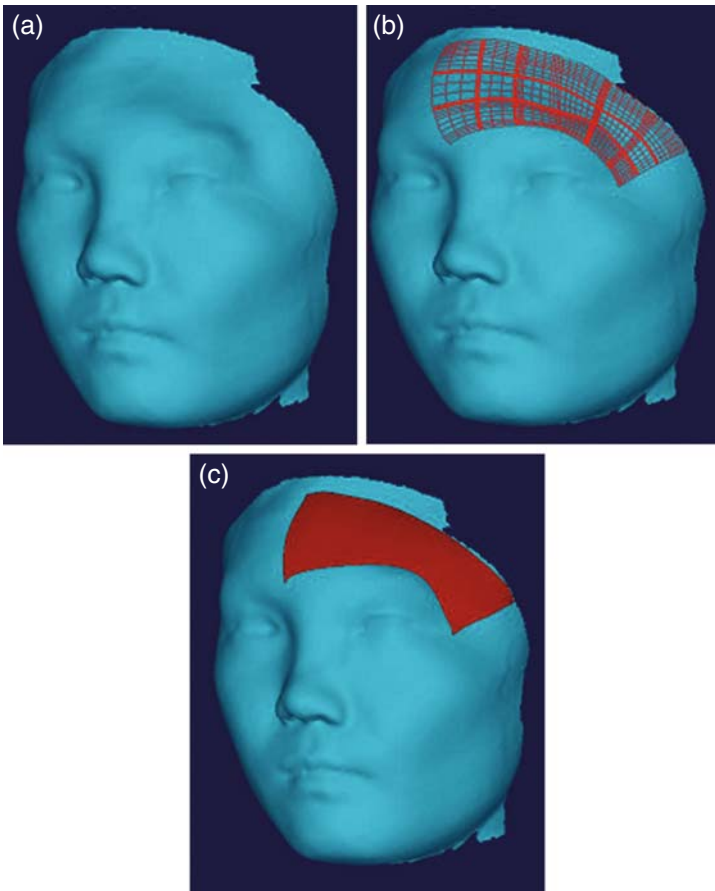
Fig. 18.11 a Subject with congenital forehead depression. **b** Creating the forehead prosthesis via three-dimensional surface patching. **c** Model of forehead prosthesis fitted onto the patient's face

of the prosthesis was enhanced with the aid of the unaffected side of the patient's forehead as the external surface contours of the prosthesis were adjusted to match the curvature of the unaffected side.

This was made possible in the Surfacer environment as the prosthetic model can be viewed from different projection angles to assess its fitting and to assist in carrying out editing works to smoothen or camouflage the junction lines formed at the edge of the prosthesis when it is worn by the patient. This was carried out to create the illusion of a continuous forehead surface. By doing so, the edges of the resulting prosthesis model are usually thin in order for the prosthesis to blend flawlessly onto the face of the patient. Figure 18.11c shows the completed model of the prosthesis fitted onto the patient's face.

As thin edges were noted on the prosthetic forehead, the direct fabrication of the prosthetic forehead would result in a RP-fabricated pattern with edges that were brittle and hence easily damaged through handling. Hence, instead of creating a master pattern to cast a silicon rubber mold to produce the actual prosthesis, the direct mold production approach was followed to produce a negative RP-fabricated mold set for casting the forehead prosthesis. Fabrication of the mold set was carried out on a SLS system (Sinterstation 2500, 3D Systems Inc., USA) to yield a negative mold that can be cast using silastic material to obtain the prosthetic forehead.

18.9 Conclusion

The biomedical applications of RP in two main fields have been explored. The applications of scaffolds fabricated using conventional, manual-based fabrication techniques are limited due to imperfections encountered in their macro- and micro-architectures and the poor flexibility and control offered by the fabrication techniques. Computer-controlled fabrication via RP technology has been recognized and established as having the potential for providing a generic solution in automating customized scaffold production. A review of the different RP methods that have been evaluated and employed for scaffold fabrication together with the work conducted by different research teams worldwide has been presented. The use of image-based approaches and CAD strategies for designing scaffold structures also has been highlighted and discussed. It is clear that the application of RP techniques in conjunction with the proposed building strategies will enable new cost-effective and rapid solutions to customize 'made-to-order' scaffold production to meet the needs of most TE applications.

Furthermore, the integration of a manufacturing system consisting of a laser digitizer, CAD software and RP system for making facial prostheses was detailed. From the case studies presented, it can be noted that the novel integrated manufacturing system is able to shorten the time needed to acquire digitized data of a patient's facial deformity, as well as to minimize patient involvement throughout the prosthesis fabrication process. The application of computer-aided techniques will undoubtedly allow reproducible and consistently high quality prostheses to be achieved.

References

1. Chua CK, Leong KF, and Lim CS. Rapid Prototyping – Principles and Applications, 2nd edn, World Scientific: Singapore, 2003.
2. Bell E. Tissue engineering in perspective, in Lanza RP, Langer RS, and Vacanti JP, eds. Principles of Tissue Engineering, 2nd edn, Academic Press: San Diego, 2000.
3. Mooney, DJ and Mikos AG. Growing New Organs, Scientific American, April, 1999.
4. Patrick CW Jr, Mikos AG, and Mcintire LV. Prospectus of tissue engineering, in Patrick CW Jr, Mikos AG and Mcintire LV, eds. Frontiers in Tissue Engineering, Pergamon, New York, 1998, pp. 1–11.
5. Thomson RC, Wake MC, Yaszemski MJ, and Mikos AG. Biodegradable polymer scaffolds to regenerate organs. Adv Polym Sci, 1995, 122: 245–274.
6. United Network for Organ Sharing Website: http://www.unos.org, 2004.
7. Scientific Registry of Transplant Recipients. Annual Report of the U.S. Scientific Registry for Organ Transplantation and the Organ Procurement and Transplant Network, Richmond, VA, 2000.
8. Whang K, Healy KE, Elenz DR, Nam EK, Tsai DC, Thomas CH, Nuber GW, Glorieux FH, Travers R, and Sprague SM. Engineering bone regeneration with bioresorbable scaffolds with novel microarchitecture. Tissue Eng, 1999, 5: 35–51.
9. Kim WS, Vacanti JP, Cima L, Mooney D, Upton J, Puelacher WC, and Vacanti CA. Cartilage engineered in predetermined shapes employing cell transplantation on biodegradable polymers. Plast Reconst Surg, 1994, 94: 233–237.
10. Davies MW and Vacanti JP. Toward development of an implantable tissue engineered liver. Biomaterials, 1996, 17: 365–372.
11. Niklason LE, Gao J, Abbott WM, Hirschi K, Houser S, Marini R, and Langer R. Functional arteries grown in vitro. Science, 1999, 284: 489–493.
12. Karmiol S. Cell isolation and selection, in Atala A. and Lanza RP, eds. Methods of Tissue Engineering, Academic Press: San Diego, 2002, pp. 19–36.
13. Freed LE, Hollander AP, Martin I, Barry JR, Langer R, and Vunjak-Novakovic G. Chondrogenesis in a cell-polymer-bioreactor system. Exp Cell Res, 1998, 240: 58–65.
14. Freed LE, Vunjak-Novakovic G, and Langer R. Cultivation of cell-polymer cartilage implants in a bioreactor. J Cell Biochem, 1993, 51: 257–264.
15. Weinberg CB and Bell E. A blood vessel model constructed from collagen and cultured vascular cells. Science, 1986, 231: 397–400.
16. Madihally SV and Matthew HWT. Porous chitosan scaffolds for tissue engineering. Biomaterials, 1999, 20: 1133–1142.
17. Mikos AG and Temenoff JS. Formation of highly porous biodegradable scaffolds for tissue engineering. J Biotechnol, 2000, 3: 114–119.
18. Temenoff JS and Mikos AG. Review: tissue engineering for regeneration of articular cartilage. Biomaterials, 2000, 21: 431–440.
19. Mooney DJ and Vacanti JP. Tissue engineering using cells and synthetic polymers. Transplant Rev, 1993, 7: 153–162.
20. Cima LG, Vacanti JP, Vacanti C, Ingber DE, Mooney D, and Langer R. Tissue engineering by cell transplantation using degradable polymer substrates. J Biomed Eng, 1991, 113: 143.
21. Cima LG, Langer R, and Vacanti JP. Polymers for tissue and organ culture. J Bioac Compat Polym, 1991, 6: 232–240.
22. Cornell CN. Osteoconductive materials and their role as substitutes for autogeneous bone grafts. Orthop Clin North Am, 1999, 30: 591–598.
23. Beruto DT, Mezzasalma SA, Capurro M, Botter R, and Cirillo P. Use of α-tricalcium phosphate (TCP) as powders and as an aqueous dispersion to modify processing, microstructure, and mechanical properties of poly-methylmethacrylate (PMMA) bone cements and to produce bone-substitute compounds. J Biomed Mater Res, 2000, 49: 498–505.

24. Liu X and Ma PX, Polymeric scaffolds for bone tissue engineering, Ann Biomed Eng, 2004, 32: 477–486.

25. Ingber DE. Mechanical and chemical determinants of tissue development, in Lanza RP, Langer RS, and Vacanti JP, eds. Principles of Tissue Engineering, 2nd edn, Academic Press: San Diego, 2000, pp. 101–110.

26. Mooney DJ, Cima LG, Langer R, Johnson L, Hansen LK, Ingber DE, and Vacanti JP. Principles of tissue engineering and reconstruction using polymer-cell constructs. Mater Res Soc Symp Proc, 1992, 252: 345.

27. Chen CS, Mrksich M, Huang S, Whitesides GM, and Ingber DE. Geometric control of cell life and death. Science, 1997, 276: 1425–1428.

28. McClary KB, Ugarova T, and Grainger DW. Modulating fibroblast adhesion, spreading and proliferation using self-assembled monolayer films of alkylthiolates on gold. J Biomed Mater Res, 2000, 50: 428–439.

29. Freed LE, Vunjak-Novakovic G, Biron RJ, Eagles DB, Lesnoy DC, Barlow SK, and Langer R. Biodegradable polymer scaffolds for tissue engineering. Biotechnology, 1994, 12: 689–693.

30. Freed LE, Grande DA, Lingbin Z, Emmanual J, Marquis JC, and Langer R. Joint resurfacing using allograft chondrocytes and synthetic biodegradable polymer scaffolds. J Biomed Mater Res, 1994, 28: 891–899.

31. Mikos AG, Sarakinos G, Lyman MD, Ingber DE, Vacanti JP, and Langer R. Prevascularization of porous biodegradable polymers. Biotechnol Bioeng, 1993, 42: 716–723.

32. Wake MC, Patrick CW Jr, and Mikos AG. Pore morphology effects on the fibrovascular tissue growth in porous polymer substrates. Cell Transplant, 1994, 3: 339–343.

33. Rout PGJ, Tarrant SF, Frame JW, and Davies JE. Interaction between primary bone cell cultures and biomaterials. Part 3: A comparison of dense and macroporous hydroxyapatite, in Pizzoferratto A, Ravaglioli PG, and Lee AJC, eds. Bioceramics and Clinical Applications, Elsevier: Amsterdam, 1988, pp. 591–596.

34. Yang SF, Leong KF, Du ZH, and Chua CK. The design of scaffolds for use in tissue engineering: Part 1-Traditional factors. Tissue Eng, 2001, 7: 679–690.

35. Thomson RC, Shung AK, Yaszemski MJ, and Mikos AG. Polymer scaffold processing, in Lanza RP, Langer R, and Vacanti J, eds. Principles of Tissue Engineering, 2nd edn, Academic Press: San Diego:, 2000, pp. 251–262.

36. Widmer MS and Mikos AG. Fabrication of biodegradable polymer scaffolds for tissue engineering, in Patrick CW Jr, Mikos AG, and McIntire LV, eds,. Frontiers in Tissue Engineering, Elsevier Sciences: New York, 1998, pp. 107–120.

37. Lu L and Mikos AG. The importance of new processing techniques in tissue engineering, MRS Bull, 1996, 21: 28–32.

38. Mikos AG, Bao Y, Cima LG, Ingber DE, Vacanti JP, and Langer R. Preparation of poly(glycolic acid) bonded fiber structures for cell attachment and transplantation. J Biomed Mater Res, 1993, 27: 183–189.

39. Lo H, Ponticiello MS, and Leong KW. Fabrication of controlled release biodegradable foams by phase separation. Tissue Eng , 1995, 1: 15–28.

40. Nam YS and Park TG. Porous biodegradable polymeric scaffolds prepared by thermally induced phase separation. Biomaterials, 1999, 20: 1783–1790.

41. Mikos AG, Thorsen AJ, Czerwonka LA, Bao Y, Langer R, Winslow DN, and Vacanti JP. Preparation and characterization of poly(L-lactic acid) foams. Polymer, 1994, 35: 1068–1077.

42. Mikos AG, Sarakinos G, Leite SM, Vacanti JP, and Langer R. Laminated three-dimensional biodegradable foams for use in tissue engineering. Biomaterials, 1993, 14: 323–330.

43. Mikos AG, Sarakinos G, Vacanti JP, Langer RS, and Cima LG. Biocompatible polymer membranes and methods of preparation of three-dimensional membrane structures. US Patent 5,514,378, 1996.

44. Thomson RC, Yaszemski MJ, Power JM, and Mikos AG. Fabrication of biodegradable polymer scaffolds to engineer trabecular bone. J Biomater Sci Polym Ed, 1995, 7: 23–38.

45. Mooney DJ, Baldwin DF, Suh NP, Vacanti JP, and Langer R. Novel approach to fabricate porous sponges of poly(D,L-lactic-co-glycolic acid) without the use of organic solvents, Biomaterials, 1996, 17: 1417–1422.

46. Baldwin DF, Shimbo M, and Suh NP. The role of gas dissolution and induced crystallization during microcellular polymer processing: a study of poly(ethylene terephthalate) and carbon dioxide systems. J Eng Mater Technol, 1995, 117: 62.

47. Shastri VP, Martin I, and Langer R. Macroporous polymer foams by hydrocarbon templating. Proc Natl Acad Sci U S A, 2000, 97: 1970–1975.

48. Whang K, Thomas CK, Nuber G, and Healy KE. A novel method to fabricate bioabsorbable scaffolds. Polymer, 1995, 36: 837–842.

49. Healy KE, Whang K, and Thomas CH. Method of fabricating emulsion freeze-dried scaffold bodies and resulting products. U.S., Patent 5,723,508, 1998.

50. Harris LD, Kim BS, and Mooney DJ. Open pore biodegradable matrices formed with gas foaming. J Biomed Mater Res, 1998, 42: 396–402.

51. Yang SF, Leong KF, Du ZH, and Chua CK. The design of scaffolds for use in tissue engineering: Part 2- Rapid prototyping techniques. Tissue Eng, 2002, 8: 1–12.

52. Ang TH, Sultana FSA, Hutmacher DW, Wong YS, Fuh JYH, Mo XM, Loh HT, Burdet E, and Teoh SH. Fabrication of 3D chitosan-hydroxyapatite scaffolds using a robotic dispensing system. Mater Sci Eng C, 2002, 20: 35–42.

53. Fisher JP, Vehof JWM, Dean D, van der Waerden, JPCM, Holland TA, Mikos AG, and Jansen JA. Soft and hard tissue response to photocrosslinked poly(propylene fumarate) scaffolds in a rabbit model. J Biomed Mater Res, 2002, 59: 547–556.

54. Lam CXF, Mo XM, Teoh SH, and Hutmacher DW. Scaffold development using 3D printing with a starch-based polymer. Mater Sci Eng C, 2002, 20: 49–56.

55. Too MH, Leong KF, Chua CK, Du ZH, Yang SF, Cheah CM, and Ho SL. Investigation of 3D non-random porous structures by fused deposition modeling. Int J Adv Manuf Technol, 2002, 19: 217–223.

56. Zein I, Hutmacher DW, Tan KC, Teoh SH. Fused deposition modeling of novel scaffold architectures for tissue engineering applications. Biomaterials, 2002, 23: 1169–1185.

57. Hollister SJ, Maddox RD, and Taboas JM. Optimal design and fabrication of scaffolds to mimic tissue properties and satisfy biological constraints. Biomaterials, 2002, 23: 4095–4103.

58. Chu T-M, Halloran JW, Hollister SJ, and Feinberg SE. Hydroxyapatite implants with designed internal architecture. J Mater Sci Mater Med, 2001, 12: 471–478.

59. Hutmacher DW, Schantz T, Zein I, Ng KW, Teoh SH, and Tan KC. Mechanical properties and cell cultural response of polycaprolactone scaffolds designed and fabricated via fused deposition modeling. J Biomed Mater Res, 2001, 55: 203–216.

60. Xiong Z, Yan YN, Zhang RJ, and Sun L. Fabrication of porous(L-lactic acid) scaffolds for bone tissue engineering via precise extrusion. Scripta Materialia, 2001, 45: 773–779.

61. Zeltinger J, Sheerwood JK, Graham DM, Mueller R, and Griffith LG. Effects of pore size and void fraction on cellular adhesion, proliferation, and matrix deposition. Tissue Eng, 2001, 7: 557–572.

62. Cornejo IA, McNulty TF, Lee S, Bianchi E, Danforth SC, and Safari A. Development of bioceramic tissue scaffolds via fused deposition of ceramics, in George L, ed., Bioceramics: Materials and Applications III, American Ceramic Society: Westerville, 2000, pp. 183–195.

63. Hollister SJ, Levy RA, Chu T-M, Halloran JW, and Feinberg SE. An image-based approach for designing and manufacturing craniofacial scaffolds. Int J Oral Maxillofac Surg, 2000, 29: 67–71.

64. Hutmacher DW. Scaffolds in tissue engineering bone and cartilage. Biomaterials, 2000, 21: 2529–2543.

65. Hattiangadi A and Bandyopadhyay A. Processing, characterization and modeling of non-random porous ceramic structures, Proceedings of Solid Freeform Fabrication Symposium, Austin, TX, 1999, pp. 319–326.

66. Steidle C, Klosterman D, Chartoff R, Graves G, and Osborne N, Automated Fabrication of Custom Bone Implants Using Rapid Prototyping, Proceedings of the 44th International SAMPE Symposium and Exhibition, Long Beach, CA, May, 1999.

67. Bose S, Avila M, and Bandyopadhyay A. Processing of bioceramic implants via fused deposition process. Proceedings of Solid Freeform Fabrication Symposium, Austin, TX, 1998, pp. 629–636.

68. Danforth SC, Safari A, and Langrana N. Solid freeform fabrication of functional advanced ceramic components. Naval Research Reviews, 1998, l: 27.

69. Kim SS, Utsunomiya H, Koski JA, Wu BM, Cima MJ, Sohn J, Mukai K, Griffith LG, and Vacanti JP. Survival and function of hepatocytes on a novel three-dimensional synthetic biodegradable polymer scaffold with an intrinsic network of channels, Ann Surg, 1998, 228: 8–13.

70. Park A, Wu B, and Griffith LG. Integration of surface modification and 3D fabrication techniques to prepare patterned poly(L-lactide) substrates allowing regionally selective cell adhesion. J Biomater Sci Polym Ed, 1998, 9: 89–110.

71. Giordano RA, Wu BM, Borland SW, Cima LG, Sachs EM, and Cima MJ. Mechanical properties of dense polylactic acid structures fabricated by three-dimensional printing. J Biomater Sci Polym Ed, 1996, 8: 63–73.

72. Xiong Z, Yan Y, Wang S, Zhang R, and Zhang C. Fabrication of porous scaffolds for bone tissue engineering via low-temperature deposition. Scripta Materialia, 2002, 46: 771–776.

73. Bose S, Darsell J, Hosick HL, Yang L, Sarkar DK, and Bandyopadhyay A. Processing and characterization of porous alumina scaffolds. J Mater Sci Mater Med, 2002, 13: 23–28.

74. Limpanuphap S and Derby B. Manufacture of biomaterials by novel printing process. J Mater Sci: Mater Med, 2002, 13: 1163–1166.

75. Taboas JM, Maddox RD, Krebsbach PH, and Hollister SJ. Indirect solid free form fabrication of local and global porous biomimetic and composite 3D polymer-ceramic scaffolds. Biomaterials, 2003, 24: 181–194.

76. Sachlos E, Reis N, Ainsley C, Derby B, and Czernuszka JT. Novel collagen scaffolds with predefined internal morphology made by solid freeform fabrication. Biomaterials, 2003, 24: 1487–1497.

77. Wilson WCJ and Boland T. Cell and organ printing 1: protein and cell printers. Anat Rec A, 2003, 272A: 491–496.

78. Boland T, Mironov V, Gutowska A, Roth EA, and Markwald RR. Cell and organ printing 2: Fusion of cell aggregates in three-dimensional gels. Anat Rec A, 2003, 272A: 497–502.

79. Mironov V, Boland T, Trusk T, Forgacs G, and Markwald RR. Organ printing: computer-aided jet-based 3D tissue engineering. Trends Biotechnol, 2003, 21: 157–161.

80. Lee GH and Barlow JW. Selective laser sintering of calcium phosphate powders. Proceedings of the Solid Freeform Fabrication Symposium, Austin, TX, 1994, pp. 191–197.

81. Lee GH and Barlow JW. Selective laser sintering of bioceramic materials for implants. Proceedings of the Solid Freeform Fabrication Symposium, Austin, TX, 1993, pp. 376–380.

82. Tan KH, Chua CK, Leong KF, Cheah CM, Cheang P, Abu Bakar MS, and Cha SW. Scaffold development using selective laser sintering of polyetheretherketone-hydroxyapatite biocomposite blends. Biomaterials, 2003, 24: 3115–3123.

83. Tan KH, Chua CK, Leong KF, Cheah CM, Gui WS, Tan WS, and Wiria FE. Selective laser sintering of biocompatible polymers for applications in tissue engineering, BioMedical Materials and Engineering, United States, 2005, Vol 15 (1–2), pp 113–124.

84. Brady GA and Halloran JW. Stereolithography of Ceramic Suspensions. Rapid Prototyping J, 1997, 3: 61–65.

85. Hinczewski C, Corbel S, and Chartier T. Stereolithography for the fabrication of ceramic three-dimensional parts. Rapid Prototyping J, 1998, 4: 104–111.

86. Brady GA, Chu TM, and Halloran JW. Curing behavior of ceramic resin for stereolithography. Proceedings of Solid Freeform Fabrication Symposium, Austin TX August 12–14, 1996, pp. 403–410.

87. Himmer, T, Nakagawa T, and Noguchi H. Stereolithography of ceramics. Proceedings of Solid Freeform Fabrication Symposium, Austin TX August 11–13th, 1997, pp. 363–369.

88. Chu TMG, Halloran JW, Hollister SJ, and Feinberg SE. Hydroxyapatite implants with designed internal architecture. J Mater Sci: Mater Med, 2001, 12: 471–478.

89. Chu TMG, Orton DG, Hollister SJ, Feinberg SE, and Halloran JW. Mechanical and in-vivo performance of hydroxyapatite implants with controlled architectures. Biomaterials, 2002, 23: 1283–1293.

90. Hollister SJ, Levy RA, Chu TM, Halloran JW, and Feinberg SE. An image-based approach for designing and manufacturing craniofacial scaffolds. Int J Oral Maxillofac Surg, 2000, 29: 67–71.

91. Fisher JP, Vehof JWM, Dean D, Waerden JPC, Holland TA, Mikos AG, and Jansen JA. Soft and hard tissue response to photocrosslinked poly (propylene fumarate) scaffolds in a rabbit model. J Biomed Mater Res, 2002, 59: 547–556.

92. Zhang W, Leu MC, Ji Z, and Yan Y. Rapid freezing prototyping with water. Proceedings of Solid Freeform Fabrication Symposium, Austin TX August 9–11, 1998, pp. 185–192.

93. Sui G, Zhang W, and Leu MC. Study on water deposit in rapid freeze prototyping process. Proceedings of Solid Freeform Fabrication Symposium, Austin TX August 7–9, 2000, pp. 342–349.

94. Zhang W, Leu MC, Feng C, Ren R, Zhang R, Lu Q, Jiang J, and Yan Y. Investment casting with ice patterns made by rapid freeze prototyping. Proceedings of Solid Freeform Fabrication Symposium, Austin TX August 7–9, 2000, pp. 66–72.

95. Feng C. Study of rapid ice prototype forming, PhD Thesis, Tsinghua University, Beijing, China, 2001.

96. Chua CK, Leong KF, Cheah CM, and Chua SW. Development of a tissue engineering scaffold structure library for rapid prototyping: Part 1. Investigation and classification. Int J Adv Manuf Technol, 2003, 21: 291–301.

97. Chua CK, Leong KF, Cheah CM, and Chua SW. Development of a tissue engineering scaffold structure library for rapid prototyping: Part 2. Parametric library and assembly algorithm. Int J Adv Manuf Technol, 2003, 21: 302–312.

98. Cheah CM, Chua CK, Leong KF, Cheong CH, and Naing MW. An automatic algorithm for generating complex polyhedral scaffolds structures for tissue engineering. Tissue Eng, 2004, 10: 595–610.

99. Chua CK, Chou SM, Lin SC, Lee ST, and Saw CA. Facial prosthetic model fabrication using rapid prototyping tools. Int J Manuf Technol Manag, 2000, 11: 42–53.

100. Chua CK, Chou SM, Ng WS, Chow KY, Lee ST, Aung SC, and Seah CS. An integrated experimental approach to link a laser digitiser, a CAD/CAM system and a rapid prototyping system for biomedical applications. Int J Adv Manuf Technol, 1998, 14: 110–115.

101. Cheah CM, Chua CK, Tan KH, and Teo CK. Integration of laser surface digitising with computer-aided design and manufacturing techniques for developing facial prostheses- Part 1: Direct rapid prototyping techniques. Int J Prosthodont, 2003, 16: 435–441.

102. Cheah CM, Chua CK, and Tan KH. Integration of laser surface digitising with computer-aided design and manufacturing techniques for developing facial prostheses- Part 2: Indirect rapid prototyping and tooling techniques. Int J Prosthodont, 2003, 16: 541–546.

103. Whitaker LA. Facial implants. Plast Surg, 1992, 1: 439–440.

104. Guerrerosantos J. Frontalis musculocutaneous island flap for coverage of forehead defect. J Am Soc Plast Surg, 2000, 106: 18–21.

Chapter 19
Manufacturing Issues

David Hill

Welcome to the jungle, which is exactly what happens once involved in medical products and their application. Sorry – it would probably be more accurate to say a minefield within a jungle, but now with someone offering a pathway through.

Patents. First step – protection from competition, ways to look after innovative ideas, advice on how to proceed.

It is not intended to provide legal or professional advice regarding the protection of completed work, just a set of simple steps which will allow the opportunity to decide what to do or when it is time to consult an "expert" or proper authorities. Should $2,000 be spent on patents for 1 year or $200 on copyright for 10; or should design registration be first?, last?, or bothered with at all? What needs to be protected?, the visual aspect, the chemical make up?, the manufacturing process? Or better still just lie and guarantee some financial reward for the next 25 years? – now recently changed so say 20, well actually really 15 to all intents and purposes if licenses and technology changes are taken into account.

The first part of being in this particular jungle is identifying what's wanted or needed and the different methods in obtaining it, the comparative costs but most of all the amount of protection given, then decide if it is suitable. Sounds a little negative? Let's just say experience.

A decision must always be made whether to actually divulge everything – as with a patent, or just keep silence and introduce the product or application to the world so staying one step ahead of everyone.

First example, James Dyson ("reinvented" the vacuum cleaner) has an aggressive philosophy on design protection and his company spends $1.1 million each year on new design applications, design registrations and lawsuits against competitors who are viewed as encroaching on those designs.

Liability. The designer/inventor is now, along with the supplier, manufacturer or company held responsible for the finished device. If he/she has acted irresponsibly, neglected to carry out rudimentary checks, was aware of any defect or potential problem and did not rectify it or acted unprofessionally in any way then he/she will end up in court. These aspects are intended to be highlighted with reference to the current legislation, and examples of actual cases show how the US and EEC courts are implementing the law.

Not a great deal to read but just enough to make sure everyone is are aware of the potential pitfalls now that there is such an increase in litigation. Also enough to ensure that no one ever becomes complacent, putting the patient first everytime – but hopefully not creating paranoia, or with too many sleepless nights.

Finally, retention of development files – this is critical, not only the obvious checking and progression of results but mandatory for audits, proof of ownership, proof of comparability and long term liability evidence.

Quality. The main thrust now, everyone is answerable for the quality produced – from lab assistant to chief executive, the records kept, methods vetted. QA audits now form the basis of all the mandatory checks carried out by the regulatory organizations (career move alert, QA hold all the cards and not a bad stepping stone once the science degree is completed). Risk assessment methods, plus the necessity of FMEA (failure modes and effects analysis) and its introduction at an early stage. "Design dossier" for class III devices, back to record keeping.

Manufacturing. Wide and diverse subject with enough to keep whole departments busy all on their own. Just to mention briefly, but be aware that the method of manufacture, be it for the individual item, a base material or final device, will dictate the final price and unfortunately money does influence quite a number of things – not least the research, application's development and ultimately any career pursued. Therefore be aware of the influence manufacturing has on the progression and final result of any design or research. Processes will be listed and methods of manufacture categorized, not as a guide but as a prompt for interpretation.

Standards. Everyone is now controlled heavily by standards within both the US (FDA) and EEC (NHS, BS, ISO) – Directives of every description. Hopefully this will clarify certain aspects and explain in words of one syllable *but watch out*; it is critical to keep abreast of new regulations plus who controls which particular regulatory body – more to the point the one which controls your chosen discipline.

Specifications. Not the regulations, not the standards, but the developed and evolved devices/process/material/applications. All materials used in any medical device *must* be specified with no deviation accepted nor tolerated (yes, been bitten on this one as well). No matter what excuses are given, no changes from the specified chemical, physical, mechanical criteria can be justified (this also applies to assembly sequence, processes, heat treatments, storage, etc.). During the development to identify suitable biomaterials all criteria are justified; there is no excuse why a buyer, production staff or "helpful" manager should even consider changing any part or item.

Increased regulations and standardization require much more during a device's development and hopefully this will help to indicate how the materials, assembly, and actual use should be justified – recorded – certified.

19.1 Patents

Very simple – "say nowt, tell nothing, lose nothing".

It is a philosophy, which in some cases works very effectively, but it must be stressed that in other cases it is only by protecting an idea will it normally remain in your control.

There are three factors to consider before attempting to protect any ideas:

Money – How much is an individual or company prepared to spend?
Time   – How much is available to ensure a fully documented claim is filed?
Return – Is it really worth the commitment, the resources to protect this gem of an idea?

Patents are granted to individuals or companies who prove a new process, product, component or improvement on existing processes. Once granted, a patent allows the individual a monopoly on the idea, either to manufacture or license or sell for a set period of time. In exchange for the selected countries' protection of the idea NOTE the details of the invention must be publicly and *fully* disclosed.

Patent applications follow a set formal pattern; below shows the UK system.

(1) Application filed on a set form with description and any diagrams, but no fee. Applicant receives acknowledgement from the patent office and allocating an "Application Number" – *critical since this forms the basis of any prior knowledge, first claim or intellectual ownership.*

Approximately 5 weeks later the application details (information given is invention title and inventors name *only*) are published in the Official Journal Patents & Designs – now on line so that any egotistical genius can show their mum.

(2) Within 1 year a more detailed form is requested by the Patent office along with an invoice for £130. This form must contain the claims and an abstract of the invention so a "Patent Search" can be carried out.

The claims form the legal part of the document and you may now start to consider the use of a Patent Agent (he's a bit like a lawyer – provided with the details he changes them by adding four or five syllables per sentence then charges an exorbitant fee). Six months later there will be a Publication of a Specification reported in the Official Journal.

(3) After 1 year it is a requirement to complete and submit a full examination form plus another £70. Normally there is a backlog between filing and information in the "public domain."

Up to 4.5 years after priority date patent may be granted – provided no one has raised an objection or the patent office has rejected the claim(s).

In the United States there are three kinds of patent – utility, design, and plant patents. Utility patents have a term of 20 years; design patents 14 years, while plant patents have a term of 17 years. They all follow very similar methods of registration as shown above. With the difference that the first off application in USA must contain greater detail, contain a declaration (or oath) stating the applicant is the sole inventor and of course some up front money.

At this point it should be noted that both the UK and USA patent offices are attempting to improve the system and protection of ideas, changes in legislation in both countries is underway. So it is worth consulting these offices via the web to keep abreast of recent, proposed and introduced changes.

Also a very important fact – any patent granted *only* gives protection in the country it is granted in, a USA registration is no good in Japan, a French one no good in USA. If the claims are considered worth the expense then separate applications need to be made per country, in their language.

Almost forgot to mention – once granted there are renewal fees each and every year through the life of the patent, increasing each year until the end. In the UK it starts at £50 and ends at £400 but the European fees are greater – first year £20 up to £723 on the 10th year and after, original filing £89, search £489, examination £1,014, finally granting £507. Unfortunately it is necessary to highlight the translation fees as well.

A common misconception is that the patent gives its owner the right to make, use, or sell the invention. It *does not, it only* gives the owner the ability to exclude others from making, using or selling the invention – *may appear harsh or pedantic but be aware of the subtle difference.* The patent owner may be forbidden from using the invention, usually due to the existence of another patent, or sometimes due to other legal restrictions.

There are essential criteria to be considered before an application for an invention can be processed,

(a) It must be new
(b) It must be an innovative step (not obvious – not something that can be considered as the next evolutionary step)
(c) it must be capable of industrial application (can be made – not an academic concept)

These three requirements should be explained more fully to avoid confusion but unfortunately best to slip in another unfortunate item here before proceeding any further. A patent is normally granted to the inventor unless he/she is employed and the invention was conceived as part of his/her normal duties, or alternatively as a direct result of specific contract agreed, in which case the invention belongs to the employer. Consolation now is that the "inventor" must by law be indicated on the application (recent developments within the UK have prompted additional payments to employees if the granted patent makes "excessive" profits for a company while the inventor is only receiving normal salary and no other financial rewards).

Back to the three criteria, "New" means not just a modification or evolution of an existing process or product. Something "not state of the art" at the time of the application. But also just as importantly not known or available prior, a patent will not be granted if the invention was prior knowledge (even if it was never used). One of the most published examples of prior knowledge is the patent application for a windsurfer. It was decided the application was invalid since years before a 12-year-old boy had built a crude version and sailed for 2 years near a caravan site in full view of anyone on the cliff tops. Although admittedly a limited audience, nevertheless in full public view. All inventions should be kept secret until applications have been filed; a presentation, paper or demonstration might be sufficient to prevent or destroy all ideas of obtaining a patent.

Which rings a bell regarding notebooks and lab records. The burden of proof is with the applicant either to confirm the idea was original or to stop anyone else from making the same or similar claim, easily done if dated entries, with details, are kept. Also most people when working on a multitude of projects tend to become a little confused; it is safer to keep precise notes than relay on your memory.

Inventive is to say not a logical progression or development of a product or process, something which could be undertaken or evolved by any other "skilled person" in the same field of work.

Industrial application, simply put, is can it physically be made? The idea must be able to take a practical form or use. The word industrial is only used as a general term and covers a multitude of practical activities (not limiting any application to machines, machinery, manufacturing or any single science).

The basic costs are not too expensive but there are a number of other factors to be considered. First, a badly constructed or worded application will afford little or no protection from those wishing to profit from other's efforts. So professional advice is available to draft and formally present an application plus suggestions on alternative methods (details later), but this advice is not free. Below shows the cost of obtaining a UK patent under normal methods. One side compares the fee payable to the Patent Office and the other the approximate cost of using a Patent Agent. Compare it against the figures discussed earlier.

Action	UK office fee	Agent fee.
Filing initial application	Free	Minimum £350, up to £600
Search by patent staff	£130	Minimum £400
Substantive examination	£70	Minimum £400
Granting of patent	£150–£450	Varies between £400 to £1,000

It must now begin to be obvious (and appreciated) the point originally raised in the first paragraph regarding money plus time investment against return.

There is also an additional cost when starting to consider protection further afield and a wish to have either European cover (EPC) or international (PCT) and a glance at the table below should show the financial problem of progressing these applications.

A normal British application to final granting (using an agent to compile, progress and monitor) will cost a minimum of £2,000 over a period of 3–4 years. This will be much greater the more complex the subject matter and any subsequent objections raised which require answering. But the cost of pursuing an European patent application to cover, say, 10 countries is likely to cost in excess of £9,000 over the same 3- to 4-year period. Then the additional cost of obtaining application patents in non-alien countries would be approximately £2,500 per country.

Action	UK	EPC	PCT	USA
Filing	Free	£260	£350 + £55 transmitted	$770 (small fee $385)
Search	£130	£815	£1030	$130 processing
Designation, per state	–	£150	£85	–
Substantive examination	£130	£1,200	National fee (per state) plus £130 international examination fee. There is also a request system via EPO @ £1,300	Use PCT
Grant	Varies between £150 and £450.	£600	National fee as applied.	

The EPC and PCT search fees are payable on filing

19.1.1 EPC Contracting Countries

Austria, Belgium, Denmark, Eire (Ireland), France, Germany, Italy, Liechtenstein, Luxembourg, Monaco, Netherlands, Portugal, Spain, Sweden, Switzerland, United Kingdom.

19.1.2 PCT Contracting Countries

Armenia, Australia, Austria, Barbados, Belarus, Belgium, Benin, Brazil, Burkin-Faso, Cameroon, Canada, Central African Republic, Chad, Congo, Cote d Ivoie, Czech Republic, Denmark, Estonia, Georgia, Kyrgyzstan, Democratic People's Republic of Korea, Finland, France, Gabon, Germany, Greece, Guinea, Hungary, Iceland, Ireland, Italy, Japan, Kazakhstan, Kenya, Latvia, Liberia, Liechtenstein, Lithuania, Luxembourg, Madagascar, Malawi, Mali, Mauritania, Mexico, Monaco, Mongolia, Netherlands, New Zealand, Nigeria, Norway, Poland, Portugal, Republic of Korea, Republic of Moldova, Romania, Russian Federation, Senegal, Singapore, Slovakia, Spain, Sri Lanka, Sudan, Swaziland, Sweden, Switzerland, Tajikistan, Togo, Trinidad and Tagago, Uganda, Ukraine, United Kingdom, United States of America, Uzbekistan, Vietnam.

PCT stands for the Patent Co-operation Treaty, which is adhered to by the countries above. PCT patent applications are administered by the World Intellectual Property Organisation. The Patent Co-operation Treaty permits an inventor to file what is called a PCT patent application. The Treaty is an attempt by the countries involved to provide some streamlining of patent applications across several countries at once.

An important part of obtaining a patent is the rewards gained from holding a monopoly for the designated time. Normally an employee can expect no additional payments for an invention which is part of his or her job, say a design engineer. But if the patent granted is deemed to provide an employer "outstanding benefit" then the Patent Act states that the employee – regardless of the job held – should be awarded compensation. Unfortunately the definition of "outstanding benefit" has yet to be quantified; to date there has not been any case taken to court to set a precedent, but this also indicates employers are recognizing the need for rewards and no one as yet has a major complaint.

Alternatively, if someone makes an invention on their own, with no connection with any company, and then decides to sell the invention to a company for an agreed single payment or "consideration," then if the benefit to that company is deemed to be considerable, in relationship to the original payment, then the Patent Act says the original inventor should be allowed another payment or renegotiated remuneration. One last example, plus point, to end this section with. The inventor of the non-intrusive method of riveting ring pulls to the top of cans arranged a license of one tenth of a penny (£0.001) per can made to the manufacturers of drinks; since Coca Cola alone sold over 150 million cans world-wide per day, it shows that glory is very nice but the money is better. The points, protect concepts and ideas, ensure confidentiality and restrict those who have access to the details of the device/product/widget.

Patents are the ultimate in protection but not the only way by any means of looking after personal interests; anything produced which is new can be referred to as "intellectual property" and can be quoted. Any product with an original or definitive shape is automatically protected by design right; no documentation or governing bodies approval is required, no registration (just good in-house records of the design concept and drawing). Design right protects a "shape" for 10 years from the date first shown on the first saleable item, although for the last 5 years of the 10 anybody can obtain a license from the design right owner. Or to put it more simply, a license must be given whether the owner likes it or not.

Other methods of protecting designs are, (1) Copyright – (2) Trade Marks – (3) Registered Designs.

19.1.3 Copyright

Protection is automatic and applies to pictures drawn, songs written, stories from a fertile imagination (remember Harry Potter or Rim Worlds – both concepts have very happy authors), computer programs written, any artistic or esthetic creation. As the name implies, it is only to stop others from "copying" the original, which is why the majority of cases involved in legal action usually come from literary, artistic and musical areas. Although in recent years, the computer world has been in the news since software programs cannot be patented and are only protected by the copyright laws.

The copyright gives certain rights to the original creators (or descendants) allowing them to control the actual use of the finished article for up to (1) 25 years for

published editions, (2) 50 years after death of the author, of musical, dramatic, literary or artistic work (including photographs); similar protection is given whether published or unpublished, (3) 50 years for films, sound recordings and broadcasts (visual or audio), (4) 50 years after the death of compiler for computer programs.

As mentioned before, the protection is automatic once published, but it does not protect the idea. If a new concept for a jet engine was drawn and shown, only the actual drawing would be secure; the idea would be lost to competitors with no comeback. Not like a work of art or music.

Even though not essential in the UK it is now standard practice internationally to mark all work with ©️ followed by the owners name. Provided all record keeping is accurate (and should be by following comments written earlier), then the ownership or proof of originality should not present any problem. Even so, sometimes as an individual, not as a company employee, it might be wise to deposit a signed and dated copy of any "innovative" work with a bank or solicitor if it is thought or viewed there may be potential problems in the future.

19.1.4 Trade Marks

These identify goods or product groups produced; consumers then come to identify the products with these symbols or words. Not only useful for sales and marketing, trade marks can sometimes become more valuable than patents or registered designs. ^{TM} entitles the owner to the exclusive use of the mark/symbol and prevents anyone else from using the same.

Registration takes place simply after a one-off payment and application presentation, but the full registration may take up to 1.5 years because of the cross checking and notices of intent to be advertised (this is to allow for any objections from other companies or presentation of duplication of symbols). The actual shape and form of trade marks are a very complex process and may involve negotiations with other manufacturers. Sometimes even professional advice may be needed to avoid conflict with other trade marks. Even owners of unregistered trade marks can obtain similar rights under common law although it is certainly more difficult (and expensive) than with registered ones. Once registered the trade mark is held in force for 10 years and renewable indefinitely at regular intervals for a nominal fee (at present 10 years @ £200).

Remember, using someone else's trademark is criminal – counterfeiting.

19.1.5 Registered Design

As with design, right shapes and product configuration have automatic protection. But if a certain shape, like a turbine blade, has a specific benefit then better and additional protection can be obtained by registering it with the Design Registry as a "registered design". Even surface textures qualify as would any similar modification designed as decoration.

The original registration lasts 5 years and is renewable every 5 years to a maximum period of 25 years. Even though the main reason for ℝ design registry is an increased protection of an automatic design drawing or concept, it is increasingly being applied to a wide range of products because of their "look." The cosmetic appeal, the look of approval. Items like pens, cases, car bodies, cameras, even computers, but the biggest influx of applications over the last few years have been toys. This is mainly because the method of attracting customers is by visual acceptance, whether the toy functions satisfactorily or even at all largely does not matter. Charges are £35 for the application and if accepted renewable fees £130 first then increasing to £450 by the fifth period renewal.

There is no automatic cover or protection outside of the country registered in for any of the above methods, Patents, Copyright Ⓒ, Design Registration ℝ, Design Right and Trade Mark ™. But agreements are available with the majority of countries, but all require addition documentation and paperwork, plus the inevitable financial considerations.

Making use of these extra interests or additional certificates is a decision an individual, a company or university must make on a commercial basis. Obviously this chapter has only been able to scratch the surface of the design protection but it will allow a slightly more informed individual to form an opinion. Check out the websites of the Patent Office, all are very informative – and free. There is complete confidentiality and the advice they provide could protect any innovative invention fully. Also they can advice on methods of presentation, requirements for complete cover, potential pitfalls and should it be necessary a list of patent agents who will supply the professional "edge" (even though it will be charged for). One other service the Patent office can always be used for, and that is information. For a nominal fee they will conduct an initial search, something which even though it may shatter any dreams initially held may save money in the long run, especially if someone has been there before and logged a claim. It also means that prior information will be available on the first application, so obtain a copy of the application or full patent, design, etc., and compare. Obtain all information assess if the original idea cannot be improved upon and then be accepted as a novel concept. Remember there are approximately 35 million patents world-wide and each one contains full technical information sometimes the only source of detail, make use of it. It has been quoted that 75% of the information contained in patents cannot be found anywhere else, so do not restrict or limit availability to three-quarters of the information available.

19.1.6 Finally Litigation

Finally litigation – legal commitment/actions for protecting what's yours. It is commonly thought that all patents; trade marks, copyright and design registrations are automatically policed by the countries involved in issuing the documents. This is not true, although the law is there to provide actual judgment and implementation of that protection; it does not keep a check on its use.

Patentees or holders must ensure their own interests are upheld and it is there-fore their own responsibility to use the law. Monitor any infringements, log the full details and then seek advice on proceeding with civil action.

If there is a possible infringement under no circumstances either speak, contact or write to the people involved; immediately contact a patent agent or solicitor. This is because in the eyes of the law it is unlawful to issue an unjustified threat against an alleged infringer. What actually is classed as a threat (in law) is never easy to decide upon and a simple contact could effect the result of any claim.

If the reverse happens and someone threatens you with the law courts the same thing applies, do not speak or contact direct leave that to those employed as intermediates.

19.2 Liability

In July 1985 the EC Directive on product liability 85/374/EEC was adopted by the EEC community; this meant that *within* three years strict liability on producers of defective goods was to be introduced. Part 1 of the consumer protection act 1987 was implemented; this was and is the most important statutory liability in the post-war era in the UK and EEC states – it was modeled on s402A of the restatement of Torts (second) of the USA. US civil liability – consider "product liability" as the responsibility placed on suppliers of products for losses or damage caused by those products, *even if misused*.

The above statement was aimed at all products from all industries, but within our own medical device area it has a special significance since the majority of products we supply are either life threatening, or detrimental to the patient if a fault occurs or is not designed out at the initial stages. More important is the development defense clause – how many times have you been informed no sales until trials complete. But the trials cannot be carried out until the product is safe and you don't know if it is safe until it's been trialed. Difficult isn't it?

There are four very important points within the liability laws that I will wish to review.

Article 3. The definition of producer is very wide and includes manufacturers of finished goods, producer of raw materials or component parts used in products. Importers, distributors and any person who puts his/her name or trade mark onto the product. But the medical profession does get a separate mention in this section as requiring particular consideration because they are the "last link between the medi-cal device or medicine supply and the patient." The mention is because doctors and health care personal may be liable under the provision of this particular article when a manufacturer of a defective product cannot be identified. However, within the UK, and for NHS staff, the supplier is considered to be the relevant health authority and *not a specific person or member of staff*, not so in the US where individuals are still liable. Although within the UK, now that the separately funded holding GPs are not employees of the NHS but classed as self-employed and under contract, they should cover themselves.

Article 6. A product is defective when it does not provide the safety to which a person is entitled to expect taking all circumstances into consideration, including presentation, instructions, labeling (new legislation introduced in 1996 and 2003) the use to which it can reasonably be put at the time the device was put into circulation. A product should not be considered defective or unsafe simply because a better or safer product is subsequently put into circulation. Now this is fine within the EEC countries where clinical judgment is still the over ruling control, not so in the US – if there is something on the market safer than that used and an accident or damage occurs the onus is on the medical staff to *prove* their selection.

Article 7. Provides six exemptions from liability for the manufacturer:

(1) "The manufacturer will not be liable if he proves that he did not put the product into circulation." This is understood to mean when a product is delivered to a second person in the course of business or when it has been incorporated into an immovable. Medical materials and devices therefore used in trials before marketing will generally be exempt under this provision – but beware because of the conflicting clause in Article 6, "provide a product fit for use."

(2) "The manufacturer or producer will not be liable if he proves that the defect which caused the damage did not exist when the product was placed in circulation." This applies to shelf life products, damage or modification by a final supplier or distributor, or where labels/instructions for use are detected or removed by a distributor.

(3) "The producer also has a defense if he can show that he did not manufacture the product for an economic purpose, nor distributed it in the course of business". Applies to private transactions and supply.

(4) "The producer will not be liable if he can prove that the defect is due to compliance of the product with mandatory regulations issued by national public authorities." However where compliance with a regulation will not necessarily discharge a manufacturer form liability, it must be shown the defect was the inevitable result of compliance with that regulation. But "mandatory regulations" are also understood to be only those imposed by the law and not contractual. An example is if a beer is brewed in Germany, it must comply with purity laws of that city and not a license between companies.

(5) Development Risk Defence, allows the evolution of new and innovative products, also the development of improvements and safety in existing products. But extremely dependant on the development file and records.

(6) "A manufacturer of a component will not be held liable if it is proven that the defect is attributable to the design of the product into which the component is fitted or to the instructions given by the manufacturer of the finished product." So off the hook! if merely following instructions from the distributor then the responsibility and liability is theirs.

A timely reminder here of Article 3, if "A" sells it, packs it, assembles it, with his name embossed on it, printed or painted, *he is* liable, NOT the manufacturer. But if "A" manufactures a device, formula or biomaterial which a second or third party

places their name or logo onto it, covering or removing "A"s, then the liability is *theirs*.

Article 17. "The directive shall not apply to products put into circulation before the implementating legislation comes into force". Basically it will not be retrospective. The above selected points show the extent and viability of the Directive, even though not fully implemented or in force with all EEC members it is on the statute books and the framework is there for all to work within. All existing and potential designers should be aware of the law, a law which may in time place them, along with their company, in the courts.

The Directive is phrased in such a way so that it is not possible to contract out of liability. There is no maximum liability set (although to discourage trivial claims individual items worth less than 500 Euros are excluded) and a defendant has a right to place a claim on co-producers or others "with responsibility." This could be the manufacturer, the head company of a group, individuals concerned with the production or design(er). It should also be noted that apart from this Directive the changing emphasize on responsibility is a growing trend for "no win, no fee" contracts with lawyers, at present some defendants or claimants do not carry on with a case for fear of costs; this is changing in Europe and following the USA format. Final point on the Directive itself. If a product is defective the defendant does not have to allocate blame, does not have to prove negligence or incompetence, he only has to prove that the product is defective, is not suitable for the purpose designed. Then leave the level of compensation to the courts.

Yet to come, Environmental Liability.

In the USA the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) deals with ecological damage and regulates the reimbursement of any clean-up costs – to be exact the officers involved "extract" the costs from those who are deemed/convicted of causing the problem. Within the EEC countries there is no one control or legislation monitoring or policing environmental contamination. In fact many experts advocate the adoption of the USA CERCLA system since it appears to work efficiently. Obviously the American system cannot be copied directly because of fundamental differences within the laws of the US and EEC, but it could be used as the basis of one.

There has been a European Environmental Bureau in operation for a number of years but it is still trying to motivate an agreeable solution to suit the vested interests of the countries that make up the EEC; this is still a laborious task. With this in mind the Bureau has proposed the following standards which it hopes will be adopted:

(a) That it be a set of European liability laws and formally a Directive on environmental liability.
(b) Clear no-fault base liability means clear strict liability.
(c) The status or meaning of the term environment must be broad enough to cover as many types of damage caused to the environment, so that it also include all types and references to the environment even those classed as non-ownership, e.g., land, air and water, also wild animal and plants.

(d) The access to justice must also be as broad as possible. This is essential for environmental protection organizations as well as individual persons, without adversely affecting individual rights. An action should not be allowed or forced to fail because of a lack of money, it should also be mandatory for those bringing the complaint to demand both an end to the environmentally harmful action and the receipt of compensation.

(e) The proof of cause and effect between the alleged action/event and the damage occurring should be made less strict. In claims of environmental damage it is sometimes very difficult or near impossible for the plaintiff to prove that individual causality or blame. It should therefore be possible and permissible for judges to prosecute an individual, an industrialist or industry on the basis of an adequately strong presumption. (Civil liberty groups will be split here between both parties, keep the lawyers employed for years – more career advice again).

(f) If there are several perpetrators, liability both jointly and severally can be brought. This will allow the plaintiffs to bring one action against the perpetrator of all of the damage in one single action, time saved and actions probably(?) faster.

(g) Guarantees of financial cover must ensure that any high-risk or new operations are not introduced via undercapitalized or shadow subsidiaries that would not possess sufficient capital to cover damage caused.

(h) There must be no limits to liability. Otherwise the incentive for prevention and avoidance will be significantly reduced.

(i) Up to a certain level damage compensation demands must be backed up by means of a cover reserve. This can be achieved by means of either money set aside on a company's annual basis or by legal guarantees. In addition, the introduction of compulsory insurance is necessary since the cost of repairing environmental damage can be so high as to exceed the perpetrator's capitalization. Ties in with (7) above, company law and start up legislation will have to be reviewed – this is not just one area of interest.

(j) Liability funds should be set up. These will be primarily for those cases where the perpetrator cannot be identified but the damage *can* be attributed to specific sectors. Similarly to the US trust funds ("Superfund" the nickname for CERCLA) they could be set up on the basis of source origins. The funds could even be arranged on a regional/national basis in order that the money benefits the area/country from which it came. The money would be supplied from a pollutant-related assessment – like the American model, only this would be a pollutant pays principle (if identified).

Regardless of the proposals mentioned above and existing laws in operation in the USA and EEC countries, there will be an ever-increasing growth in the environmental controls administered to manufacturing. Which means that devices that require materials developed, processed, refined or just plain machined will be monitored and required to "pay the price."

New designs and evolution of biomaterials, plus their eventual use, will result in the use of expensive materials, technology to process them, sacrificial loss

of materials, use of resources which will effect others, emotive issues that "extremists/zealots/concerned individuals" will pass judgment upon. Be aware that academic excellence, "a better mouse trap", could be stifled by an environmental objection and that all product designers have now entered a new realm where other influential external factors must be considered.

19.3 Quality, Standards, Specifications

These are all interlinked and can be covered by a few well-chosen sentences and peals of wisdom (which is why this author is surprised someone asked him).

Three separate areas cover quality during your biomaterial development – an audit, a "Design Dossier" and use of Failure Modes and Effects Analysis (FMEA) sometimes mistakenly refereed to as risk analysis.

19.4 Audit

An independent review of the materials development, the independency guaranteed by any QA department (who as everyone knows are a law unto themselves anyway). The audit should normally be stringent, objective and open minded, the overall objective is to provide assurance that the material (1) meets all requirements laid down in the original brief, (2) is safe to manufacture and use, and (3) does not present any danger to the patient. Not only during any testing or initial assessment but for the full duration of its working life. Specific points to consider are:

Quantity produced, batch to batch
Quality limits (acceptance/rejection)
Manufacturing costs
Manufacturing methods
COSHH (or whatever health and safety regulations apply in each country)
Method of use
Environmental effects, during and after manufacture

Personnel attending such audits should comprise of members from:

Design
Aspects of the design explained
Reasons for particular actions explained
Background information on use
Provide documented records and justification notes if required.
Production
Equipment required and why
Capital budget requirements

             Assembly methods
             Component supply
    QA
             Test methods and initial results
             Manufacturing QC test methods
             Document recordal
             Staged inspection levels
   Marketing
             Verify design's interpretation of use
             Quantify production numbers required
             Confirm market sales forecast

As more and more documented evidence is being called for by various regulating bodies a design audit is virtually becoming mandatory.

19.4.1 Design Dossier

This document is a detailed specific file (separate from the normal development file retained in a design office which contains all correspondence, quotes, references etc.), which is maintained by a designated responsible person within the company and must provide evidence and/or documented proof of a design, in such detail, that the manufacture, performance and function of a product can be evaluated. In this instance to comply with the MDD 93/42/EEC as well as mandatory that all class III devices must comply, but other requirements may apply to meet CE mark, manufacturing regulations, or international export/import regulations.

"Interpretation of Technical Documentation Required for Class I, IIa, IIb and III Devices by MDD" is 14 pages long. Basically detailed information necessary on the design and development of a medical device governed by the Medical Device Directive 93/42/EEC Article 11.1a for Class III these also require conformance to QA requirements EN ISO 9001/46001. Sounds complicated but a simpler method is to justify your selection, manufacturing method, design, document the calculations and test results, copy everything and place these "hard copies" into a folder. The design dossier then becomes very easy and not a paper chase when required. Yes it was *hard copy* and listed below are the requirements (suggested but best adhered to):

(A) Description of the actual design, it's manufacture and end use performance. Plus details and documents which conform to the Directive, Annex II, 4.2. Namely those referred to in Annex II 3.2c.
(B) Details of any planned changes or developments which will change the product.
(C) Design brief plus all functioning requirements. The FMEA and actions implemented as a result.
(D) Methods of control and design verification. Plus how the design's progress was checked and measured during development.

(E) Proof of conformity to additional regulations if connected to other equipment to aid, increase or allow its use.
(F) Specific statement relating to substances referred to in Sect. 7.4 of Annex 1 of the Directive, plus any applicable data in connection with testing carried out.
(G) Label copy and operating instructions (if any).
(H) All clinical data which was compiled during the development stage.
 (I) Details of the manufacturer, the group or type the device falls within.
 (J) Device classification and the justification regarding its selection.
(K) Details of any similar devices already in circulation.
 (L) Full device description and technical features.
(M) Catalogue details, numbers and descriptions.
(N) Material data for all component builds.
(O) Specific manufacturing methods.
(P) Sterilization, bioburden, microbiological testing information – where applicable.

One last thing the directive suggests that the dossier is retained for a minimum of 5 years.

19.5 FMEA

A formalized approach to this problem is the Failure Modes and Effect Analysis or FMEA, a systematic method that considers the potential failures of each component, sub-assembly and final product build. The actual analysis starts by considering the failure effects on the output on one level and their resultant domino type influence on higher levels. It is considered to be a bottom to top analysis working from a possible effect to the possible cause.

It is an established design technique for assessing the effect and suitability of a design in matching the requirement brief.

Although each FMEA chart should be individually marked and recorded for its own specific product they will all contain the same data:

Part Name
Description of Part Function
List all Potential Failure Modes
Estimate Potential Effect of Failure
Score the Severity of the Effect
Consider the Probable Causes of Failure
Assign Probability of the Cause Occurring and Part Value
List Method(s) of Detection
Mark a Detection Rate
Calculate an Overall Risk Priority Value
Recommend Remedial Action to Correct Potential Failures

19.5.1 Standards

Quite simply, to manufacture and supply medical devices for use in any country there are standards with which you must comply. Engineers, scientists, manufacturers all refer to these standards and codes of practice continually because they provide an acceptable set of guidelines; although many are not actually legally binding or enforceable they are what is termed quasi-legal.

The International Standards Organisation (ISO) promotes an acceptance of certain standards. These standards normally originate from collaboration between the ISO and member countries own legal sections such as BSI from the United Kingdom, ASTM from USA and DIN form Germany. In 1993 the Medical Device Directive (MDD) was published; this was a direct development of the new approach described above.

For sales to the USA the following must be adhered to: Comply with plus audit to the USA form and GMP (Good Manufacturing Practice) standard set out by the FDA, before any sales of product.

For product market clearance, the device selected requires either a 510(K) to prove substantial equivalence to a pre-arranged device (e.g., a device currently on the market which is acceptable and safe) to prove safety and effectiveness of the device. There is a short form entry which should proceed through the regulatory department in 90 days, Or a PMA (pre market approval) for new devices that are proven not substantially equivalent under the 510(K). A lengthy 3-year process, also expensive.

For information the following Draft International Standards have been released:

> 95/563336 ISO/DIS 12891-1 "Retrieval & Analysis of Implantable Medical Devices".
> 95/563133 DC ISO/ 11737-2 "Sterilisation of Medical Devices. – Tests of sterility performed in the validation of a sterilisation process".
> EC Packing & Waste Directive (MDA Bulletin 16). The directive requires many types of packing to contain certain percentages of recycled materials and to be manufactured of recyclable materials.

Documents you might consider:

- Guidance Notes for Completion of PCA 1 & PCA 2 – 005375
- Guidance on Biocompatibility Assessment – 005373
- Guidance Notes for Manufacturers of Dental Appliances – 005379
- Guidance Notes for Manufacturers of Class I Medical Devices – 005378
- Guidance Notes for Manufacturers on Clinical Investigations to be Carried Out in the UK – 005374
- Information for Clinical Investigations – 005376
- Silicone Implants and Connective Tissue Disease – 005380
- Framework Document – 005188
- Business Plan 1994/95 – 005187
- Doing No Harm – 005189

Listed below are biocompatability ISO references.

Evaluation test	Description	ISO 10993 reference
Cytotoxic	Lysis of cells (cell death), inhibition of cell growth and other toxic effects on cells	10993-5
Sensitisation	Potential for contact sensitisation	10993-10
Irritation	Irritation potential to, for example, skin, eye or mucosal membranes	10993-10
Intracutaneous reactivity	Localised reaction of tissue; applicable whenever irritation by dermal or mucosal tests are inappropriate	10993-10
Acute systemic toxicity	Potential harmful effects of exposure usually by single administration	10993-11
Subchronic (Subacute) toxicity	Harmful effects following repeated systemic exposure, for example, up to three months in rats	10993-11
Genotoxicity	Potential for mutagenic activity, normally up to three in vitro tests (for example, Ames, chromosome aberration and mouse lymphoma)	10993-3
Implantation toxicity	Local tolerance effects following implanted contact up to, for example, three months in rats	10993-6
Haemocompatibility	Possible effects on blood or blood components; haemolysis tests determine the degree of red blood cell lysis and the release of haemoglobin in vitro	10993-4
Chronic toxicity*	Harmful effects following longer term repeated systemic exposure, for example, more than three months in rats	10993-11
Carcinogenesis*	Tumorigenic potential over major portion of animal's life-span, for example, two years for rats	10993-3
Reproductive and developmental toxicity*	Potential effects on reproduction function, embryonic development (teratogenicity) plus prenatal and early postnatal development	10993-3
Biodegradation	Processes of absorption, distribution, biotransformation and elimination of leachables and degradation products where potential for resorption and/or degradation exists	10993-9

* Rarely performed for medical devices (may be needed if specific concerns are raised by other safety studies and/or literature findings for the same/similar materials).

19.5.2 Specification

A detailed description, including construction, materials and quality required. Simple really, not complex, not difficult; then why does it cause so much aggravation and trouble?

Specifications should be regarded as cast in stone with no variation allowed or sanctioned. Rigid controls are necessary to enforce compliance, with set formal structure to follow if changes are thought to be necessary at a later date. Once this idea is accepted at all levels from Director to Floor Sweeper, then risks are minimized. A very short message, but probably the most important within the book. The formal specification should be controlled by an independent department with no vested interests or conflicting directives. If possible the manufacturing head should have no influence nor say in how the specification is written, modified nor enforced – but he/she should be answerable when the standards are not maintained or adhered to. Once "the lunatics start to control the asylum" and can change the standards to suit what they produce or supply then the end product – be it a material, a component or product – is at risk from the very beginning.

Any changes to be introduced must have a formal request checked and if accepted then signed for by the relevant departments, unless all parties agree the proposed change should not be introduced.

New issues of the specification (along with information updates, instruction for use changes, drawings, etc.) must be dated and the issue numbers changed. Details of why the change was introduced must be recorded and kept as well as the cancelled issues retained or archived. The recommended retention of the originals is for 5 years. It may sounds complicated and only suitable for large companies but lots of small organizations or departments get into trouble because of lax record/specification keeping.

19.5.3 Manufacturing

The methods vary obviously depending on the final product required, which might just be a chemical process, a raw material, an ingenious and complex component, an insert or multi part assembly. Electrical, electronic, mechanical, don't know? Bioproducts can and do vary in their construction and end use.

Every method available must be verified and validated though, to ensure consistency and quality of product – more so for repeatability. There is no point in just listing the methods of supply, machining, surface treatment or joining, nor the methods of assembly, heat treatment, post treatment, etc., unless designers can actually understand their application. To do that would need a large number of books in addition to this one (all at reasonable discounted prices of course). Sufficient it to say, there are facilities available that will supply details for the manipulation of materials, compounds and elements in any and all ratios, mixes and alloys a manufacturer could ever wish for, all that has to be done is ask! search!! look!!! listen!!!! Search

engines on the web, colleagues, college lecturers, research notes/publications, trade
and professional lectures and, God forbid, libraries.

Depending upon an engineer's discipline, this will determine the methods freely
available or advisable for use. But always keep alternatives in mind, look at other
peoples work, view how their parts were produced, not to "copy" but to utilize their
methods of manufacture. The same methods will react differently with different
materials and produce a multitude of different features – features that can made
use of.

Do not be put off by cost; all new ideas will inevitably have to pay the price
of development, price of small batch manufacture, price of experimental testing.
This is always going to happen but if the idea is worth looking at then make sure a
manager, a director provides or arranges the finance – that's their job.

If there is potential in an application of a gene therapy which will prompt the car-
tilage of the knee to start and regenerate itself, then lobby for the back up to obtain
materials, test facilities or just simple funding. If on the other hand the best way
forward, either from an individual's or group's viewpoint, is a radical alloying of a
polymer and quasicrystals (flame sprayed mild steel substraight using quasicrystal
surface modification have produced a virtual wear proof bearing surface suitable
for knee replacements), then push for the access, money and of course the glory.
Nanoparticles, nanomachining, nanotechnology, larger micromachining, microtech-
nology – all techniques that should be known or at least be aware of, but have you
considered their application to any specific or particular environment? Research and
development of new biomaterials is fine but where are the applications? Consider
the uses as well as the biological interface and acceptance of the body; somewhere
there must be a grain of thought as to their manipulation into a feasible product,
device, enhancement which will allows the patient to derive some benefit. Remem-
ber that there will always be a method of manipulating any material for use; it might
be a big hammer or a scanning microscope or a multi faceted self focusing laser.
But if it does not exist yet, it will.

Never let the problems of today's technology worry, halt or restrict anyone's
enthusiasm nor place restriction on creative imaginative ideas.

Index