John N. Coupland

An Introduction to the Physical Chemistry of Food

Food Science Text Series

The Food Science Text Series provides faculty with the leading teaching tools. The Editorial Board has outlined the most appropriate and complete content for each food science course in a typical food science program and has identified textbooks of the highest quality, written by the leading food science educators.

More information about this series at http://www.springer.com/series/5999

John N. Coupland

An Introduction to the Physical Chemistry of Food

John N. Coupland
Department of Food Science
Pennsylvania State University
University Park
Pennsylvania
USA

To Jennifer, Michael and Anna.

Preface

There are many ways to study food and to be considered an expert: a chef at a fine restaurant, a family cooking together at home, a hunter dressing a deer, an engineer designing a grain mill, a winemaker pressing grapes are all food experts. What distinguishes a food scientist is that they aspire to some level of understanding of *why* the food behaves as it does. Roald Hoffman (1998) talks about understanding as being either "vertical" or "horizontal"; vertical understanding offers a mechanism for a phenomenon in terms of more fundamental ideas, horizontal understanding is analysis within the terms of the existing discipline. A horizontal understanding of making an omelet would be a recipe; a vertical understanding would be in terms of protein chemistry and network formation. I agree with Hoffmann that the most useful understanding draws on both. For a food scientist to "understand" an omelet, the changes in the egg as it cooks must at be related to both changes in protein conformation and, at the same time, to the conditions in the pan controlled by the cook.

What then are the more fundamental ideas the food chemist should look for vertical understanding? Clearly the properties of food emerge from the molecules that make it up and introductory science courses are quite good at preparing students to think in terms of a molecular world. General chemistry courses teach the basics. Organic chemistry gives some functional groups and molecular transformations, and biochemistry provides the molecules of life and enzymatic catalysis. This course progression provides a reasonable background for vertical understanding for many aspects of food chemistry (e.g., browning reactions, lipid oxidation). However, many other food properties depend on the physical, non-covalent interactions of molecules in foods; topics touched on briefly in the most introductory general chemistry classes and then ignored. The student's understanding of the physical properties of foods is therefore fundamentally unscientific—they learn that a low water activity means the water is "bound" by the food components or that emulsion droplets "tend to" coalesce. This is, at best, science as proverbs with no possibility for real vertical understanding.

Another progression of courses would prepare a student to understand the properties of food in terms of physical chemistry. However that pathway is a difficult one, firstly because "real" physical chemistry courses often draw on a stronger background in chemistry and mathematics than is typical for a food science undergraduate. Secondly, many of the fundamentals of physical chemistry, especially quantum mechanics, are very demanding yet have only very limited use in understanding the physical properties of foods. Other

material, for example activity coefficients, colloid science, and phase dia-grams are immensely important in foods yet often mentioned only in passing in general physical chemistry classes. Lastly, general physical chemistry is in its nature general so few examples will be relevant to foods. (Biophysics, when it is available, is often a better option both in content and in examples).

The primary goal of my book is to help food science students reach a useful vertical understanding of the physical chemistry of foods within the context of their typical educational path. I have tried to introduce the impor-tant phenomena, the food science, but at the same time provide a mechanism for why they occur, the physical chemistry. An explanation is an argument, "because this, therefore that" and some ways of making these arguments are more helpful than others.

The arguments of physical chemistry are rigorously mathematical and, for the rare student that can master the mathematics, deeply satisfying. The rare student. More common is the student who learns the proof but loses the meaning, and more common still the student who loses both[1]. However, without mathematics the reasoning behind physical chemistry is reduced to imaginary models, "cartoons." Many physical chemists become instinctively uncomfortable at this point as they fear their subject will disappear in a flurry of hand waving. I argue though that the thoughtful use and refinement of physical models can provide a real pathway to understand the physical chem-istry of foods.

We build our physical understanding of the world around us through a series of representations, of models, that we continue to either refine or reject based on their usefulness. We do this as a scientific culture, Newton built a theory of motion—Einstein refined it, but also importantly at individual and pedagogical levels. A child might wonder that the sky is blue and be satisfied to be told it reflects a blue sea, a useful model and an appropriate under-standing at an early stage. Later, as an undergraduate, they could replace that model with a better one incorporating theories of Rayleigh scattering and structures of the lens and retina and yet later with sophisticated models of how photons interact with matter and how electrical stimulation leads to sen-sory perception. At each stage of their education, the individual understands the phenomenon at some level. None of the models is complete, but at each point the fact there is an argument, "the world is like this *because* of that, means the individual has something to argue against rather than just facts to accept. The process of rejecting models and building better ones is the pro-cess of both discovery and of learning.

I base the structure of the book around a very basic model of molecules attracting and repelling one another in the context of the randomizing effects of heat. This approach has the advantage of being deeply intuitive, the mol-ecules making up the food can be understood as classical particles, and also of starting from the simplest pictures of solids, liquids, and gases from high school science. I use the first three chapters to set up this foundational physical model. The first chapter deals with the basic rules of thermodynamics and is

---

[1] One of my least favorite student questions is "Do we need to learn the equations?" No, but you do need to understand them.

likely to be a repetition for many readers (although in my experience, students can readily repeat a definition of entropy, enthalpy, and Gibbs free energy without really understanding what these terms mean). I have approached the topic from a molecular perspective where enthalpy is expressed as bonding and entropy as disorder. In Chap. 2 thermodynamic properties are expressed more explicitly in terms of the structure and interactions of molecules and intermolecular bonds are introduced in some depth. Chapter 3 uses the ideas of molecular interactions and the presence of a high-energy intermediate as a way to explain measurable rates of change.

The next two chapters use the basic model to address the very general problem of ingredient miscibility and its consequences. Phase behavior is central to the properties of most real food and rarely considered in any depth in general physical chemistry classes. Chapter 4 uses the thermodynamic rules from Chap. 1 to relate the molecular interactions from Chap. 2 to the properties of mixtures of molecules. This chapter contains the longest mathematical derivation in the text to calculate phase boundaries and to provide a more solid mental model of the roles of enthalpy and entropy to the central question of ingredient miscibility. Once phases have separated, the properties of the interface between them become important. Chapter 5 introduces a mechanical and energetic definition of surface tension and then discusses the properties that derive from it.

The remaining chapters apply the basic model and the resulting properties of multiphase materials to understand the structure and properties of specific types of matter important in foods (Crystals, Polymers, Dispersions, and Gels). However, having made the decision to focus on structures as the organizing principle of the book, some topics are necessarily split between chapters. In particular rheology is covered in Chap. 6, 7, and 8. Newtonian and non-Newtonian rheology is introduced in the context of viscous polymer solutions and refined in the context of dispersions. The rheological properties of solids are covered in the final chapter on gels.

The book is designed to be read as a narrative and as an introduction to a broader topic. Each proposition is developed from simpler concepts so the flow of chapters makes a logical sequence for a course of study. To make it easier to read I have tried to minimize in-text citations and used a bibliography at the end of each chapter to describe the material I found most useful writing the text and where the reader might look for a deeper understanding. Specific information from a particular source is cited as normal in the text.

I have used some of this material in my undergraduate food chemistry classes and found qualitative explanations to questions like why there is a delay before the onset of crystallization and why are polymer solutions viscous helpful. I use the text much more directly in my graduate course in "Food Physical Chemistry," but for this group how the theory is applied in the process of scientific discovery is much more important. I have included some examples of this as boxes in the text. I have found a useful format for graduate students is to ask them to read a section in advance and then give them some data from a paper and ask them to draw cartoons to explain how the changing organization of the molecules causes the changes seen, or, "The Reverse Problem" where they use a physical molecular model to predict

the results of an experiment. I am deeply grateful to the students who have worked with me on iterations of the text and approaches to teach from it.

I am grateful to many of my colleagues, friends, and former students for helpful discussions and criticisms of drafts of parts of this work. Many of the good ideas in the book came from them; the remaining mistakes are mine alone. In particular Claire Berton-Carabin, Eric Dickinson, Ibrahim Gulseren, Rich Hartel, Denny Heldman, Julian McClements, Brent Murray, Perla Relkin, Don Thompson, Umut Yucel, Jochen Weiss made valuable contributions. I chose a single-author approach to achieve a greater unity of vision for most of the book but I am thankful to Rammile Ettelaie and Allen Foegeding for sharing their expertise and co-authoring the chapters on polymers and on gels.

I am also grateful to the growing online community of scientists, most of whom I only know as Twitter IDs, who were generous in helping me track down data, references, or even just offering encouragement.

I am indebted to the staff at the University of Leeds and at the University of Hohenheim for their hospitality as I worked on this book over two sabbatical leaves and to Penn State University for allowing me to take two sabbatical leaves. I am also grateful to my editor at Springer, Susan Safren, for her continued faith that this book would one day be written.

Lastly, I offer my inadequate thanks to my family for their patience and support as I worked through this long project. I could not have done it without them. Writing a book takes time, and your perspective changes as you write. What seemed important shifts, and it's hard to keep track of the essential narrative. About half way through this project I remember watching my baby daughter rolling around on a mat and thinking that would be a great analogy for a random walk. She could roll left and she could roll right but she would surely never leave the safe confines of the blanket. Now I have to rush to meet her from the school bus. Further corrections can surely wait for the second edition.

Contents

Basic Thermodynamics

1.1 Introduction

We expect food to change over time; recipes suggest cooking times, packaging states a shelf life, and we will pay more for a 10-year-old than a 5-year-old whiskey. Some changes occur over a fraction of a second and others over several years; some improve the quality of the food while others harm it. Whatever the mechanisms involved, controlling changes in foods to optimize quality and ensure safety is the primary task of the food technologist. We must answer two important and distinct questions about change—what can happen, and will it happen fast enough to be relevant to the food we eat? The first question is concerned with the thermodynamics of the system and the second with the kinetics. If we observe a change, then we know that it is both thermodynamically possible and kinetically viable. If we see nothing, it could be either thermodynamically impossible or thermodynamically possible but kinetically too slow to be important. For example, during baking, bread will brown rapidly, i.e., we can conclude that the browning reaction is both thermodynamically possible and kinetically viable. However, if the same dough is held at room temperature, it remains the same pale color over several weeks; either the reaction is thermodynamically impossible under these conditions or thermodynamically possible but too slow to be seen.

We will return to the questions of kinetics in Chap. 3, and in this chapter, introduce the basics of thermodynamics. Thermodynamics is an axiomatic subject; that is to say, it is based on a few simple statements of how the universe is observed to behave. These statements (the laws of thermodynamics) are based on observation rather than on any external proof, but if they are accepted, the entire edifice follows logically. No one seriously expects the laws of thermodynamics to be overturned, and as much as one can say for any human endeavor, the theory is always right. From our point of view, the most important laws of thermodynamics are the first and the second[1]: for any change, energy is conserved (first law), but the total entropy of the universe increases (second law). Clearly, to do anything useful with this, we will need to better understand what energy and entropy mean. To start this chapter, we will try to define our terms on the level of the molecules that make up our food. We will then see how we can use the laws of thermodynamics to understand change and equilibrium in that small part of the universe important to us.

---

[1] For completeness, the zeroth law of thermodynamics (oddly numbered because its position in the logical scheme was not accepted until quite late in the development of the science) states that if object A is in thermal equilibrium (i.e., in contact with and at the same temperature) with object B and with object C, then objects B and C are also in thermal equilibrium with one another. This seems a somewhat tortured presentation of an obvious statement, but it is necessary to introduce the ideas of temperature and thermometry. The third law is that the entropy of a perfect crystal at 0 K is zero and gives us a scale for entropy.

1.2 Energy

"Energy" is used so commonly in regular conversation that it is easy to get distracted from the formal scientific definition. Something with energy has the capacity to do work, that is, to move a mass against an opposing force. So, for example, lifting a weight against gravity is work, but so is pulling an iron bar away from a magnet (i.e., work against a magnetic field) or blowing a soap bubble (i.e., work against the surface tension, see Chap. 5). Whenever we talk about energy, we should be able to imagine a way that it could be harnessed, however impractically, through pistons, levers, and pulleys to lift a weight. For example, a compressed spring has energy because we could imagine using it to fire a ball in the air, and similarly a flying ball has energy because we could imagine it hitting a seesaw apparatus and launching a second ball upward. Other more complex examples are also valid. Gasoline is the energy source for a car; it is burnt in the engine to power the expansion of the cylinder to turn the drive shaft and eventually to spin the wheels. We could use that energy, however inefficiently, to move the massive car up a hill against the force of gravity. Similarly, when sugars or fats are digested and broken down in muscle cells, they power the contractions that could propel you up the same hill on foot against the opposing force of gravity. Although both operations are inefficient, with most of the energy wasted as heat, they justify us saying that gasoline and food have a certain amount of energy because both can be harnessed to move a mass against an opposing force.

Energy occurs in two forms: potential and kinetic. Potential energy is due to the position of a mass relative to an applied force, and kinetic energy is the energy of a mass due to its velocity. The most familiar and important example of potential energy in the world of the large heavy objects we can see and touch is gravitational potential energy, $E_{gravity}$:

$$E_{gravity} = mgh \qquad (1.1)$$

where m is the mass of the object, g is the acceleration due to gravity ($\sim 10 \ ms^{-2}$), and h is its height. Another example of potential energy is a stretched rubber band held in position against its tendency to contract. We could write an equation similar to Eq. 1.1 to describe the potential energy of the rubber band, but we would need to replace g and h with terms describing the strength of the rubber and its degree of stretching. Indeed, whatever the nature of the force (e.g., gravitational, electromagnetic, surface, or nuclear), a mass will have energy as a result of its position relative to it, and the form of the energy equation will be similar to Eq. 1.1.

We are not usually concerned with heavy objects like balls and elastic bands but rather with combinations of molecules that make up our food. What does potential energy mean at this scale? The masses of the molecules are too small for gravitational potential energy to be important (this makes sense—a ball will fall to earth under gravity but fructose does not "settle out" of fruit juice), and we can also ignore changes in nuclear energy if we are not planning any nuclear reactions (and this is probably fair in most food systems). We are left with most of the potential energy in chemical systems residing in the bonds within and between molecules. The nature and strength of these bonds will be a major focus of Chap. 2, but at this stage, we can imagine if two molecules are chemically attracted to one another (i.e., tend to form some sort of bond), then it takes energy to hold them apart and the nonbonded molecules will have a higher potential energy. An analogy would be the two masses separated by an elastic band; moving them apart would require work against the elastic properties of the rubber, so the more separated state would have a higher potential energy. Conversely, two molecules that repelled one another would have a higher potential energy if they were pushed closer together.

The second important form of energy the kinetic energy, $E_{kinetic}$, of a mass is proportional to the square of its velocity, v:

$$E_{kinetic} = \frac{1}{2}mv^2 \qquad (1.2)$$

So a heavy object will move slower than a light object with the same energy. At the molecular scale, kinetic energy lies with moving and vibrating molecules. We will return to the movement of molecules in Chap. 2, but for now, it will suffice to note that the kinetic energy of molecules is proportional to the absolute temperature.

Having defined energy mathematically, we can use the first law to make predictions. Introductory physics classes are filled with trivial examples of this applied to individual objects. For example, consider a 1 kg ball thrown upward at 10 ms^{-1}. Using Eq. 1.2, we can show $E_{kinetic} = \frac{1}{2} \times 1 \times 10^2 = 50$ J. As the ball moves upward, it will slow and eventually instantaneously stop before returning to the earth. Energy is conserved, so the kinetic energy of the ball had at the start must equal the potential energy achieved at the top of its arc. From Eq. 1.1, $E_{potential} = 50$ J $= 1 \times 10 \times h$, and so the maximum height reached is 5 m. This analysis neglects the important losses due to friction with the air, but we know from the first law that if in an experiment the ball did not reach its expected 5 m height maximum, then the energy shortfall would exactly be matched by the thermal energy due to frictional heating. Such simple Newtonian mechanics are complicated at the molecular level by the disordering effects of heat as described by the entropy.

1.3 Entropy

The first law of thermodynamics tells us that energy can convert from one form to another but does not deal with the fact that some energy conversions happen spontaneously while others do not. It made sense that the ball in the previous example would not go as high as we expected, as some of the kinetic energy was "wasted" as frictional heating of the air rather than converted to potential energy (height). However, the reverse process—some of the energy of the air molecules spontaneously transferring to the ball and starting it moving—would be unthinkable. Our everyday observations are so filled with this sort of asymmetry that we rarely stop to notice it: an ice cube will slowly melt in a warm drink but never grow; milk stirred into tea will never "un-mix" if the direction of the spoon is reversed. In each of these examples, both the forward and reverse reactions obey the first law, but only the forward reaction is seen. The second law of thermodynamics provides the distinction between these cases—the only processes that occur spontaneously are those that increase the total entropy of the universe.

Entropy is a measure of statistical probability, and the second law tells us that over time, the universe will progress toward its most probable state. We can see this tendency toward the most probable state in a thought experiment that, as we shall see later, shows some parallels to real molecular systems. Take ten dice, each initially showing six and throw them in sequence. The motion of an individual dice is described by Newton's laws, and the result should be predictable. However, in practice, the multiple collisions with the table surface effectively scramble the motion, and the individual rolls give random results. Is it possible to get all sixes? Yes, but very, very unlikely; the odds of rolling ten sixes in a row is $(\frac{1}{6})^{10}$—about 1 in 100 million.

In a chemical system, there are very many molecules moving and interacting with one another. Even if we neglect molecular interactions and bonding for the moment, there will be constant exchanges of kinetic energy between the molecules as they bounce off one another and off the container walls just like the dice bounced off the table surface. Each collision is described by Newton's laws of motion, and the properties of a molecule (e.g., position, energy) should be predictable. However, the multiple collisions effectively randomize the outcome, and we have no real way of computing the properties of a large number of individual molecules over a reasonable time. Imagine placing ten gas molecules in a cylinder and then opening the stopcock to a system of five other similar but empty cylinders (Fig. 1.1). After sufficient time for them to move around, each molecule will have gone through so many collisions that its position will be effectively random and the molecules will be distributed between the six chambers. Is it possible all the molecules returned to the starting cylinder? Yes,
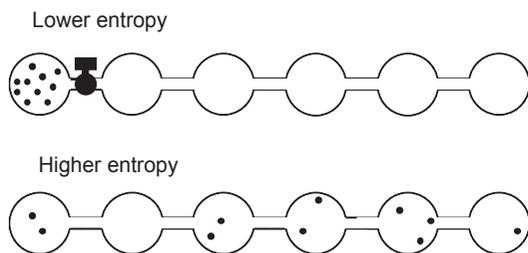
Lower entropy

Higher entropy

Fig. 1.1 Schematic illustration of gas molecules expanding from one cylinder to five other empty cylinders. The molecules are shown as points and illustrate a snapshot of their position. Statistically, the expanded gas molecules will be evenly distributed between the six cylinders

but very, very unlikely; the outcome is governed by the same statistics as the rolling dice examples. Indeed, if we use more realistic numbers of molecules, say a mole ($=6 \times 10^{26}$), the odds of them all returning to the starting state over the lifetime of the universe become negligible and we can say that the forward reaction (the expansion of a gas into a vacuum) proceeds spontaneously, but the reverse reaction (spontaneous contraction of a gas to leave a region of vacuum) is impossible. Both expansion into a vacuum and contraction back to the starting state are permissible under the first law, but the overwhelming improbability of the second case is the basis for the irreversibility of the process and the second law.

The relationship between entropy, S, and statistical probability is given by

$$S = k \ln \Omega \qquad (1.3)$$

where k is the Boltzmann constant ($= 1.38 \times 10^{-23}$ JK^{-1}), and Ω is the number of microstates, a measure of statistical probability of a given configuration. The number of microstates was defined by Richard Feynman (1963) as "the number of ways the inside of the system can be put together without changing the outside." In the gas expansion case, we could seal off the stopcock and measure the gas pressure in the first cylinder. If all the molecules returned to their starting state, then the pressure would have a certain (maximum) value, but any other configuration of the molecules would lead to a lower pressure. There is only one way to put together the inside of the systems (the location of the molecules)

without changing the property measured from the outside (the pressure); the entropy is 0. If the gas expanded out from the starting cylinder, the measured pressure would be lower. There are 6^{10} other ways the molecules could be arranged among the cylinders that would lead to a lower measured pressure so the entropy of the expanded gas is 18 K.

The simple gas molecules we have looked at so far were just treated as points, and we measured their entropy based on the constraints on their position (i.e., translational entropy, Fig. 1.2a). Real molecules have other types of entropy depending on the state of variability vs. order in any property that can affect the overall measured properties of the system. So, for example, if we had more than one molecular species (Fig. 1.2b), there would be an entropy of mixing term to describe the relative position of one type of molecule to the others. The entropy of mixing will tend to drive molecules to diffuse from regions of high concentration to regions of low concentration. For example, when a steak is marinated, the flavors from the spices will spontaneously diffuse into the meat driven by the concentration gradient and the entropy of mixing. If the molecules are anything other than spherical, they can vary in orientation, and molecules aligned with one another have lower orientational entropy than randomly oriented molecules (Fig. 1.2c). Overcoming the orientational entropy is one of the major difficulties in forming a crystal from a liquid (Chap. 5). Flexible molecules will have configurational entropy depending on the distribution of bond angles seen (Fig. 1.2d). For example, many polysaccharides have a compact coil shape in solution rather than an extended chain shape (Chap. 7).

1.4 The Boltzmann Distribution

In our dice example, the outcome of each roll was random, but we would expect over a large number of rolls to see approximately equal numbers of ones, twos, threes, etc., and an average result of 3½. If we saw a clear preference for one number over the others, we would conclude

Lower entropy   Higher entropy
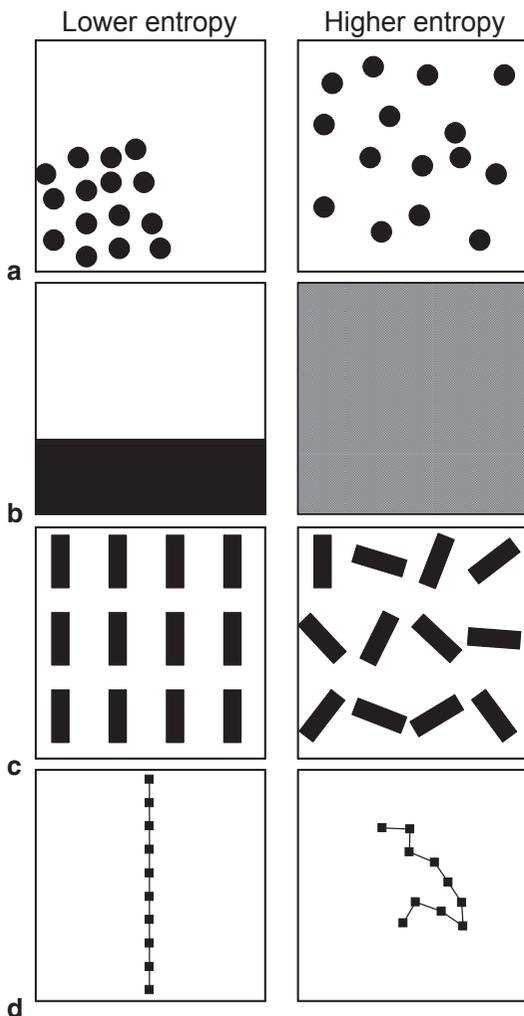
Fig. 1.2 Examples of molecular entropy. **a** Molecules have lower translational entropy if their position is somehow fixed. Molecules in a liquid or particularly a crystal have lower translational entropy than molecules in a gas. **b** A mixture of molecules will have lower entropy of mixing than two separate phases. **c** Nonspherical molecules have lower orientational entropy if they align with one another. **d** Conformational entropy drives flexible molecules to take on a random configuration. Note that in all cases, one ordered and one disordered state is shown, but in practice, there will be many more disordered states

the dice were weighted in some way to bias the outcome, and depending on whether the bias is for high or low numbers, the long-term average would be more or less than 3½. In this case, the preference of the dice for one outcome over the others means the most probable, highest entropy outcome is not the even distribution of numbers but rather the biased distribution of numbers.

What would we conclude if at the end of our expanding gas experiment, we found most of the molecules still in the starting cylinder and the remainder distributed about the other cylinders? Clearly, this outcome is very different from the even distribution expected from the second law, and just like with the weighted dice, something must be biasing the result. If there was some sort of attraction, a bond, between the molecules, then that could account for the bias. For example, if we repeated our experiments with sucrose molecules tightly bonded into a crystal, we would not expect any redistribution of molecules when the stopcock was opened, and the most likely outcome would be the highly biased one of all the starting molecules remaining in the starting cylinder. Similarly, if we used water molecules for our experiment, we might expect some water vapor in the other cylinders, but the majority would remain as a liquid in the first cylinder. There are bonds between water molecules, so we would get some preference for the closely associated state, but they are weaker than the intermolecular bonds in a sugar crystal.

Our original experiment shown in Fig. 1.1 is based around molecules free to exchange kinetic energy on collisions but with no potential energy to hold them together or push them apart (i.e., no chemical bonds). In the presence of a potential energy term, the second law does not predict an even distribution over all available states but rather a preference for the low-energy configurations. In the case of a system with two possible states separated by an energy difference of ΔE, the expected outcome is given by the Boltzmann distribution:

$$\frac{n_i}{n_0} = \exp\left(\frac{-\Delta E}{kT}\right) \qquad (1.4)$$

The ratio of the number of molecules in the high-energy state (n_i, e.g., nonbonded) to the number of molecules in the low-energy state (n_0, e.g., bonded) is given as the exponential of the ratio of the energy difference between the states (ΔE) to the thermal energy of the system (kT where

Fig. 1.3 The Boltzmann distribution. The ratio of high to low energy molecules (n_i/n_0) decreases as the energy difference (ΔE) increases or as the temperature (T) decreases. *Inset* shows distributions of molecules between two possible states separated by high or low energy differences
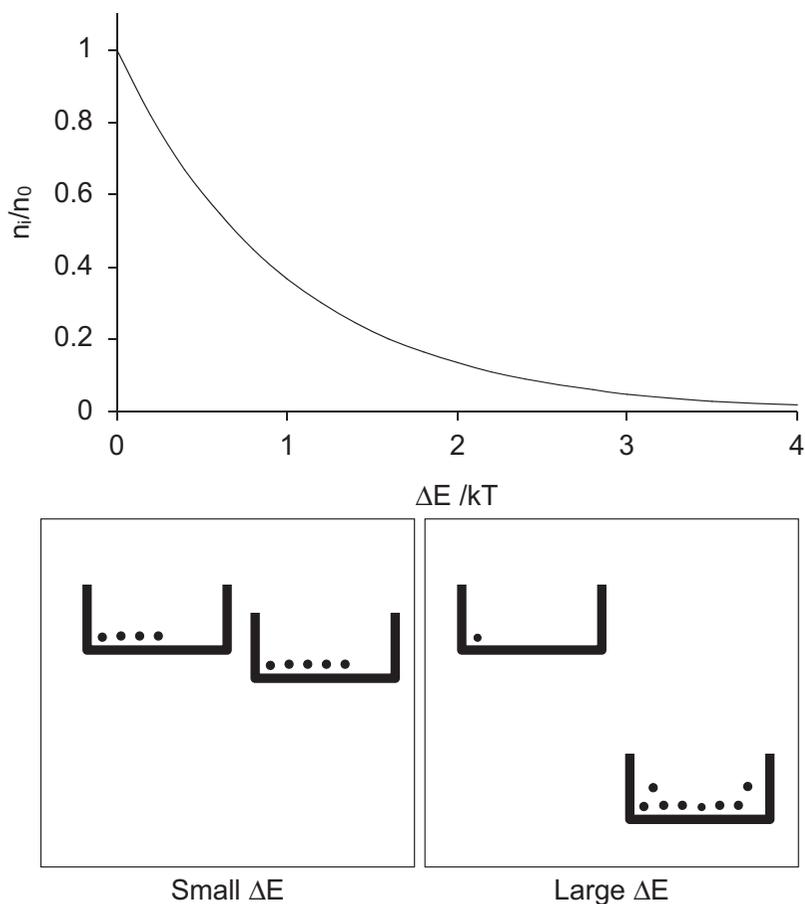
k is the Boltzmann constant and T is the absolute temperature[2]). If there is no energy difference, n_i/n_0 is 1 and, as we would expect from the second law in the absence of bonds, an even distribution of molecules between the two states (Fig. 1.3). However, as the energy difference increases or temperature decreases, we will see a greater and greater proportion of molecules in the low-energy state as potential energy (ΔE) becomes more importantly relative to kinetic thermal energy (kT). Again, this makes sense, for if we did our expanding gas experiment with water molecules at high temperatures, we would expect more water vapor and less water remaining in the liquid state.

Entropy brings a sense of time and irreversibility to physics, but we have to be cautious. Boltzmann's picture of the universe is of structure and order melting away to an inevitable grey chaos, but all around us, we can see order emerging from disorder in apparent violation of the second law: water can freeze into ice, plants grow and form structures from the elements surrounding them, and alcohol can be concentrated by distillation. In all of these cases, there is an obvious local decrease in entropy, but this is somehow facilitated by a movement of energy. In contrast to the examples we used to illustrate the inevitability of entropy increase, there were no such energy flows. To understand how our observations of *local* entropy decreases associated with energy changes can be reconciled with the

---

[2] Sometimes Eq. 1.4 is seen written with the gas constant R (=8.314 JK^{-1} mol^{-1}) in place of the Boltzmann constant. The gas constant is the product of Avagadro's number (=6.02×10^{23}) and the Boltzmann constant and is useful when the energy difference is expressed on a per-mole basis.

second law's requirement for *universal* entropy increase, we need to first define our local system.

1.5   Focusing on a System—the Enthalpy

The way we have approached the laws of thermodynamics so far is too broad to be really useful. We know that the total amount of energy is conserved and the total amount of entropy increases with any change, but these terms refer to the total energy and entropy in the universe. These are numbers we cannot hope to know when all we are really interested in are the contents of our test tube or package of food. We need to focus our concern on that tiny part of the universe we care about, the system, clearly define its boundaries, and then account for what moves across them. For example, if we were interested in the thermodynamics of boiling vegetables, we could choose to define our system as one plant cell, or as a single piece of vegetable, or as the vegetables plus the water they are boiling in, or as the vegetables, water, pan, and stove, etc. These and many other choices are all valid as long as we can clearly define a boundary and measure what moves through across it. The best choice often depends on where the measurements are easiest to make, so in this case, we might take the vegetables plus boiling water as our system. As steam is leaving as heat is applied, we must measure both mass loss and energy flow in. Alternatively, if we were heating a sealed can of vegetables, our system could still be the vegetables and water inside the can, but now it is isolated (i.e., no mass can cross the system boundaries) but not adiabatic (i.e., energy can still cross the system boundaries). At this stage, we will only consider isolated systems.

Energy can be transferred from one place to another as heat or as work. Therefore, because of the first law, the only way the total energy of our system (i.e., the internal energy, U) can change is if there is some flow of energy across its boundaries:

$$dU = dq + dw \qquad (1.5)$$

where dU is the change in internal energy, dq is the heat flow into the system, and dw is the work done on the system. Note the sign convention; flow of energy into the system is given a positive value, and flows out, a negative value. The power of Eq. 1.5 gives us something to measure. We may never know the total energy of our system (e.g., what is the energy of a can of vegetables?), but we can measure heat flows and we can measure work so we can calculate *changes* in internal energy.

From a practical point of view, heat flow is relatively easy to measure, as it can be inferred from changes in temperature, while work in its various forms is harder to keep track of. If it were possible to set up a system so no work is being done (i.e., $dU = dq$), then by measuring the heat flow across the system boundaries, we can measure the parameter we want, energy change. However, some forms of work are more easily discounted than others. Electrical work (movement of charged mass against an electrical potential) is important in batteries and sometimes in living cells but can usually be neglected in foods. Surface work (the enlargement of surface area against interfacial tension) can be important in some foods (e.g., foams and emulsions), but again in many cases it can be neglected. However, most materials, particularly gasses, expand and contract in response to changes in temperature, and this form of work is harder to avoid and must be accounted for.

If we could develop some complicated equipment to hold the volume constant during a reaction by changing the pressure, we could prevent the system doing work by expanding and contracting. We could then directly use measurements of heat flow to measure changes in internal energy. In practice, however, most changes we are interested in occur at atmospheric pressure, and some expansion and contraction work will occur. We must account for this in our measurements of internal energy change. The work done in expanding an ideal gas by a small amount (dV) against constant pressure p is $-p\,dV$. If expansion is the only work occurring, then we can substitute $-p\,dV$ for dw in Eq. 1.6, i.e.,

$$dU = dq - pdV \qquad \text{At constant pressure} \quad (1.6)$$

Now we cannot directly match the measured heat flow (dq) to internal energy changes (dU), and to get around this, we must instead define a new parameter—the enthalpy, H:

$$H = U + pV \qquad (1.7)$$

This definition initially seems arbitrary, but will prove useful. For small changes,

$$\mathrm{dH} = \mathrm{dU} + \mathrm{d(pV)} = \mathrm{dU} + \mathrm{pdV} + \mathrm{Vdp} \quad (1.8)$$

and since we are interested in constant pressure systems, $Vdp = 0$. Rearranging and substituting into Eq. 1.6 gives $dH = dq$, i.e., the heat flow at constant pressure is the change in enthalpy. We have related a thermodynamic property of the system (the enthalpy) to a parameter we can measure (heat flow) but this time under constant pressure conditions that are more experimentally realistic. The enthalpy term we created to enable us to do this is almost the same as internal energy for systems whose volumes change little (e.g., reactions in solid or liquid foods). In the following example, we will show how these definitions allow us to measure enthalpy changes associated with a change in food structure.

Fig. 1.4 Common corn starch granules imaged using optical microscopy
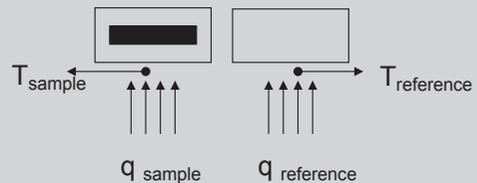
Fig. 1.5 Schematic illustration showing heat flux in DSC

Enthalpy changes in foods are often measured using differential scanning calorimetry (DSC). A small (~ 10 mg) portion of the sample is sealed into a cell and placed in a furnace and a second empty-cell blank in a similar adjacent furnace (Fig. 1.5). Our system is the pan and its contents (i.e., the metal of the pan, about 3 mg of starch, and 7 mg of water), so we must measure energy exchanges across its boundaries. The DSC is programmed to heat the sample and blank at the same controlled rate. The reference pan needs a certain amount of heat to warm the metal of the empty pan, but the sample pan needs more to heat the starch and water as well. The difference between the heat required by the two pans is measured, and from this the energy required by unit mass of the sample to cause unit change in temperature is calculated, i.e., the specific heat C_p:

Example: Calorimetry of Starch

Starch is made up of linear and branched polymers of glucose naturally existing in plants as semicrystalline granules several micrometers in diameter (Fig. 1.4, see also Sect. 7.5). When starch is heated to a critical temperature in the presence of water, the viscosity suddenly increases as the crystallinity is lost and the granules suddenly swell to many times their original size. This process is known as gelatinization and is an essential step in the cooking of starchy foods. If we were interested in the thermodynamics of gelatinization as most processing is done by altering the temperature at constant pressure, the relevant parameter is the enthalpy change.

Fig. 1.6 Specific heat of a 30% starch suspension in water calculated from a DSC measurement of differential heat flux. The peak corresponds to gelatinization and the loss of the granular structure seen in Fig. 1.4. The area under the curve corresponds to the enthalpy of gelatinization

$$C_p = \frac{dq}{dT} = \frac{dH}{dT} \quad \text{at constant pressure} \quad (1.9)$$

If an endothermic reaction occurs requiring an additional energy to facilitate the change, the instrument will measure the additional heat flux needed to maintain the programmed heating rate in the sample and record as a peak in the apparent C_p. On the other ha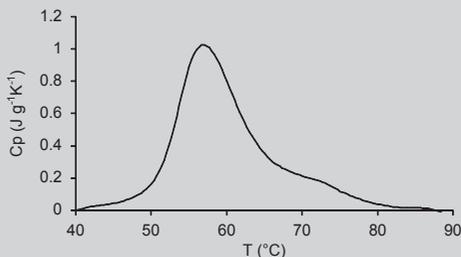nd, the heat released by an exothermic reaction would mean the sample requires less heat to warm it and the apparent C_p of the sample would be reduced.

Measured specific heat flow for a 30% suspension of corn starch granules in water is shown in Fig. 1.6. Between about 50 and 85 °C, there is a peak, implying that between those temperatures, the sample required additional heat energy to increase in temperature. This additional energy must have been used to drive some reaction; in this case, the gelatinization of the starch and the total amount of additional heat energy is given by the area under the peak ($\Delta H = \int C_p \, dT = 14.4 \text{ J g}^{-1}$). By carefully designing our system so that we could measure heat flow across its boundaries; we can measure the enthalpy changes associated with a change in food structure. Importantly, classical thermodynamics is a science of bulk observations—we can state with confidence that there was an enthalpy

change in the system at a characteristic temperature but make no inferences about what changes in starch structure correspond to this change in enthalpy. We only "know" the peak corresponds to granule melting when we open the cell after the experiment and examine its contents under the microscope.

By focusing on a defined system, we can use the first law to make measurements of the energy, or more precisely, enthalpy, changes in our system. We are now in a position to return to the confusing examples from earlier where we saw local entropy can decrease if there are energy transfers across the systems boundaries.

1.6 Free Energy and Equilibrium

A system at constant pressure can be defined in terms of its enthalpy and its entropy. It is helpful to combine these as a free energy, the fraction of the total energy available to do external work[3]. At constant pressure, this is defined as the Gibbs free energy[4]:

$$G = H - TS \quad (1.10)$$

A system with a high internal energy (large H) will be capable of doing a lot of external work unless it is very incoherent (high S) or hot (high T). From common experience, the only observed spontaneous changes are the ones capable of doing external work (e.g., balls falling, springs contracting). As Gibbs free energy is the maximal amount of work the system can do, the only change we can expect to see is those which result in a decrease in reducing this number. This idea is easier to understand if we rewrite Eq. 1.10 as

---

[3] Technically, nonexpansion work. A system at constant pressure will expand and contract to do work against the gas surrounding it.

[4] The Gibbs free energy is used for constant pressure systems; an analogous Helmholtz free energy is used for constant volume systems.

the change in Gibbs free energy of a system in response to a reaction (e.g., water freezing, plants growing, alcohol being distilled) at constant temperature:

$$\Delta G = \Delta H - T\Delta S \qquad (1.11)$$

Changes from the higher to lower Gibbs free energy ($\Delta G < 0$) are thermodynamically permissible, while changes from lower to higher Gibbs free energy ($\Delta G > 0$) will not occur.

Eq. 1.11 allows a decrease in system entropy (ΔS negative so $-T\Delta S$ positive) provided sufficient energy flows out of the system (ΔH is negative). We can test Eq. 1.11 with the special case where no energy crosses the system boundaries (i.e., $\Delta H = 0$ and $\Delta G = -T\Delta S$). In this case, the only way for ΔG to be negative is for ΔS to be positive, and we have returned to the simple model of the second law we used above in the absence of energy flows.

The Gibbs free energy is the single most important quantity we will use in the thermodynamics of food systems, as it points the direction in which changes will occur and the conditions when there is no free energy difference between the two states defines the point they are at equilibrium.

Example: Calorimetry of Starch (continued)
Knowing that at equilibrium, $\Delta G = 0$ and $T\Delta S = \Delta H$ (Eq. 1.11), we can use calorimetric measurements of enthalpy changes to infer the associated entropy changes. For example, in the aforementioned DSC measurements of starch gelatinization, we saw a peak in heat flow to the starch–water system with an onset of $50\,°C$ (323 K) and a total enthalpy change associated with the transition of 14.4 Jg^{-1}. At low temperatures, no gelatinization is seen, so $\Delta G > 0$, while at higher temperatures, the gelatinization reaction proceeds spontaneously, so $\Delta G < 0$. If we take the onset of the DSC peak as the temperature where gelatinized and granular starch are at equilibrium, then at 323 K, $\Delta G = 0$ and $T\Delta S = \Delta H$, so $\Delta S = 0.04$ J K^{-1} g^{-1}. The entropy increases

as a result of gelatinization, and this agrees with our model of partly crystalline starch granules melting and the polymers becoming more disordered.

1.7 Chemical Potential

Free energy is the most important parameter describing the driving force for chemical change; it provides the landscape over which chemical reactions proceed, ever seeking the lowest point. However, we are more used to talking about chemistry in terms of the composition of the system. We can relate Gibbs free energy to composition by dividing it into a fraction for every component present[5]—so if a glass of milk had a certain free energy, then a portion would come from each water molecule, a portion from each lactose molecule, a portion from each serum protein molecule, etc. The contribution of component A is expressed as the change in the total Gibbs free energy (δG) when a small amount of component A is added (δn_A moles), assuming the temperature, pressure, and concentration of all other components remain constant:

$$\mu_A = \left(\frac{\delta G}{\delta n_A}\right)_{T,p,n_{B,...}} \qquad (1.12)$$

Chemical potential is most useful to us in calculating how molecules will tend to move. There are many examples in foods where molecules move from one place to another, water will evaporate from a food surface during drying, oil will diffuse out from the nuts in a candy bar and soften the chocolate coating, and ice crystals will form from water as food is frozen. As we will see in all of these cases, the molecules

---

[5] Free energy, like mass and volume, are *intensive* parameters and thus depend on the amount of material present and can be divided up on a "per molecule" basis. Other parameters, e.g., density and temperature, are independent of the amount of material present (i.e., *extensive* parameters), and so it does not make sense to try and assign a fraction to each molecule.
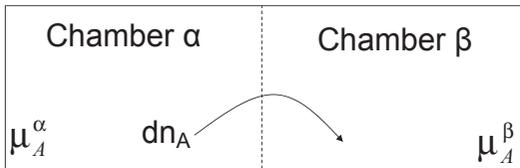
Fig. 1.7 Molecules will move from a region of high chemical potential to that of low chemical potential

are moving from a region of high chemical potential to a region of low chemical potential. We can generalize all of these problems as a container divided into two chambers (α and β) by a membrane that allows only one type of molecule through (the "A" molecules) while keeping everything else out (Fig. 1.7). The two chambers have different compositions, so that the chemical potential of the A molecules in the first chamber (μ_A^α) is greater than the chemical potential in the second chamber (μ_A^β). In our examples of transport processes in food, the two chambers are respectively the moist food and the air surrounding it (only water is allowed to move between them), the nuts and the surrounding chocolate (only nut oil is allowed to move between them), and the cold food and the ice crystals inside it (only water is allowed to move between them). The initial Gibbs free energy of the system is given by

$$G_0 = \left[\sum \mu_i^\alpha n_i^\alpha + \mu_A^\alpha n_A^\alpha\right] + \left[\sum \mu_i^\beta n_i^\beta + \mu_A^\beta n_A^\beta\right]$$
(1.13)

where the $\mu_i n_i$ terms refer to the chemical potential of all the other non-A components in the system. Next, we allow a small amount of A (dn_A) to move from the chamber α to chamber β. As we are only moving a small amount of A, we can assume the properties of all the non-A compounds in the system are unchanged and so the new chemical potential is

$$G_1 = \left[\sum \mu_i^\alpha n_i^\alpha + \mu_A^\alpha (n_A^\alpha - dn_A)\right] + \left[\sum \mu_i^\beta n_i^\beta + \mu_A^\beta (n_A^\beta + dn_A)\right]$$
(1.14)

Subtracting Eq. 1.13 from Eq. 1.14 gives the free energy change due to the transfer:

$$dG = (\mu_A^\beta - \mu_A^\alpha)dn_A \qquad (1.15)$$

So for the movement of molecules to be spontaneous (i.e., $dG<0$), $\mu_A^\alpha > \mu_A^\beta$. As we set out to show, molecules move from a region of high chemical potential to a region of low chemical potential. In our aforementioned diffusion, the chemical potential of the water in the air was lower than that of the water in the food for drying to occur, the chemical potential of the nut oil in the chocolate was lower than that of the nut oil in the nuts for diffusion to occur, and the chemical potential of the water molecules in the ice was lower than that of the chemical potential of the water molecules in the unfrozen liquid for freezing to occur.

A helpful way to think of chemical potential is as a "tendency to leave," and because it is related to the Gibbs free energy, it has an enthalpic as well as an entropic component. Molecules will try to maximize entropy while minimizing their bonding energy. We can see this balancing effect at work in the following example.

Example: Aroma of Food Emulsions
As we smell with our noses, perceived aroma is related to the concentration of aroma-active molecules in the headspace gas above the food but not to the same molecules that remain trapped in the food matrix. For example, even if a low-fat food is formulated with the same flavors as a high-fat product, it will frequently taste different because fats tend to bind up aroma-active molecules. Ghosh et al. (2006) studied this phenomenon using a series of model food aroma compounds, including ethyl heptanoate (EH). They added different amounts of EH to water and allowed it to come to equilibrium with a known volume of air, then used gas chromatography to measure the headspace concentration. Whatever the amount
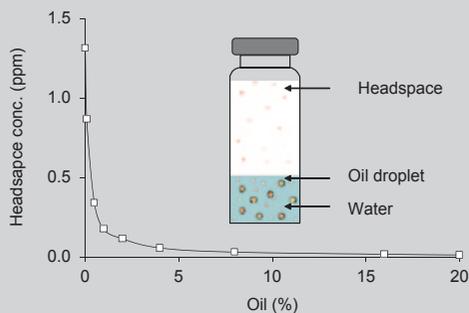
Fig. 1.8 Concentration of EH in the headspace above an emulsion as a function of emulsion oil content. *Inset* shows a diagram of the experimental system. Adapted with permission from Ghosh et al (2006). Copyright 2006 American Chemical Society

of aroma or volume of water or gas used, the ratio of concentration of EH in the two phases studied was a constant, that is, the equilibrium constant of the reaction expressed as a partition coefficient,

$$K_{hw} = \left[EH_{headspace} \right] \big/ [EH_{water}] = 1.8 \times 10^{-2}.$$

When the same experiment was conducted with oil in place of water, the concentration in the headspace was much lower, but again there was a characteristic ratio between the concentrations of EH in the different phases,

$$K_{ho} = \left[EH_{headspace} \right] \big/ [EH_{oil}] = 6.0 \times 10^{-6}.$$

If there were no chemical interactions, the maximum entropy would be the case where the EH is evenly distributed between the phases and both partition coefficients would equal 1. Clearly, this is not the case; there must be some sort of bonding that reduces the chemical potential of EH in the liquid phase and limits its partitioning into the headspace gas. As $K_{ho} < K_{hw}$, there must be stronger bonding between EH and oil than that of between EH and water. The consequence of the strong "preference" of EH for oil over water was seen when Ghosh measured the concentration of EH in the headspace above emulsions prepared with different oil-to-water ratios (Fig. 1.8).

Even relatively small amounts of fat acted as a "sink" for the aroma and reduced the amount partitioning into the headspace gas. EH has a strong, slightly unpleasant smell. If it was present as a contaminant in full-fat milk ($\sim 4\%$ fat), its headspace concentration and, therefore, its impact would be much less than that if it were present in skimmed milk ($< 0.1\%$ fat).

The chemical potential rather than the absolute concentration of the molecules in our food emerges as the important parameter defining a tendency to react. However, we prepare our foods with recipes that specify concentrations of ingredients. We need a way to relate concentration with chemical potential.

1.8 Solutions

Our approach to calculate the chemical potential of a solution as a function of its composition will be somewhat indirect. We start with the chemical potential of ideal gas mixtures, not a food but something we can understand well from the gas laws, then by analogy relate their properties to those of first ideal and finally real solutions and foods. From the gas laws, the chemical potential of an ideal gas increases with pressure as

$$\mu_A = \mu_B + RT \ln \frac{p_A}{p_B} \qquad (1.16)$$

where μ_A and μ_B are the chemical potential at pressures p_A and p_B. As usual, T and R are the absolute temperature and the gas constant respectively. At least qualitatively, this makes sense: molecules will move from regions of high pressure to regions of low pressure, and so a higher pressure means a higher chemical potential. The chemical potentials of many gases have been measured at 1 atm pressure and are tabulated as the standard state of the gas (designated by o):

$$\mu_A = \mu_A^\circ + RT \ln p_A(atm) \qquad (1.17)$$

Using Eq. 1.17 and a known value of the chemical potential of the gas in the standard state, we can calculate the chemical potential at any other pressure.

If we had a mixture of gases, the chemical potential of each component would depend on its partial pressure, that is, the pressure of the gas due to that component alone. The partial pressure of a component in a gas is proportional to its concentration on a mole fraction scale, that is $p_A = x_A p_{total}$, where p_{total} is the total (measured) pressure of the gas, and x_A is the mole fraction of component A. In the ideal case, we are assuming that molecular interactions are not important, and as the average molecular separations in a gas are large, the molecules do not interact and their behavior is reasonably ideal.

However, we are more interested in the properties of mixtures of solutes and solvents than with mixtures of gases. As pressure makes little difference to the properties of a condensed phase, we can neglect the total pressure term and rewrite Eq. 1.17 to give chemical potential of a component of a mixture as a function of its concentration on a mole fraction scale:

$$\mu_A = \mu_A^\circ + RT \ln x_A \quad ideal\ solutions \qquad (1.18)$$

where μ_A° is the chemical potential of the pure A. Again, we can test this equation against our everyday qualitative observation that molecules tend to diffuse from regions of high concentration to regions of low concentration, which corresponds to high chemical potential to low chemical potential.

This formulation is valid for ideal solutions where chemical interactions play no role. In a solution, this is less realistic than in a gas mixture because the molecules are close to one another and do interact. We can modify Eq. 1.18 to take account of chemical interactions by replacing the concentration term with activity (a_A):

$$\mu_A = \mu_A^\circ + RT \ln a_A \quad real\ solutions \qquad (1.19)$$

Activity is the "effective" or thermodynamic concentration. In Eq. 1.19, we are saying that whatever the actual concentration of A in our mixture (x_A), it behaves the same as an ideal solution of concentration a_A. Activity is sometimes expressed as the product of the actual chemical concentration and an activity coefficient (γ_A):

$$a_A = \gamma_A x_A \qquad (1.20)$$

The activity coefficient is not a constant, but an empirical parameter describing the extent to which reactivity is affected by molecular interactions. Usually, $\gamma_A < 1$, for example, if we tried to fortify a drink with calcium but half of our added ions were bound by protein in the drink, then the effective solution concentration of Ca^{2+} would be half of what we expected, that is, $\gamma_{Ca^{2+}} = 1/2$. Activity coefficient tends to decrease with increased concentration as molecular interactions become more important. In very dilute solutions, activity approaches concentration and $\gamma_A = 1$, Eq. 1.19 reduces to Eq. 1.18, and the solution behaves ideally.

Molecules will move until, at equilibrium, their activities and thus chemical potentials in all phases are equal, even though their concentrations may not be. Thus, in the aroma partitioning example discussed earlier, there was much greater concentration of EH in the water than in the headspace while the activity and chemical potential of EH in both phases was the same. Furthermore, as the EH in the gas phase was presumably behaving ideally (large molecular separations, relatively low concentrations), then the measured concentration in the gas was equal to the activity in the gas and hence at equilibrium the activity in the water phase. Perhaps the most commonly exploited activity measurement in foods is water activity, and this is the topic of the next section.

1.9 Water Activity

By controlling water, many of the spoilage reactions in food, including microbial growth, cannot occur, and the shelf life is extended. Because it is a measure of chemical potential, the activity of water in a food is more relevant to its availability for spoilage reactions than the actual amount of water present. The rate of reactions in the aque-
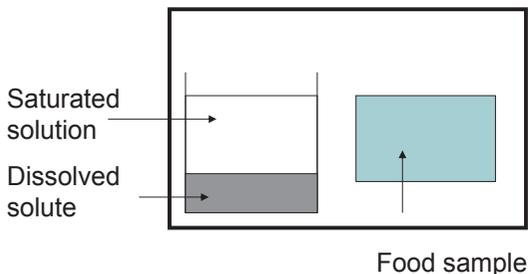
Fig. 1.9 Schematic illustration of the processes used to bring food to a defined water activity. The container is sealed, and the contents are allowed to come to equilibrium so the water activity in all phases is constant and buffered to the water activity of the saturated salt solution

Table 1.1 Equilibrium relative humidity of selected saturated salt solutions at 25 °C (from Wolf et al. 1984)

Chemical	RH(%)
Barium chloride	90.3
Lithium chloride	11.2
Magnesium chloride	32.8
Potassium acetate	22.6
Potassium carbonate	43.8
Potassium chloride	84.3
Sodium bromide	57.7
Sodium chloride	75.3

ous phase, including enzyme–catalyzed reactions and microbial growth, typically decreases rapidly with water activity.

Water activity is readily measured by sealing a few grams of food in a small pan and allowing the food to come to equilibrium with the air surrounding it. At equilibrium, the activity of water in the food is equal to the activity of water in the air, and as most gases are reasonably ideal, the activity of the water in the air is equal to the partial pressure of the water vapor in the air. Different water activity meters measure the humidity of air in different ways, but a popular design is to chill a mirror electrically until it reaches the dew point of the air (a function of air humidity) and fogs over.

Food can be brought to a known water activity by allowing it to reach equilibrium with air of known humidity. This is commonly done by placing a sample of food in a sealed chamber also containing a saturated salt solution with some excess undissolved salt crystals present (Fig. 1.9). The activity of water in a solution decreases with the mole fraction of the solution in a solute-dependent manner (Eq. 1.19). Thus, saturated solutions of different salts have different water activities (Table 1.1). The presence of the undissolved salt crystals serves to buffer the composition of the solution phase; if some water evaporates, more crystals will form and the solution will remain saturated and the water activity is unchanged. Alternatively, if some water is absorbed by the solution, some crystals will dissolve yet the solution remains saturated and again the water activity is unchanged. The satu-

rated salt solution fixes the relative humidity of the air in the chamber, and the food will either gain or lose moisture in response until it reaches the same water activity as the saturated salt solution.

Example: Moisture Sorption Isotherms of Apples

Sá et al. (1999) freeze-dried apple slices to remove all the water present. They then sealed them in chambers containing different saturated salt solutions (Fig. 1.9) and allowed the chambers to come to equilibrium. By measuring the amount of mass gained per gram of dried apple and plotting it against the water activity of the salt solutions, Sá et al. could measure a moisture sorption isotherm[6] for the apple (Fig. 1.10).

A first thing to note about Fig. 1.10 is that it is incomplete—it does not show the fresh, fully hydrated fruit. Fresh fruit has perhaps 85 % moisture (or about 5.6 g of water per gram of solids on a dry weight basis), but the *y*-axis in Fig. 1.10 only reaches approximately 1.4 g/g (about 58 % water). Therefore, most of the water in the apples is removed only by reducing the

---

[6] This is a sorption isotherm, as they started from a dry material and added water. It would have been equally valid to start with a moist product and allow it to dry down to the different water activities, in which case they would measure a moisture desorption isotherm. There is often a hysteresis between sorption and desorption isotherms.
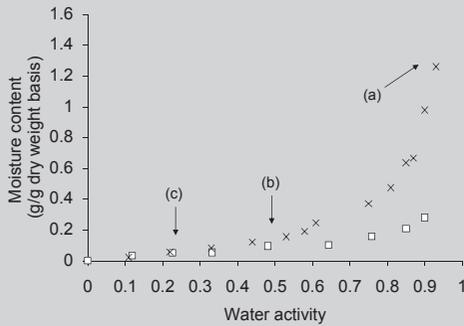
Fig. 1.10 Moisture sorption isotherms for apple slices (×, from Sá et al. 1999) and cornflakes (□, from Azanha and Faria 2005). The letters indicated correspond to the highly idealized microstructures proposed in Fig. 1.11
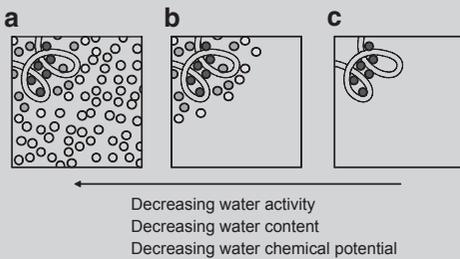

Fig. 1.11 Schematic diagram showing the changing interactions between water and solute molecules as a food is dried. Water molecules are represented as *circles* with the *depth of shading* representing the degree to which they interact with the food molecule. Note that this diagram shows a static and idealized cartoon of a real food. In a real food, water is highly mobile and individual molecules will move between different states of binding very quickly. **a** In a moist food, most of the water molecules are far from the solute molecules and unaffected by their presence. The water activity of these molecules is close to that of pure water. **b** In a semi-dried food, the remaining water molecules are somewhat affected by the solute and have a lower activity/chemical potential than pure water. **c** In a "dry" food, the only remaining water is very tightly bound by the solute molecules and will be extremely difficult to remove

water activity of the environment to 0.93 (the highest water activity saturated solution used). The majority of the moisture in most fresh fruits and meat has an activity close to 1 and properties similar to pure water. If fresh-cut produce is left exposed

on anything but the most humid of days, it will quickly lose this moisture to the air, dry out, and wilt. We could imagine the moist food as few food molecules (e.g., sugars, acids, pectin, cellulose) surrounded by a large amount of water (Fig. 1.11a). In this case, the vast majority of the water molecules are far from the solutes, and thus their properties are unaffected by them. These water molecules have chemical potential close to pure water and are easily driven off if the food is dried (i.e., a high water concentration means a high water chemical potential and a high "tendency to leave").

The isotherm itself shows how strongly the apple slices bind the remaining water. The different chambers can be seen as different "stresses" on the water; a lower buffered humidity in the gas means water with a lower chemical potential can be removed. The isotherm is relatively flat at lower moisture contents, so while there is little water left in the food, it is increasingly tightly bound by the foods, i.e., has a lower chemical potential. This is illustrated schematically in Figs. 1.11b and c; water molecules closer to food molecules are more tightly bound, and the variety of chemical structures present in the apple tissue means that some water is weakly affected by the fruit and some is strongly affected and has a very low water activity.

Moisture sorption isotherms can be used to compare the interactions of different foods with water. In Fig. 1.10, the moisture sorption isotherm of cornflakes is shown alongside that of apple. At a given moisture content, apple typically has a lower water activity than cornflakes. Therefore, the apple molecules are binding the water molecules more tightly and lower their chemical potential more than the cornflake molecules. If the chemical potential of water in the apples is lower than that in the cereal, moisture will tend to diffuse from one to the other even though the concentrations are the same.

Example: Moisture Migration in a Breakfast Cereal

Breakfast cereal is often packaged with dried fruit pieces. An entrepreneur with limited understanding of food chemistry might try to make a prototype product by just buying dried apples and cornflakes and packaging them together. What problems would they encounter and how might they be solved? Dried apple slices bought from the store typically have a moisture content of about 30 % (0.44 g water per gram of dry solids), while cornflakes typically have a moisture content of about 3.5 % (0.03 g water per gram of dry solids). According to the moisture sorption isotherm (Fig. 1.10), the water activity of the apple would be about 0.75, while that of the cereal would be close to 0. Water would diffuse from the apple to the cereal until the water activities were the same and there is no chemical potential gradient. A small gain in water content would lead to a large increase in water activity for the cereal because the isotherm is relatively flat in that region, while a similar loss in water from the apples would lead to a smaller loss in water activity (the actual magnitude of the water transfer would depend on the relative amounts of apple and cereal). When cereal gains water, it tends to soften and the prototype product would quickly become unacceptable. If the apples could be dried to a lower moisture content, then it might be possible to match the water activity of the cereal and eliminate the diffusion problem. This solution is sometimes seen in products where freeze-dried pieces of fruit are mixed with cereal. The very low moisture fruit is hard and brittle, but the freeze-drying process maintains an open and crunchy texture quite acceptable to the consumer. If the manufacturer wants to use a soft and chewy fruit, it would be necessary to lower its water activity in some other way, and this might be done by adding a humectant (e.g., glycerol) that will bind up the water in the fruit and lower its chemical potential without unacceptably changing its taste or texture. An alternative solution would be to coat the fruit with some water-impermeable layer (e.g., a thin coating of oil or sugar) to slow down the diffusion process and provide a kinetic limitation to a thermodynamically favorable process.

1.10 Summary

We began this chapter by identifying entropy as a measure of disorder and energy in terms of the bonds between molecules. The laws of thermodynamics mean that any change must be accompanied by a net increase in the former while conserving the latter. Within a system, a defined region of the universe, entropy, and enthalpy can be combined as the free energy. Any reaction within the system must be accompanied by a decrease in free energy. Chemical potential is a way of distributing the free energy between the components of the system and gives us a more useful rule for predicting chemical change: molecules will move from a region of high to a region of low chemical potential. Finally, we set about measuring chemical potential by defining standard states and calculating activity coefficients—a measure of the interactions between molecules.

This chapter was concerned with understanding the driving forces for molecules to react and reach equilibrium. We did this with only the vaguest description of what molecules are and how they interact. In the next chapter, we will look at the molecular building blocks of food in more detail.

1.11 Bibliography

Thermodynamics is a subject where it is easy to lose the wood in the trees, and it is useful to find some combination of formal mathematical

argument with more qualitative explanation of what it all means. "Atkins' Physical Chemistry" (Atkins and De Paula 2006), especially Chaps. 1–7, provide detail on the arguments of classical thermodynamics while maintaining a clear, readable narrative. For the statistical, molecular perspective, Dill et al.'s (2003) "Molecular Driving Forces" is excellent. While both of these books provide a strong and complementary grounding in the basics of thermodynamics, it is easy to lose the "big picture" importance of the subject. Two good popular science books capturing the power and meaning of thermodynamics are "The Four Laws," again by Peter Atkins (2007), and "Into the Cool" by Schneider and Sagan (2005).

Other books look at the thermodynamics of biological systems, and this focus is obviously more relevant to foods. I have found "Physical Chemistry—Principles and Applications in Biological Sciences" (especially Chaps. 2 and 3) by Tinoco et al. (2002), "Biological Thermodynamics" (especially the biological examples in Chaps. 1–5; Haynie 2001), and "Thermodynamics and Kinetics for the Biological Science" (Hammes 2000) particularly helpful. Walstra (2003) gives a general introduction to thermodynamics of food in Chap. 2 of "The Physical Chemistry of Foods" and describes water activity in much more detail in Chap. 8.

Molecules

2

2.1 Introduction

> Everything is made from atoms. That is the key hypothesis. The most important hypothesis in all of biology, for example, is that everything that animals do, atoms do. In other words, there is nothing that living things do that cannot be understood from the point of view that they are made of atoms acting according to the laws of physics
>
> The Feynman lectures in physics, Vol. 1, 1963 (pp. 1–8).

Richard Feynman's hypothesis is the core of our goal as food chemists; we want to be able to relate all of the properties of foods to the atoms they contain. For some questions, this is fairly straightforward ("this fat is harder because it is more crystalline") while others are so complex we struggle to even frame them in terms of chemistry ("why does this sauce taste creamier than that one?"). However, in principle if we can properly understand how the atoms are behaving, we should be able to explain any behavior of food.

A common approach to many problems in science is to divide the subject up into a hierarchy of structures and focus only on the most relevant. For example, an engineer might notice that when a building collapses individual bricks are still intact in the rubble. From that observation, it would be sensible to study the cement holding the bricks together rather than the strength of the bricks themselves. By analogy, most of the physical changes in foods involve changes in the arrangements of molecules rather than the breaking and making of bonds within molecules. Therefore,

the atomic scale is far less important to most of our physical problems than the molecular scale and we can treat molecules as the building blocks of our food, reframing Feynman's hypothesis as:

Everything that food does, molecules do.

So what do molecules do? Their behavior is governed by the laws of thermodynamics described in the last chapter but to properly relate chemical behavior to chemical structure we need to understand the nature of kinetic and potential energy at the molecular level. Molecules have kinetic energy because of their masses and velocities while potential energy results from intra- and intermolecular bonding. In this chapter, we will start by considering molecular movement then look at bonding. We will finally return to Feynman's hypothesis and look at some ways that molecular properties can be related to bulk properties of a food.

2.2 Molecular Motion

Each molecule has a kinetic energy equal to ½ kT in each direction (x, y, and z) or 3/2 kT overall ($\sim 2.75 \times 10^{-21}$ J). In a gas, this energy leads to very fast molecular motion, approaching the speed of sound for many molecules at room temperature, but in a liquid the molecules are very densely packed and their movement is limited by interactions with their neighbors. They will move away from their starting position, but only slowly as they collide frequently with other molecules exchanging momentum and changing

direction. The net effect of the multiple collisions is that the trajectory of a moving molecule is a random walk—a series of small steps where the direction of each is not affected by previous steps (Fig. 2.1a). Because each step is in a random direction, it is as likely to take a molecule up as it is down or left as it is right. Therefore the average net displacement is zero after a random walk of any length. However, the combination of steps is unlikely to take the molecule exactly back to its starting point and most random paths will end a certain distance from the starting position. In a random walk, the average displacement is proportional to the square root of the number of steps taken and hence the square root of time. Figure 2.1b shows the average displacement for a molecule after 1, 5, and 10 steps; the direction of movement is unknown but molecules will move slowly away from their starting positions.

Despite the random progression of an individual molecule, the net effect will be to move from regions of high concentration to those of low concentration. (As we saw in the previous chapter, activity rather than concentration is the real driving force for diffusion as it also incorporates molecular interactions that can hold molecules together or force them apart. However, for our treatment of molecular motion we will continue to discuss concentrations for the ideal case where there are no molecular interactions.) This effect is merely statistical—imagine a box containing two chambers separated by a window, any one of the molecules has a statistical possibility of moving through the window into the other chamber over a given time period and there will be a constant exchange of molecules between the chambers (Fig. 2.2). If there was, say, a 1 % chance of the random walk of a given molecule taking it through the window in a given second and there were 100 molecules in the right-hand chamber and 1000 in the left then in an typical second 1 molecule would move right to left and 10 from left to right—a net movement of 9 molecules from left to right. A smaller concentration gradient would lead to a smaller rate of mass transfer, for example if there were 600 on one side and 500 on the other, the net rate of exchange would only be only 1. Each exchange reduces the

a

b
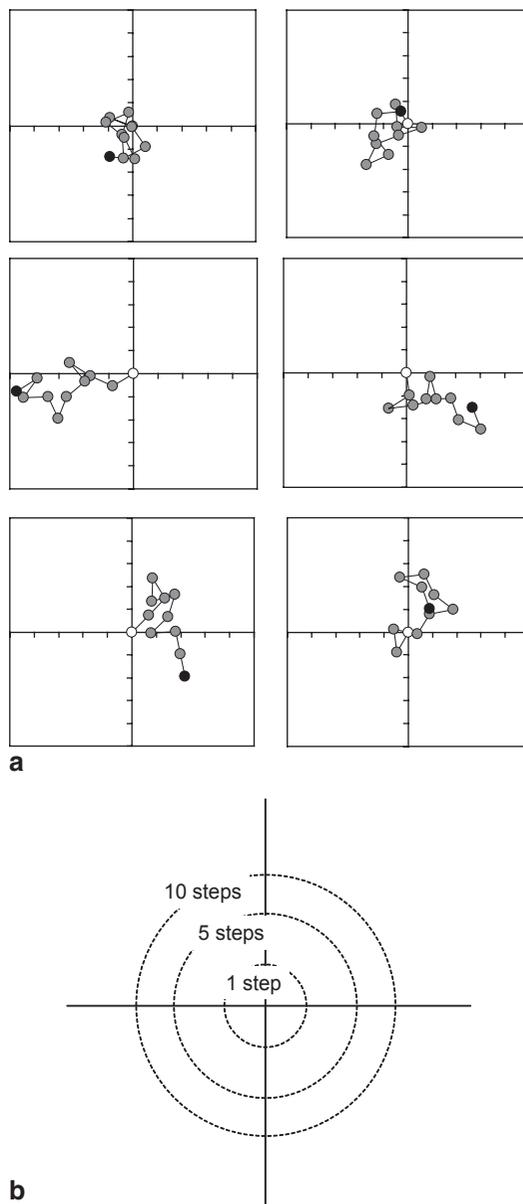
Fig. 2.1 a Example 10-step molecular random walks. The starting position shown at the origin as an *open point*, the final position as a filled point and intervening steps as *shaded points*. The molecule moves in a straight line until it collides with another molecule (not shown) and moves off in another random direction. **b** The average distance away from the starting position increases with the square root of the number of steps taken but the average net displacement after the walk is zero as random movements in one direction are cancelled by random movement in the opposite direction

concentration gradient so the rate of exchange will decrease over time. The net movement of
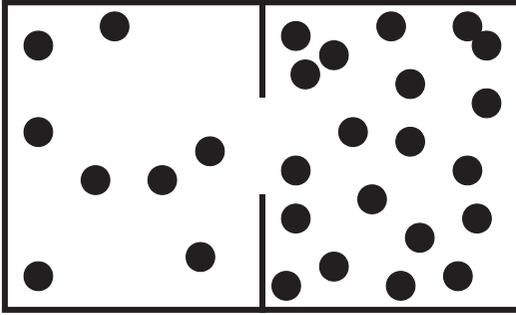
Fig. 2.2 A box of molecules separated into two chambers by a window. Each molecule is following a random walk so each has an equal chance of passing through the window in a given time period. The rate of diffusion through the window is proportional to the concentration difference between the chambers

molecules through a window of unit area per unit time is the flux (*J*) and, as our thought experiment has shown, is proportional to the concentration gradient (*dc/dx*) (i.e., Fick's first law):

$$J = -D\frac{dc}{dx} \tag{2.1}$$

(Note the negative sign because flow is from high to low concentration.) The proportionality constant, *D*, is the diffusion coefficient of the molecules. The diffusion coefficient can be measured experimentally, typically by measuring changes in local concentration over time or by gradient field nuclear magnetic resonance (NMR).

Albert Einstein related the macroscopic phenomenon of diffusion to the microscopic random walk:

$$D = \frac{kT}{6\pi\eta r} \tag{2.2}$$

where η is the viscosity of the material the molecule is diffusing through, *r* the effective molecular radius and kT the thermal energy of the system (i.e., the product of the Boltzmann constant and absolute temperature). A molecule diffuses more slowly through a viscous material and a larger molecule will feel more drag than a smaller one and diffuse more slowly. Equation 2.2 can be used to relate measurements of diffusion coefficient to molecular dimensions. For example, the self-diffusion coefficient of water (the capacity

of a water molecule to diffuse in other identical water molecules) is measured by NMR as about $2.5 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$. Taking the viscosity of water as 9×10^{-4} Pa s, the diameter of a water molecule comes out as 2 Å, which is reasonably close to the value of about 3 Å from molecular modeling. However, care must be taken using Eq. 2.2 in this manner. Firstly, polymers and ions are frequently highly hydrated so several water molecules will be entrained with the diffusing molecule and move along with it. Consequently, the effective size measured for the diffusing polymer or ion will be that of the molecule of interest plus the hydration layer. Secondly, viscosity measured at the bulk scale with fluid flow measurements (see Chap. 7) may not correspond to the viscosity causing drag on the diffusion molecules at the microscopic level. For example, the diffusion coefficient of sucrose molecules decreases with sucrose concentrations due to the increasing solution viscosity. However, adding a small amount of xanthan gum causes no significant change in sucrose diffusion coefficient despite a large change in measured viscosity (Basaran et al. 1999). This discrepancy is probably because the xanthan polymer can spread out and make the bulk solution viscous (see Chap. 7), there are still large pores and the sucrose molecules are free to move through the gaps and are not affected by them.

As solution viscosity increases, the diffusion coefficient and hence the mobility of the molecules will decrease. Viscosity increases with increasing concentration and decreasing temperature. At a characteristic temperature and concentration, viscosity reaches a level that no molecular translational movement is possible at which point the liquid is said to have entered a glassy state. Glassy materials are hard and brittle because the molecules cannot flow past one another in response to applied force and instead just shatter. The rates of chemical reactions are very slow in the glassy state as for molecules to react they must first diffuse through the solution to come into contact with one another. We will return to the glass transition in the context of crystallization in Chap. 6.

2.3   Bonding and Molecular Structure

Molecular motion is random and will tend to increase entropy by evenly distributing molecules in space. Any structure we see must therefore arise from forces acting between atoms and holding them in a preferred arrangement. We used gravity as an example of a force in the first chapter but gravity, although very long range, depends on the masses of the objects involved. While it dictates the movement of heavy objects we see at the macroscopic level (e.g., throwing a ball), the tiny masses of atoms means the gravitational contribution to chemical bonding insignificant. However, gravity is just one of the four fundamental forces of the universe. In addition, the strong and weak nuclear forces act with great strength at very short ranges and are responsible for the properties of the atomic nucleus. However, as the atomic nuclei do not change in foods, nuclear forces are irrelevant and we are left with electrostatic forces, the mutual attraction of like charges and repulsion of unlike charges, as the sole remaining interaction responsible for all chemical bonding.

Electrostatic forces are responsible for all of the different types of chemical bonds. Whether they occur between atoms within a molecule (e.g., covalent bonds) or between different molecules (e.g., Van der Waals forces) are all just manifestations of this same underlying interaction. Having acknowledged the central mechanism for bonding, it is still helpful to divide bonds into different subcategories and then focus only on those most important to the problem in hand. With this in mind, we will divide the general phenomena of bonding into bonds holding atoms together as molecules and bonds between molecules (i.e., intramolecular and intermolecular bonds). We will briefly review the bonds holding a molecule together with a view to understanding the types of building blocks that will interact with one another via intermolecular forces to produce food structure.

Atoms consist of a tiny, massive, and positively charged nucleus associated with sufficient negatively charged electrons to neutralize the overall charge. (An atom or molecule whose positive nucleus is not balanced with electrons is an ion). The position of the electrons cannot be stated precisely, but quantum mechanics can predict the atomic orbital—the region of space close to the nucleus where the electron is likely to occur. There is one first-level orbital ($1s$), four second-level orbitals ($2s$, $2p_x$, $2p_y$, and $2p_z$), and four third-level orbitals (3_s, $3p_x$, $3p_y$, and $3p_z$) sometimes known as the first, second, and third electron shells. Each orbital can contain up to two electrons, and as the atom gets larger it will fill up the orbitals from lower to higher energy, for example, hydrogen has one electron which is typically in the 1s orbital, helium has two electrons so both are in the 1s orbital—filling it. Carbon has six electrons so the $1s$ orbital is filled and the remaining four electrons half fill the second-level orbitals. A bond is when the orbitals from two atoms combine to form a molecular orbital with the pair of electrons distributed between the atoms. The properties of the bonding orbitals for simple molecules can, in principle, be calculated using quantum mechanics but here it will suffice to take a simple approach and merely note some of the important features of covalent bonds.

- **Fixed Valency.** Each type of atom tends to form a characteristic number of bonds (i.e., the valency) governed by the number of electrons needed to fill the outer electronic shell. Thus, hydrogen with one electron needs a second to fill its first-level orbital and achieves this by forming one bond, helium has two electrons so its first level orbital is already full and tends not to form bonds. Carbon has four electrons in its outer shell and must form four bonds to fill it. It is possible to form multiple bonds between two atoms when more than one pair of electrons is shared between them. For example, carbon can form one, two, or three bonds with another carbon atom to form the backbone of ethane, ethene, or ethyne (Fig. 2.3).

- **Polarization.** If the electron pair in the bonding orbital is evenly distributed between the two atoms, the bond is nonpolar, but if one atom has a greater affinity for electrons it will tend to draw them closer, leaving the
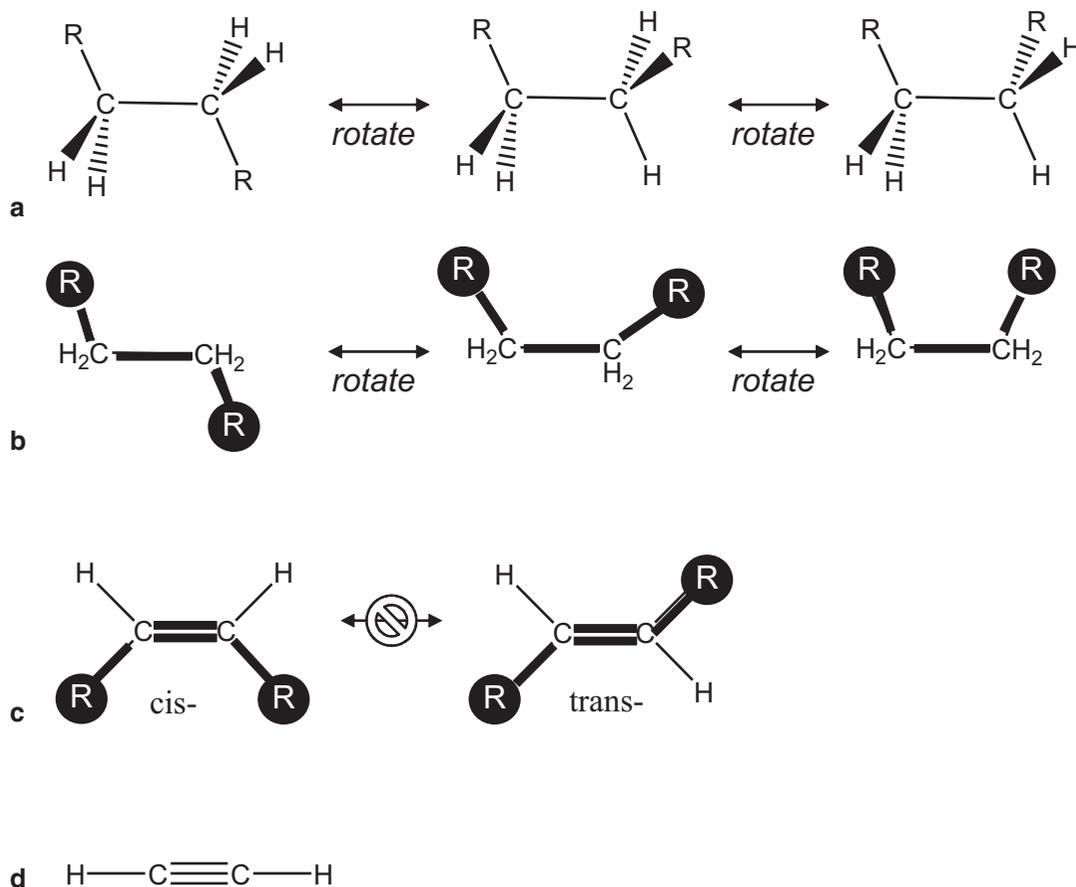
Fig. 2.3 Single, double, and triple bonds structures as illustrated by **a** ethane, **b** ethene, and **c** ethyne. Some of the hydrogens are labeled as R-groups to illustrate the changes in conformation due to rotation about the carbon–carbon bond. The single bonds in ethane are free to rotate, profoundly changing the shape of the molecule (shown as a 2D projection in b). An energy barrier restricts rotation about double bonds

distribution skewed and the bond polarized. The atom with the greater share of the bonding electrons accumulates a fractional negative charge ($\delta-$) leaving the other atom with a slight positive charge ($\delta+$). The partial charges on the atoms in a molecule are readily calculated by most chemical drawing programs (e.g., the charges on water in Fig. 2.4 were calculated using MarvinSketch program from ChemAxon Kft., Hungary). The electron affinity of atoms can be expressed as electronegativity on the Pauling scale (Table 2.1); if a bond links two atoms, the electrons will tend to accumulate on the atom with the higher Pauling value and gain a partial negative charge. If the bond is very highly polarized,

the electrons will be effectively entirely associated with the more electronegative group which will gain a permanent negative charge (i.e., an anion) leaving the other group with fewer electrons than needed to provide charge neutrality (i.e., a cation). The degree of ionic character to a bond can be calculated as half the absolute value of the difference between the electronegativities of the atoms involved. For example, a carbon–hydrogen bond is $|2.6-2.2|/2 = 20\%$ ionic while a carbon–oxygen bond is $|3.5-2.6|/2 = 45\%$ ionic and a sodium–chloride bond is $|0.9-3.15|/2 = 112.5\%$ ionic (note—values greater than 100 % are taken as completely ionic bonds).
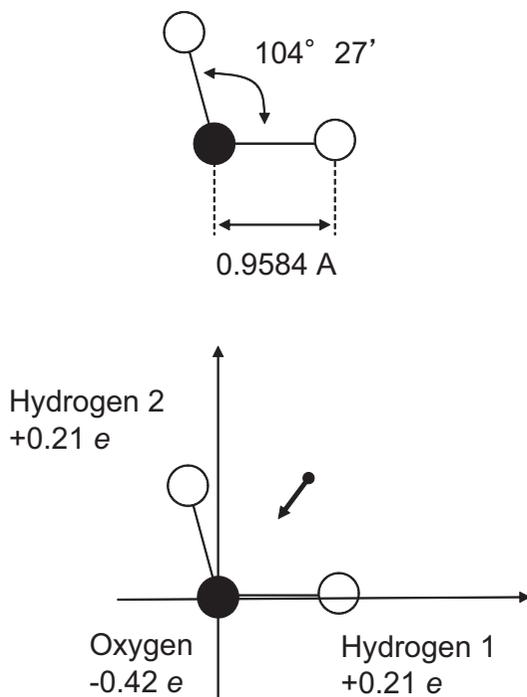
Fig. 2.4 Water (including dipole and bond angles). The partial charges on each atom are calculated by the structure drawing program (in this case Marvin from ChemAxon Kft., Hungary). *Inset arrow* is the equivalent dipole

Table 2.2 Covalent bond strengths and lengths (Haynes et al. 2013; Israelachvilli 1991).

	Bond strength		Bond length
	(kJ mol^{-1})	(kT at 300 K)	(Å)
C–C	360	144	1.54
C=C	600	241	1.34
C=O	340	136	1.23
C–H	430	172	1.09
H$_2$O H-bond	6–23	2.5–9	1.97

from the electrons in the bonds repelling one another. Thus, the carbon–hydrogen bond and the carbon–carbon triple bond in ethyne repel one another to give the bond angle of 180° (Fig. 2.3d). Similarly, the carbon–carbon double bond and the two carbon–hydrogen single bonds in ethene also repel one another resulting in a planar molecule with bond angles of 120° (Fig. 2.3c). Not all molecules are flat; the four bonds around each carbon in ethane repel one another to give a tetrahedral shape (bond angle 109.5°, Fig. 2.1a). Lone pairs of electrons (full outer shell orbitals not contributing to covalent bonds) also repel to one another as well as any bonding electrons so the bond angle in water (104.5°, Fig. 2.4) is closer to tetrahedral than to linear because oxygen has two lone pairs of electrons as well as two bonds.

We can get a sense of the strength of covalent bonds by comparing the bond energy to the thermal energy of the system. Bond energy means the amount of energy you need to put in to break the bond and thermal energy is the kinetic energy of molecules due to heat. As we saw in the previous chapter, thermal energy is given by kT, so at room temperature it is about 4.1×10^{-21} J ($= 1.38 \times 10^{-23}$ JK$^{-1} \times 300$ K). The energy of a mole of carbon–carbon bonds is 360 kJ (Table 2.1) so the energy of each bond can be calculated by dividing through by Avagadro's number: 6×10^{-19} J ($= 360{,}000/6.02 \times 10^{23}$). The energy of the bond is 144 times that of the thermal energy at this temperature so we would expect thermal motion to have little effect; the probability of a bond breaking due to thermal energy using the Boltzmann distribution:

Table 2.1 Pauling scale values for electronegativities (Haynes et al. 2013).

Pauling number	
C	2.6
H	2.2
O	3.5
N	3.1
Na	0.9
Ca	1.0
Fe(II)	1.8
Fe(III)	1.9
Al	1.5
Cl	3.0

- **Fixed Geometry.** Covalent bonds are short (~ 1–2 Å) and very strong while multiple bonds tend to be shorter and stronger still (Table 2.2). The angles between bonds are fixed and depend on which orbitals are involved in bonding. We can again take a simplified approach and imagine the shapes result
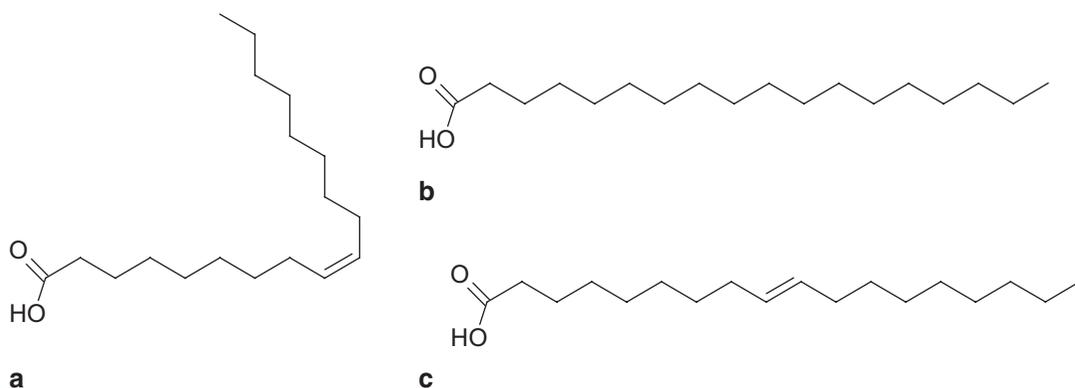
Fig. 2.5 a Oleic acid, **b** Stearic acid, and **c** Elaidic acid

$$\frac{n_i}{n_0} = \exp\left(\frac{-\Delta E}{kT}\right) \qquad (2.3)$$

where n_i is the number of molecules in the high energy state (e.g., nonbonded) and n_0 the number in the low energy state (e.g., bonded). In this case $n_i/n_0 = 1.86 \times 10^{-63}$, a vanishingly small number, and we can be certain that unless the temperature is enormous, thermal energy alone will never break covalent bonds. Even as we heat a food and the molecular motions become faster, they are never likely to reach an intensity that the covalent bonds will spontaneously break and so, for our purposes, we can treat them as "fixed" linkages between atoms. Of course, making and breaking covalent bonds is important for many reactions in foods (e.g., the rancid aroma in oxidized fat results from the cleavage of carbon–carbon double bonds and the formation of carbon–oxygen bonds), and when we argue that bonds are fixed, we mean they will not break down by heat alone—there must be some sort of chemical mechanism proposed to allow the reaction to proceed. For the most part though, we will not deal with covalent bond reactions in this work and instead study the ways that intact molecules to build larger structures within food.

Although we can regard covalent bonds are permanent, we should not see them as rigid; they flex and vibrate elastically about their mean angles and lengths to a greater extent as they are heated. Importantly, single bonds are free to rotate about their axis. Bond rotation can be responsible for dramatic changes in the shape of a molecule as illustrated in Fig. 2.3b which shows a 2D projection of the rotated forms of a substituted ethane compound (seen in Fig. 2.3a). The only significant restrictions to single bond rotation are interactions between substituent groups that may favor one configuration over another.

It is more difficult to rotate about a double bond as this would require breaking one of the bonds, rotating about the residual single bond, and then reforming the double bond in the opposite configuration. We can therefore treat the *cis*- (i.e., adjacent hydrogens on the same side of the molecule) and *trans*-isomers (i.e., adjacent hydrogens on the opposite side of the molecule) as different molecules with different properties. For example, most of the double bonds in natural vegetable oils are in the *cis*-configuration (e.g., oleic acid, Fig. 2.5a). To turn the liquid oils into solid fats for margarine, hydrogen is added across the double bonds to turn them into single bonds (i.e., hydrogenation, adding hydrogen to oleic acid converts it to stearic acid as shown in Fig. 2.5b). A by-product of this reaction is significant amounts of *trans*-fats (e.g., elaidic acid is the *trans* version of oleic acid, Fig. 2.5c). The original oleic acid has a kink in the chain due to the *cis* double bond while the saturated stearic acid and *trans* elaidic acid are straighter molecules. Although oleic and elaidic acids have the same chemical composition, the *cis* to *trans* isomerization raises the melting point from 4 °C to 46.5 °C.

Because of their strength and permanence, we can describe covalent bonds in terms of length, characteristic angles, and polarity and then treat the resulting molecules as more or less fixed building blocks from which we will assemble food structure. However, the bonds between molecules are usually much more tenuous and to understand them properly we will need a clear picture of how electrostatic forces acting at a distance give rise to a bond.

2.4   Intermolecular Forces

The closer you push the north poles of two magnets together, the more strongly they repel one another. Similarly, if you try to move the north pole of one magnet toward the south pole of a second, they will attract one another more and more strongly as they get closer. This is an everyday manifestation of the same electrostatic forces that are responsible for bonding. Rather than moving directly to a mathematical description of electrostatic forces, it is instructive to use an analogy to see how forces acting at a distance can give rise to bonds. Rather than pushing two magnets together, we will imagine pushing one ball across the sloping surface of a table toward a second, fixed ball. Various shapes of surface are shown in Fig. 2.6, the left hand figures show the height of the surface as a function of the separation between the two balls and the right-hand figures are the forces required to hold the moving ball at a given separation. If the table were flat (Fig. 2.6a), there would be no force needed to move the ball to any separation. If the surface sloped towards (Fig. 2.6b) or away from (Fig. 2.6c) the fixed ball, then the second ball would tend to roll away from it or towards it and would require a positive or negative force respectively to remain in a given position. The magnitude of the force required to hold the ball at a given position depends on the slope of the surface so in Fig. 2.6b and c the same force is needed to hold the moving ball at any separation from the fixed ball. Figure 2.6d shows a curved surface, the strength of the repulsive (positive) force increases as the separation

decreases. Figure 2.6e shows a complex surface, with an attractive force at long separations and a repulsive one at short separations. At an intermediate separation, the energy minimum, there is no net force acting on the moving ball and that position represents the equilibrium separation of the two balls. The moving ball will tend to roll into the energy minimum and stay there.

These trivial examples show how heavy balls will move according to a gravitational potential but we can reimagine the left hand figures as the electrostatic potential between two molecules as a function of separation. The right hand figures show the force acting on the moving molecule as it approaches the fixed molecule; if the force were negative at any point, it would tend to pull the molecules closer and if it was positive, the molecules would tend to repel one another. Figure 2.6a shows the potential for noninteracting particles (i.e., an ideal gas). Figure 2.6d shows a potential that gets steeper at shorter range. At long separations, there would be no forces between the molecules but, as separation decreases, the repulsive force gets stronger. This example corresponds to two similarly charged ions. Figure 2.6e represents a bonding potential. At long separations, there are no interactions between the molecules and they are free to move uninfluenced by one another. However, as they approach one another, the potential starts to curve downwards toward an energy minimum that tends to trap the molecules at a fixed separation from one another. The bond length is given by the separation at the energy minimum, that is, the separation when the slope of the potential, and thus the forces acting is zero (shown as s^* in Fig. 2.6e). The strength of the bond (shown as ΔE in Fig. 2.6e) is the energy needed to pull the molecule out from the energy minimum and drag it to a range at which it no longer interacts with the fixed molecule. The bond energy at any separation can be expressed as either the depth of the energy minimum or the area under the force distance curve as shown in Fig. 2.6e.

To understand the interactions between molecules we must calculate the shape of the electromagnetic potential. In the next few sections,

Fig. 2.6 Energy potentials (*left*) and corresponding force–distance plots (*right*) between a fixed particle at separation zero and a second moving particle. **a** A flat potential—no forces acting. A potential with a constant **b** positive or **c** negative slope results in a repulsive (positive) or attractive (negative) force respectively. **d** Curved potential means force (negative slope) also changes with distance. **e** A complex potential with negative and positive forces as a function of distance. The minimum energy/zero force is the equilibrium separation (s^*) of the particles and ΔE is the energy required to move them from this separation out to a range where they no longer interact
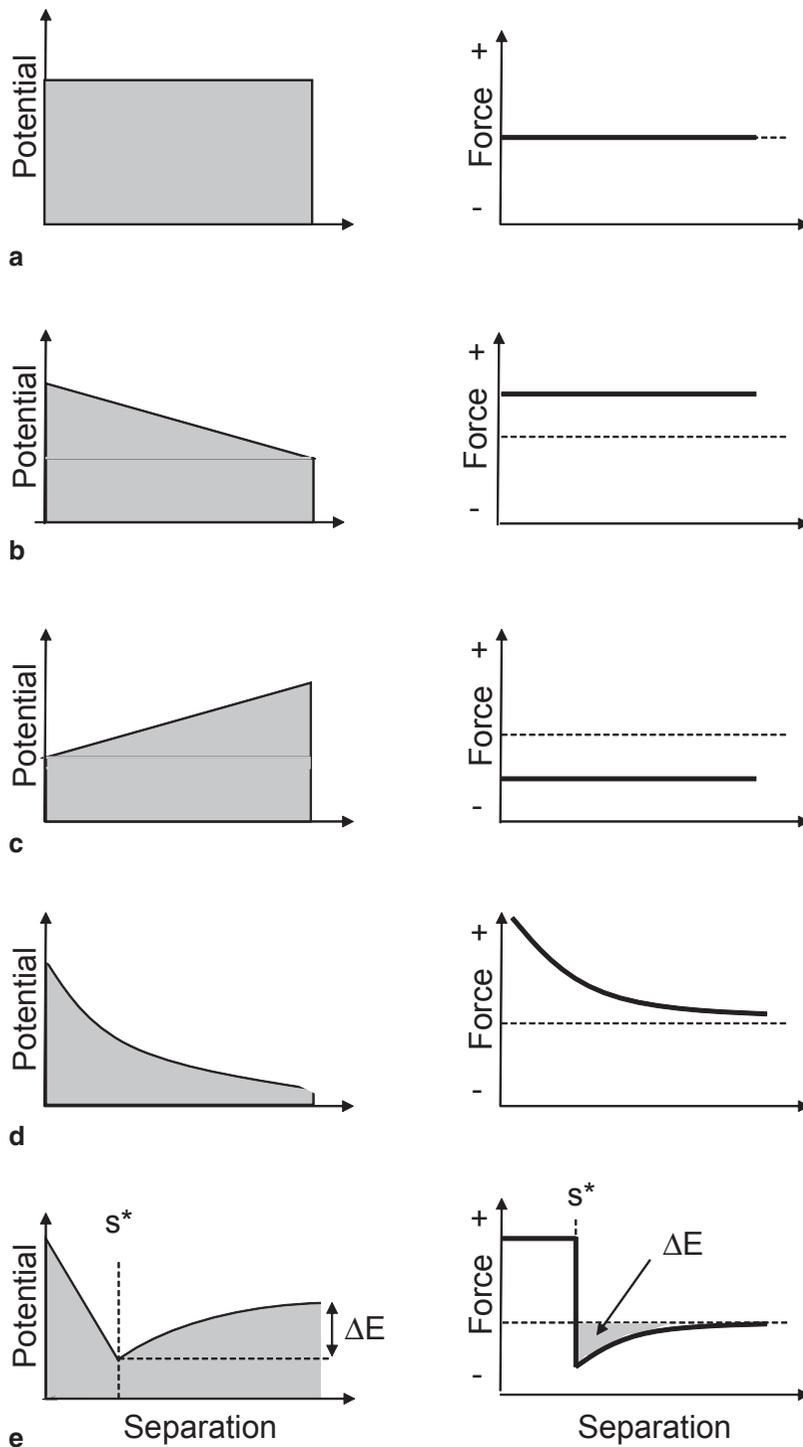
we will look at various types of intermolecular interaction that might contribute to the overall interactions (i.e., types of bond). We will then sum the individual interactions to get the full electromagnetic potential function and calculate the bond strength and length.

Fig. 2.7 Interaction potential between a sodium and chloride ion in water (e_r=78, *left axis*) and oil (e_r=2, *right axis*)
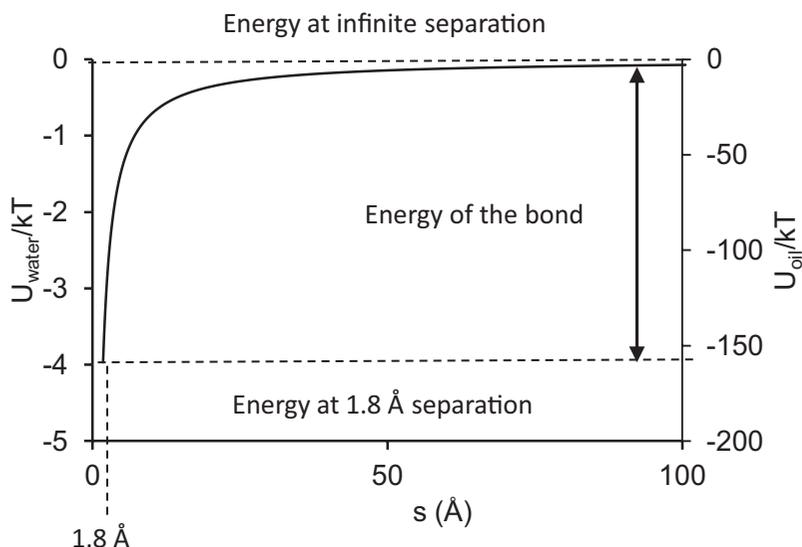
2.5   Ion–Ion Interactions

The interaction energy between two charges a distance s apart is given by Coulomb's law:

$$U_{ii}(s) = \frac{q_1 q_2}{4\pi \varepsilon_0 \varepsilon_r s} \qquad (2.4)$$

where q_1 and q_2 are the magnitudes of the change in coulombs (the charge on an electron is 1.602×10^{-19} C) and ε_0 and ε_r are the dielectric permittivity of a vacuum ($= 8.85 \times 10^{-12}$ c^2 N^{-1} m^{-2}) and the relative dielectric constant of the medium separating the charges, respectively. The importance of equations such as this is to concisely and precisely state what we know about the interaction. For example, the common observation "like charges repel one another" is contained within Coulomb's law: If the sign of q_1 and q_2 are similar then $U(s)$ is positive and there would be an energy cost to bring the charges together. Coulomb's law helps explain why sodium and chloride ions can sit alongside one another in a salt crystal but sodium and potassium ions cannot. Another common observation "salt dissolves in water but not in oil" can be quantitatively understood in terms of Coulomb's law as the relative dielectric permittivity of oil is much less than that of water (approximately 2 and 78, respectively). The interaction potential between two dissimilar charges as a function of separation distance in oil

and water are shown in Fig. 2.7. The potential is negative in both cases and the ions attract one another, but the magnitude of the potential is much greater at a given separation in oil than in water. For example, taking the radius of a sodium ion as 1 Å and a chloride ion as 1.8 Å, their minimum separation should be 1.8 Å; if we wanted to dissolve them in a solvent, we would have to move the point charges from this separation out to an infinite distance. If we move the ions apart in water, the energy cost would be about 15 kT, a large energy barrier but not insurmountable. If we tried to move the ions apart in an oil solvent, the energy cost would be a prohibitive 600 kT. (A word of caution: In all of these calculations, we are assuming that the solvent can be described as a continuum with a dielectric permittivity equal to its bulk measured value. This probably reasonable at wide separations where there are many solvent molecules between the charges and their many different conformations tend to cancel each other out. However, when the separation between the ions is small, the exact arrangement of the few atoms and local charges on the solvent molecules will make a huge difference to the effective permittivity and the results from Eq. 2.4 will become unreliable as the essential graininess of matter becomes important.)

The range of the interaction is given by the functional dependence of the potential on separation distance. In this case, the potential is pro-

Fig. 2.8 Sodium ion activity (i.e., fraction of sodium ions added effective as free Na⁺) as a function of sodium sulfate and sodium chloride concentration. *Inset* shows a schematic representation of the ion pairs that form at higher concentrations due to the long-range effectiveness of electrostatic interactions
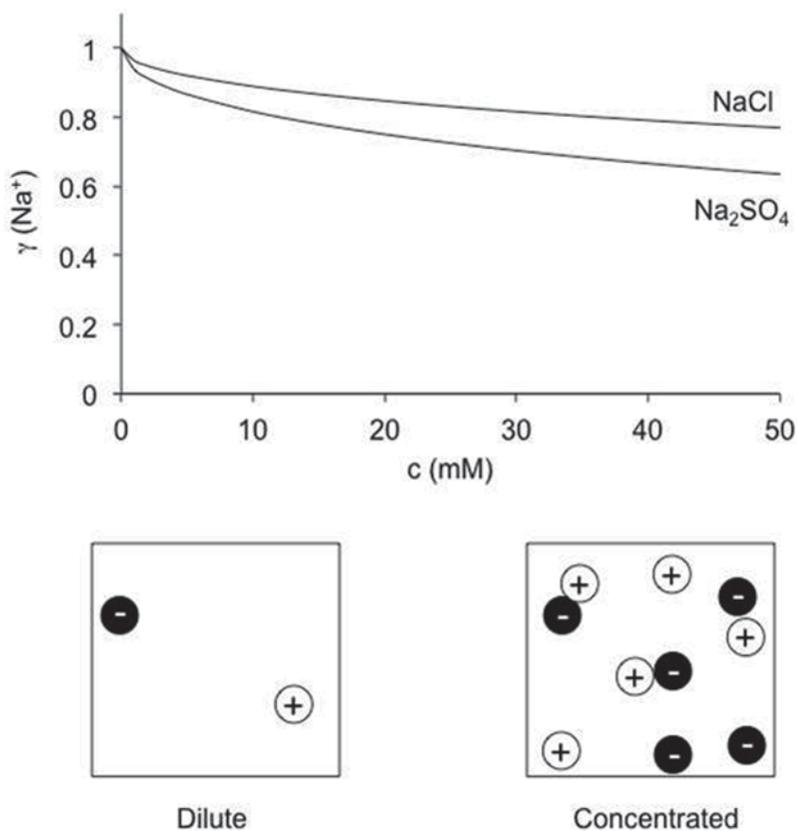
portional to *1/s* and the interaction will be felt a long way away from any ion. (We will meet other interactions later where the potential is proportional to s^{-6} or even s^{-12}.) Oppositely charged ions will tend to attract one another and form pairs. These pairs are not permanent associations like the ionic bonds in a crystal lattice but even their transient presence means the effective concentration of free ions in solution is less than the amount added. Walstra (2003) suggests an estimate of the activity coefficient, γ, of an ion as:

$$\ln (\gamma) = 0.8 z_i^2 \sqrt{I} \qquad (2.5)$$

where z is the charge of the ion and I is the ionic strength. Ionic strength is a measure of the total concentration of charge in the system and is given by:

$$I = 1/2 \sum c_i z_i^2 \qquad (2.6)$$

where c_i and z_i are the concentration and charge on ion i. For example, Fig. 2.8 shows the activity of sodium ions as a function of sodium chloride and sodium sulfate solutions. As concentration increases, more and more of the ions are bound in pairs and consequently contribute less to the effective free ion concentration. Similar losses of effective concentration (i.e., reduction in activity) are seemingly for nonionic molecules but not until much higher concentrations as their intermolecular interactions are shorter range.

2.6   Ion-Dipole Interactions

Equation 2.4 describes the interactions between ions, but because of bond polarization, many nonionic food molecules carry several partial charges distributed about their structure. In principle, the interactions due to partial charges can be calculated from every interaction pair, for example,
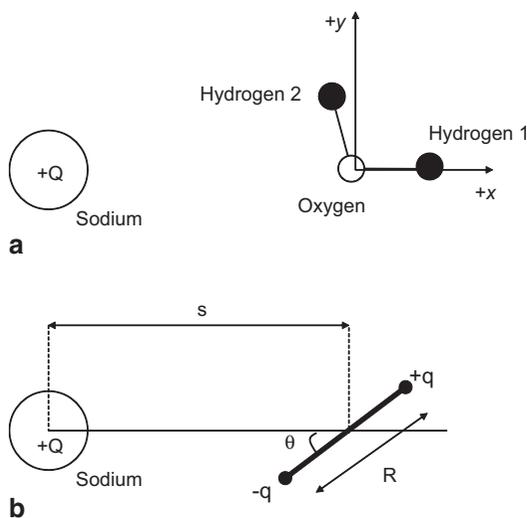
Fig. 2.9 Diagram showing the electrostatic interactions between a sodium ion and a water molecule. The water molecule is represented as **a** three point charges or **b** an equivalent dipole

Fig. 2.9a shows a water molecule in a fixed distance away from a sodium ion. There is a partial positive charge on each hydrogen atom and a partial negative charge on the oxygen atom to maintain overall neutrality. The net interaction between the ion and the molecule would have three separate terms: one oxygen–sodium attraction and two distinct hydrogen–sodium repulsions each depending on the distance of the water from the ion and its relative orientation. Larger molecules are more complex. It is more practical to combine the contributions of all the charges on the molecule and to treat the complex structure as a dipole.

A dipole is two equal and opposite charges separated by a distance (Fig. 2.9b). It is a vector quantity in that it has a magnitude and a direction. The magnitude of a dipole, the dipole moment, is given by $\mu = qR$ where q is the magnitude of the charges and R their separation. The unit for a dipole moment is coulomb-meters but the commonly used non-SI unit is the Debye (D) where 1 D$=3.335 \times 10^{-30}$ Cm. A dipole moment of 1 D, typical of many small polar molecules (Table 2.3) corresponds to a charge of $+1$ electron and -1 electron separated by a distance of 1 Å.

Table 2.3 Dipole moments of selected molecules (Haynes et al. 2013)

Compound	Dipole moment (D)
Propane	0.08
Propanol	1.58
Methanol	1.7
Propanoic acid	1.75
Octanol	1.76
Water	1.85
1–Chloropropane	2.02

A simple molecule made from two different atoms is clearly a dipole— the bond has a fixed length (R) and because it is polarized, one atom carries a partial positive charge and the other a partial negative charge (q). The dipole moments of larger molecules can be calculated by choosing a convenient reference point on the structure and summing the contributions of all of the partial charges on all of the atoms:

$$\mu_x = \sum_i q_i x_i \tag{2.7}$$

where q_i is the charge on atom i and x_i the distance from the reference point in the x-direction. Because molecules have 3D shapes, there is a corresponding vector for the y and z directions. The overall dipole moment is then given by Pythagoras' theorem: $\mu = \sqrt{\mu_x^2 + \mu_y^2 + \mu_z^2}$. For example, Fig. 2.4 shows the geometry of the water molecule with the partial charges on the oxygen ($-0.42\ e$) and hydrogen ($+0.21\ e$) atoms. Taking the oxygen atom at the origin of the axes and aligning one of the hydrogens with the $+$x axis, the overall dipole can be calculated in each direction:

$$\mu_x = \left[0.958\,\text{A x } 0.21e \right]$$
$$- \left[0.958\,\text{A.cos}(60).0.21e \right]$$
$$\mu_y = 0 + \left[0.958\,\text{A.sin}(60).0.21e \right]$$

The first term on the right-hand side of these expressions is the contribution of hydrogen #1 and the second to hydrogen #2. There is no contribution from the oxygen atom as it lies at the origin of the axes with no displacement in either the x- or y-directions. There is no z-component to the

water dipole because it is a planar molecule. The overall dipole moment of water is then calculated to be 1.5 D, reasonably close to the experimentally measured value of 1.85 D (by microwave spectroscopy). According to these calculations, the electrical interactions of water can be modeled as those of a rod of length 1 Å with a charge of 1.5 e on each end or a rod of length 1.5 Å and charges 1 e (Fig. 2.4c).

The electrical interactions of the dipole derived from the molecular structure are the same as the sum of the individual partial charges on the original structure (Fig. 2.9). Having determined the dipole moment, the interaction potential between an ion and a dipole is easily calculated from a variation of Coulomb's law:

$$U_{id}(s) = \frac{q_1 \mu_2 \cos\theta}{4\pi\varepsilon_0\varepsilon_r s^2} \qquad (2.8)$$

where θ is the angle of orientation of the dipole to the ion (see Fig. 2.9b). Equation 2.8 is very similar to Eq. 2.1 with two important differences. First, the strength of the interaction depends on the angle of orientation of the dipole. A polar molecule will orientate itself to an ion so the unlike partial charge is closest and the like charge is furthest away. Fig. 2.10 shows the energy cost of bringing a water molecule toward a sodium ion at a fixed orientation:

- If the positive hydrogens are facing the positive sodium, the net interaction is repulsive.
- If the negative oxygen is facing the positive sodium, the net interaction is attractive.
- If the water is aligned at 90° to the sodium, the repulsive interactions cancel one another out and there is no interaction.

Second, Fig. 2.10 also shows the tendency of a dipole to align itself with an ion. At large separations, the energy of the different orientations differ only by a few kT so while the attractive, lower energy orientations will be most common, the dipole will still be relatively free to rotate. On the other hand, water molecules very close to an ion have a strong preference for the low energy conformation and will tend not to rotate out of it.

Second, the range of an ion–dipole interaction is much less than an ion–ion interaction. At close ranges, the attraction between an orientated di-
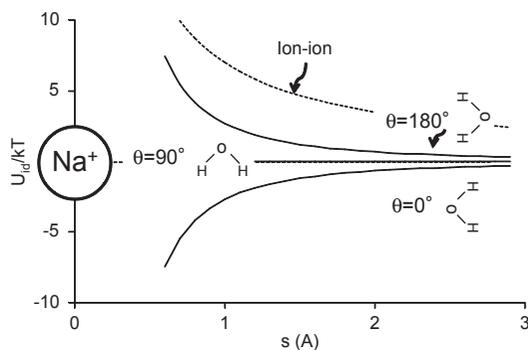

Fig. 2.10 The ion–dipole interaction potential between sodium and water at different fixed orientations. A sodium–sodium interaction is shown for comparison (*dashed line*)

pole and an ion is strong because the attraction of the unlike charge is greater than the repulsion of the like charge. However at long ranges, the like and unlike charges on the dipole are close together (compared to their distance from the ion) that they tend to cancel one another out.

2.7 Dipole–Dipole Interactions

The interactions between two dipolar molecules can be visualized at a macroscopic scale in bar magnets freely rotating on a string interacting with a second bar magnet that is fixed and not free to rotate. At large separations, they will not be affected by one another but as they are brought closer they will tend to feel a rotational force (to line up opposite poles) and an attractive or repulsive force depending on the current alignment (Fig. 2.11). Some examples:

1. The magnets/molecules are at right angles to one another so the attractive and repulsive forces balance one another out and there is no net force between them. However, the magnet/molecule will tend to rotate to bring the opposite poles closer together.
2. The magnets/molecules are antiparallel to one another so there is a net attractive force. There is no turning moment in this configuration.
3. The magnets/molecules are parallel to one another so there is a net repulsive force. There is no turning moment on the molecules in this exact configuration but any slight perturbation
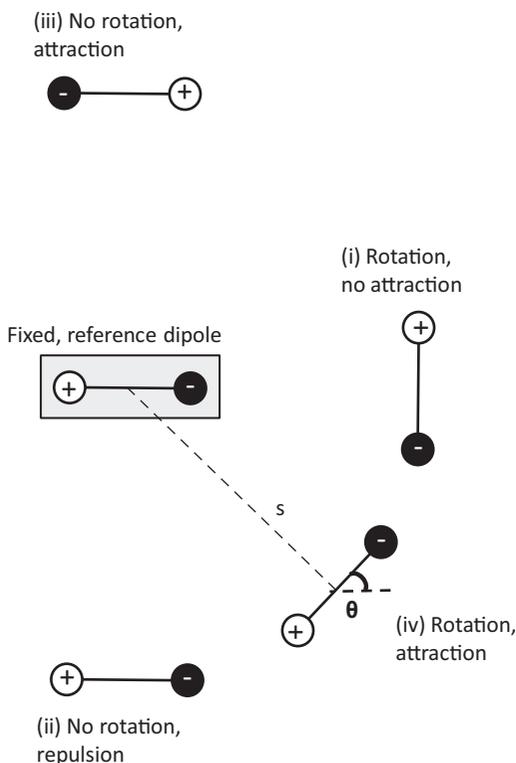
(iii) No rotation,
attraction

Fixed, reference dipole

(i) Rotation,
no attraction

s

θ

(iv) Rotation,
attraction

(ii) No rotation,
repulsion

Fig. 2.11 Illustration of the interactions between pairs of dipoles at different angles to one another

However, in the liquid state molecules are free to move so their orientations are not fixed. If the molecules were completely free to rotate, the attractive and repulsive orientations would cancel one another out and the net dipole–dipole interaction would be zero. However, the attractive orientations are preferred to the repulsive ones and polar molecules in solution will tend to line up next to one another and provide a weak attractive force, the Keesom force. The range of this interaction is the shortest we have seen to so far ($\sim s^{-6}$) because the orientation effect is only quite weak, that is, the molecules are rotating freely with only a slight preponderance of time in the attractive conformations. The inverse sixth power dependence on separation is characteristic of Van der Waals forces, and the Keesom forces are one contribution to that group of interactions. Importantly, because the attractive orientations are preferred, the net effect of the Keesom forces is always attractive. Polar molecules will always attract one another, but only over quite short ranges.

2.8 Van der Waals Interactions

The ion–ion, ion–dipole, and dipole–dipole interactions are typically weaker than covalent bonds but can provide structure to assemblies of molecules with charges. There are also electrostatic interactions between nonpolar molecules but these tend to be weaker still and very short range. Intermolecular attractions between nonpolar molecules are responsible for keeping small lipids, for example, hexane liquids rather than gases at room temperature.

Nonpolar molecules lack any separation of charge. Symmetrical compounds cannot be dipoles, for example in hydrogen gas (H_2) the electrons are pulled equally in both directions along the bond and there is no dipole formed. Even when different atoms are involved, if the molecule is symmetrical it cannot be a dipole. For example the electrons in the bonds in carbon tetrachloride (CCl_4) are drawn toward the electronegative chlorine atoms but because the pull in one direction is cancelled out by pulls in the

may allow them to flip into the antiparallel alignment.
4. At other angles, θ, there are both turning and attractive forces.

Assuming a fixed orientation (e.g., for two molecules fixed in a crystalline lattice), the magnitude of the attractive/repulsive force is:

$$U_{dd}(s) = \frac{\mu_1 \mu_2 (1 - 3\cos^2\theta)}{4\pi \varepsilon_0 \varepsilon_r s^3} \qquad (2.9)$$

Note the range of the interaction is shorter ($\sim s^{-3}$) than either the ion–ion or ion–dipole interactions because at long range the positive and negative parts of both dipoles tend to cancel one another out. The magnitude of the force is less than the corresponding ion–dipole interaction because while one end of the dipole feels a net attraction, the other feels a net repulsion.
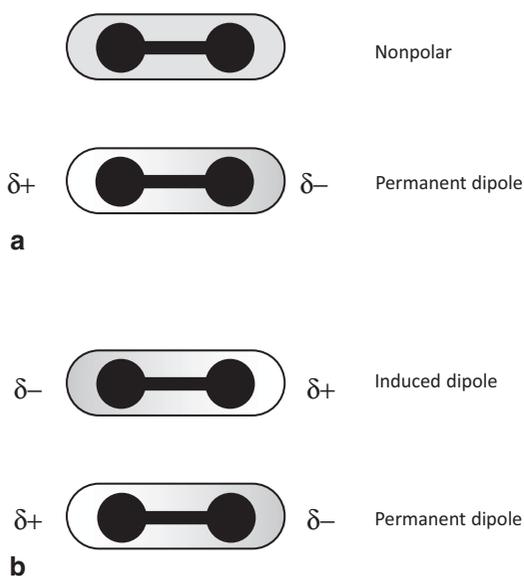
Fig. 2.12 Highly schematic diagram illustrating the basis for attraction between a permanent and induced dipole. Depth of shading is a schematic representation of average electron density. **a** A dipole approaches a nonpolar molecule and **b** induces the formation of an oppositely aligned dipole. Because the permanent and induced dipoles have an opposite orientation, they attract one another

Fig. 2.13 The electron distribution (illustrated schematically by depth of shading) on even a nonpolar molecule will instantaneously fluctuate between many states, several of which are polar and can induce a dipole in an adjacent molecule by a mechanism similar to that illustrated in Fig. 2.12

opposite direction, the net dipole moment is zero. The alkane and alkene building blocks of lipids are symmetrical and nonpolar (Fig. 2.3) and, although the oxygen atoms will generate a small dipole, most simple lipids can be treated as effectively nonpolar (Fig. 2.5).

Even though a molecule is not itself polar, it can be polarized. For example, if a nonpolar molecule approaches the negative end of a permanent dipole, the electrons will tend to move from their typical uniform distribution and accumulate on the atoms furthest away from the negative charge leaving the atoms closest to the charge slightly positive (Fig. 2.12). The newly induced dipole is aligned with the permanent dipole and so attracted to it. The magnitude of the attractive force depends on the permanent dipole moment and also on how easily the electrons on the nonpolar molecule can be moved from their equilibrium distribution, that is, the polarizability. The interaction between a permanent and induced dipole are known as induction forces. They are short ranged, depending on the inverse sixth power of
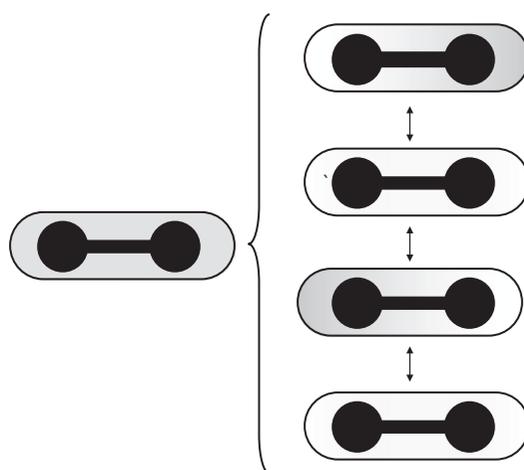
separation, and are a second contribution to the Van der Waals forces. A third contribution arises from the interactions between nonpolar molecules.

Even though a molecule is not polar, it can have an instantaneous dipole moment. The dipoles calculated above are based on the average distribution of electrons. In fact, the electron distribution on the bonds of an individual molecule fluctuate rapidly and at any given instant, there may be an accumulation of charge on one part of the molecule giving a nonpolar molecule and instantaneous or transient dipole moment (Fig. 2.13). The presence of a transient dipole on one molecule can induce a dipole in a second nearby molecule in a similar manner to the permanent-induced dipole interaction illustrated in Fig. 2.12. The instantaneous-induced dipole interaction results in attractive forces (London or dispersion forces) that are a third contribution to the Van der Waals interaction (i.e., $\sim s^{-6}$). The magnitude of the London forces depends on the polarizability of both groups as they require both the spontaneous movement of electrons in the first molecule and the induced movement in the second. Alkanes and hydrophobic molecules are highly polarizable. London forces require no per-

manent charges in the system so are particularly important in describing the properties of nonpolar fluids such as oils.

Van der Waals interactions are the sum of the Keesom (permanent dipole-permanent dipole), orientation (permanent dipole-induced dipole), and London (transient dipole-induced dipole) forces all decreasing with the sixth power of separation. The overall magnitude of the Van der Waals potential is given by:

$$U_{VDW}(s) = \frac{-C}{s^6} \qquad (2.10)$$

where C is a positive constant with contributions from the dispersion, induction, and orientational components of the interaction and is related to the properties of the molecules involved. As C is always positive, the Van der Waals potential is always negative so the forces acting are always attractive. The range of the Van der Waals forces is very short, decreasing with the sixth power of separation distance, but although they are weak, they act between any pair of molecules. In highly polar molecules, their contribution is small compared to the electrostatic interactions, but in nonpolar liquids they are effectively the only attractive forces acting.

2.9 Steric Interactions

Electrostatic interactions draw together molecules with unlike charges, and their weaker cousin the Van der Waals interactions brings together molecules without permanent charge. However at very small separations, as the molecules themselves come into contact, their electron orbitals will start to overlap. This is forbidden both in the world of quantum mechanics (i.e., the Pauli exclusion principle) and in the everyday world (i.e., you cannot force two objects into the same space) and leads to a strong steric repulsion at between the molecules at very short ranges. For spherical molecules of radius σ, a distance s apart the steric interaction is given as:

$$U_{steric}(s) = C\left(\frac{\sigma}{s}\right)^n \qquad (2.11)$$
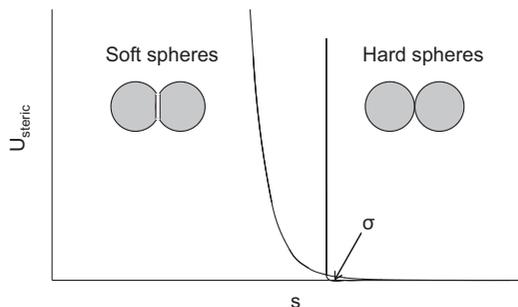

Fig. 2.14 Steric repulsion potential between molecules modeled as hard ($n=\infty$) and soft spheres ($n=12$) of radius σ

where n is a large number (often taken as 12) and C is a positive constant. When the separation between the molecules is greater than twice their radii (i.e., not touching), $\sigma/s<1$ so the steric force is negligible. However, as the separation between the molecular centers decreases below one diameter (i.e., molecules pushing into one another, $\sigma/s>1$), the magnitude of the repulsive force increases very rapidly with distance (Fig. 2.14). Sometimes the n parameter is adjusted the species as more or less compressible but these adjustments are relatively minor compared to the important consequence of steric forces—a minimum distance of approach between two molecules.

2.10 Bonding in Water—Some Special Cases

Water is a unique material and we owe the existence of life on earth to its plentiful supply and unusual properties. Indeed, in many of the recent news stories where astronomers claim to find moons and planets where the conditions are suitable for life, they usually mean that the temperature and pressure are such that there may be liquid water present. Some of the unusual properties of water that make it so important in living systems include:

- A high specific heat ($C_p=4.18$ JK^{-1} g^{-1}) to buffer cells against changes in temperature
- A solid form less dense than the liquid. Most solids are denser than their melts but ice will

float on water. If water behaved as a typi-
cal fluid the bottoms of the ocean and lakes
would quickly fill up with sunken winter ice
that would not melt in the spring.

- Water is a good solvent for ions and polar
molecules but a poor solvent for nonpolar
materials. Many biologically important mol-
ecules can therefore react in a water solvent
but nonpolar molecules are excluded to form
important structures. For example, phospho-
lipids will spontaneously aggregate to form
membrane bilayers and proteins will coil up
so their more hydrophilic amino acid residues
are at the surface (in contact with water) and
the more hydrophobic in the dry core.

Most of the unusual properties of water arise from
its highly hydrogen-bonded structure. In general
a hydrogen bond forms between two functional
groups AH and B where A and B are highly elec-
tronegative and B has a lone pair of electrons.
The hydrogen is left with a partial positive charge
and because it is so small it can get very close
to the electronegative B. Because the molecules
approach one another so closely, there is a cer-
tain degree of overlap of the atomic orbitals of
A, H, and B to form a partial covalent bond. Hy-
drogen bonds are much stronger than most other
non-covalent interactions (~ 10 kT) and they are
unusual because their partially covalent nature
gives them a preferred orientation (a straight line
linking A-H-B) and length (~ 1 Å). Each water
molecule has two hydrogens and two lone pairs
of electrons on the oxygen: so can form up to four
separate hydrogen bonds with other water mol-
ecules. If water were fully hydrogen bonded, it
would arrange into a hexagonal 3D lattice corre-
sponding to the crystalline structure of ice. There
is good evidence that a majority of these hydro-
gen bonds survive melting into the liquid state
and that water is in fact a highly cross-linked
material. Although structured, water is a highly
dynamic material with the intermolecular bonds
forming and breaking very quickly to allow flow.

The highly structured nature of water is im-
portant because it has lower entropy than would
be expected for a liquid. (See Chaps. 1 and 3
where entropy was introduced as the absence of
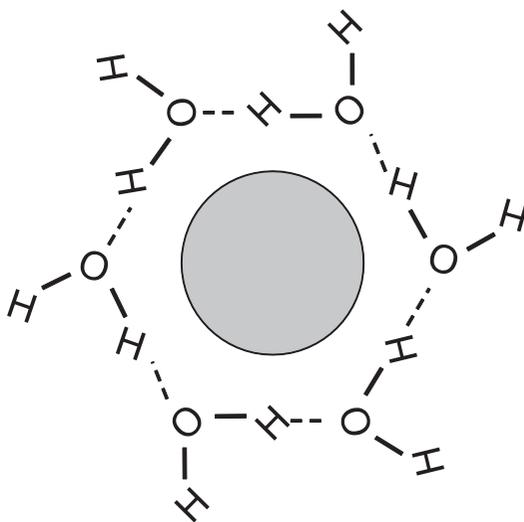order. Because the water molecules are to a de-

Fig. 2.15 Highly schematic diagram showing hydrogen-
bonded water as a clathrate cage around a nonpolar mol-
ecule. The water adjacent to the surface of the solute has
different properties than the water in the bulk (i.e., more
hydrogen bonds, lower entropy)

gree orientated with respect to one another, they
have low entropy). When a solute dissolves in
water, the water molecules must rearrange their
bonding patterns, both with the solute molecule
and with each other, to accommodate the inclu-
sion. Some solutes favor structuring of the water
(i.e., more hydrogen bonds, less entropy) while
others oppose it. Particularly important is when
a nonpolar molecule dissolves in water; there are
few significant intermolecular interactions be-
tween the solute and solvent but the water forms
a highly ordered *clathrate cage* around the solute
to accommodate it (Fig. 2.15). The formation of
the additional hydrogen bonds involves a loss of
energy (i.e., exothermic) but the entropy cost in
orientating the water molecules opposes the dis-
solution. As a consequence polar molecules dis-
solved in water are attracted to one another by a
hydrophobic force; the force is not enthalpic like
the electrostatic interactions seen in this chapter
but entropic, because bringing the two molecules
together will reduce the amount of ordered water.
This distinction is important, because entropic
interactions become more important at higher
temperatures (Eq. 1.10, $G=H-TS$). While an en-
thalpy-based "bond" can be overcome with heat
(e.g., more salt dissolves in water at high tem-
peratures), an entropy "bond" cannot (e.g., oil

actually becomes less soluble in water at higher temperatures).

2.11 Effects of pH on Molecular Interactions

Intermolecular interactions involving ions are stronger than those between polar groups. However, many molecules are weak acids or weak

Example: Solubility of Amino Acids

Amino acids are sometimes added to foods as nutrient supplements or as reagents to generate Maillard flavors and colors during cooking. Needham and co-workers (1971) measured the solubility of different amino acids by adding excess water, allowing the system to reach equilibrium, then withdrawing a small volume of the solution phase and measuring the dissolved amino acid concentration (Fig. 2.16).

The amine and carboxylic acid groups of an amino acid are polar, and interact well with water. However, increasing the number of carbons in the side chain increases the number of nonpolar groups and the strength of the hydrophobic interactions reducing the solubility.
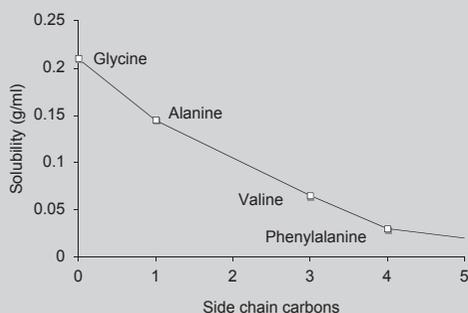

Fig. 2.16 Solubility of amino acids as a function of number of carbons in the side chain. A benzene ring was assumed to be equivalent to three (CH_2) groups so phenylalanine is treated as $C=4$. Adapted from Needham et al. (1971)

Table 2.4 Typical pK values of some functional groups. In real molecules, the actual pK value will vary depending on the structure of the rest of the molecule as well as the solution conditions

Protonated form	Deprotonated form	pK
$-CH_2COOH$	$-CH_2COO^-$	5
Protonated pyridine	Pyridine	5
$-CH_2SH$	$-CH_2S^-$	8
$-CH_2NH_4^+$	$-CH_2NH_3$	10

bases, capable of binding or releasing a proton depending on the pH and so shifting between ionic or nonionic forms:

$$AH \rightleftarrows A^- + H^+ \quad \textit{weak acid}$$
$$BH^+ \rightleftarrows B + H^+ \quad \textit{weak base}$$
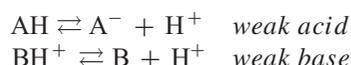
At high proton concentration, (i.e., low pH) the proton-bound form is seen (i.e., AH, BH^+) and vice versa. The affinity of a functional group for a proton is given by a binding coefficient:

$$K_A = \frac{[A^-][H^+]}{[AH]} \ \text{ or } \ K_B = \frac{[B][H^+]}{[BH^+]} \quad (2.12)$$

Typical values of pK (i.e., $-\log_{10}K$) are given in Table 2.4. Taking the definition of $pH = \log_{10}[H^+]$ and substituting for $[H^+]$ in Eq. 2.12 gives the Henderson–Hasselbalch equation:

$$pH = pK_A + log\frac{[A^-]}{[AH]} \ \text{ or }$$
$$pH = pK_B + log\frac{[BH]}{[BH^+]} \quad (2.13)$$

Knowing the total concentration of the functional group $c = [A^-] + [AH]$ or $c = [B] + [BH^+]$, we can calculate the proportion present as either the protonated or deprotonated form as a function of the properties of the solution (i.e., pH) and the properties of the functional group (i.e., pK) (Fig. 2.17). Some important features of Fig. 2.17:

- At $pH = pK$ there are equal amounts of protonated and deprotonated forms present,
- At low pH (i.e., high $[H^+]$), the protonated, more positively charged (i.e., BH^+) or neutral (i.e., AH), form is seen,
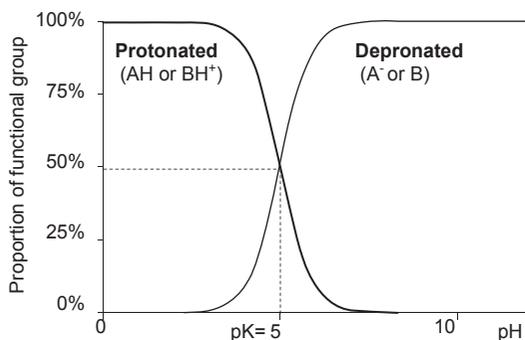
Fig. 2.17 Proportion of a weak acid/base present in the protonated (*bold line*, AH, or BH⁺) or deprotonated (*fine line*, A⁻, or BH) form. For both groups pK = 5

- The sigmoidal function only changes over about two pH units either side of the pK. Outside this range, only one form is seen.

Example: Solubility of Amino Acids

In the same study discussed above, Needham and co-workers (1971) measured the effects of pH on the solubility of amino acids. Glycine solubility is low between 4 and 8 but increases at higher and lower pH values (Fig. 2.18). At very high pH, there are very few protons in solutions and all the functional groups capable of donating a proton do so. The carboxylic acid carries a negative charge and the amine group carries no charge, so the molecule has a net charge of +1. Ions interact well with water and repel one another so have good aqueous solubility. As pH decreases, the amine group starts to gain a proton two pH units above its pK (=9.6), is 50% protonated at pH=pK, and is fully protonated two pH units below its pK. Thus by pH 8, the molecule has no net charge (i.e., zwitterionic) and is less water soluble. As pH is further reduced, the acid group starts to gain a proton two pH units above its pK (=2.3) is 50% protonated at its pK, and is fully protonated two pH units below its pK. Thus at very low pH, the molecule begins to
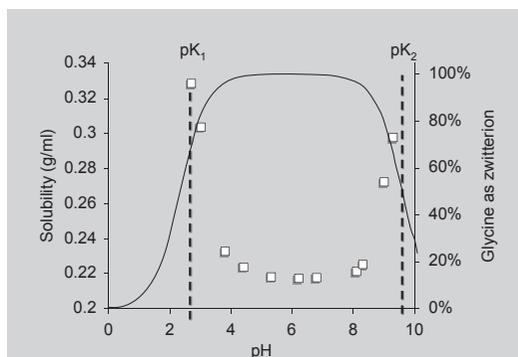

Fig. 2.18 Solubility of glycine as a function of pH. Adapted from Needham et al. (1971). *Line* shows calculated proportion of the glycine molecules present as the low solubility zwitterionic form

develop a net positive charge, interacts better with water, and its solubility increases once more.

2.12 Combined Interaction Potentials

The overall electrostatic energy potential acting between two molecules is the sum of all the interactions (i.e., electrostatic plus Van der Waals plus steric, etc.). While the exact potential between two real molecules is hard to calculate, it is instructive to examine a simple model as an example. For the widely used Lennard-Jones potential, we imagine two molecules interacting via Van der Waals attraction and steric repulsion:

$$U_{LJ}(s) = -As^{-6} + Bs^{-12} \qquad (2.14)$$

where A and B are appropriate constants for the attractive Van der Waals and repulsive steric interactions, respectively. By selecting values for A and B, we can calculate a net interaction potential between the molecules (Fig. 2.19). At long separations, both forces have decayed to zero and there is no interaction between the molecules. So for example, the properties of a gas are relatively independent of the interaction between the molecules as the average inter-
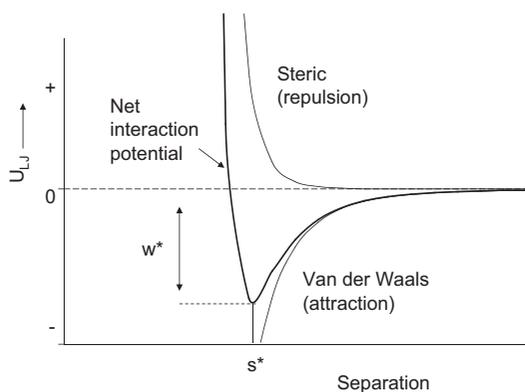
Fig. 2.19 Lennard-Jones potential showing the combination of an attractive Van der Waals term and a repulsive steric term to give a bond of characteristic length s* and energy w*

molecular separation is so large (i.e., all gases behave as ideal gases at low pressure). At intermediate separations, the steric repulsion has decayed to zero while the electrostatic attraction remains important and the molecules are attracted to one another. As the separation approaches the molecular size, the steric repulsion becomes increasingly important. At very small separations, steric repulsion dominates and the molecules will repel one another. Note the similarities between the Lennard-Jones potential and our sketched bonding potential in Fig. 2.6e. At a critical value s*, the steric repulsion exactly matches the Van der Waals attraction. At this point, there is no net force acting on the molecules and they will tend to remain at this equilibrium separation. The length s* represents an effective bond length for non-covalent interaction and will determine how tightly liquids pack and hence their density. The energy w* represents the bond energy.

2.13 Relating Bond Energies to Bulk Properties

Bond energies can be cautiously related to real molecular processes by carefully keeping count of the number and strength of bonds formed and broken. So, for example, to vaporize one molecule of a liquid, we would have to break every bond
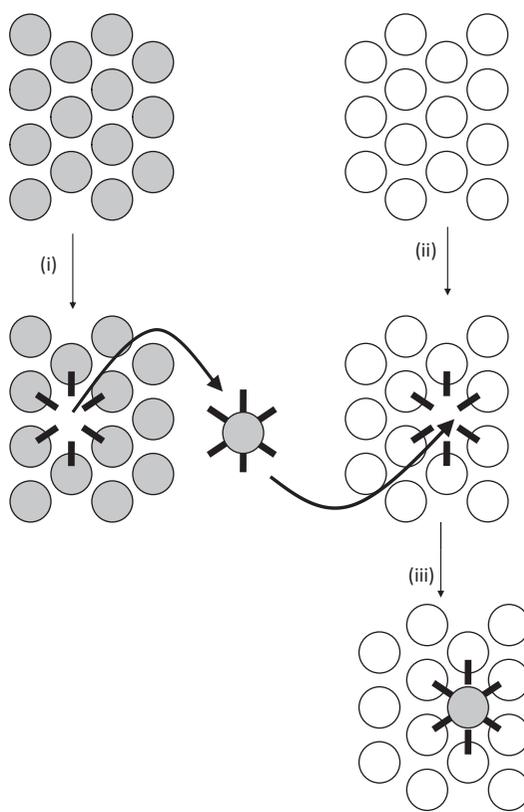

Fig. 2.20 Schematic diagram showing how molecular processes can be related to bond energies. For example (i) vaporizing a gas involves breaking a certain number of bonds between like molecules. Dissolving a solute in a solvent invoices (i) separating a solute molecule from its own phase (breaking bonds between like molecules), (ii) making space in the solvent for the solute by breaking solvent bonds, and (iii) adding the solute to the gap in the solvent and making solvent solute bonds

holding it to other liquid molecules (Fig. 2.20). Because bonds are short range, we can assume that the only significant interactions are between nearest neighbors and we can only count bonds between adjacent molecules. So if the coordination number is z (i.e., each molecule has z nearest neighbors), then the total vaporization energy for one mole of gas would be $\frac{1}{2}.z.w.N_a$ (the $\frac{1}{2}$ term is to avoid double counting molecules—breaking one bond frees two molecules). In other processes, keeping track of the number of bonds can be more difficult and it is necessary to generate an imaginary mechanism that isolates the changes more clearly. For example, to dissolve one liq-

uid in another, we first have to separate the solute molecules from contact with one another, next break bonds to make a gap in the solvent to accommodate the addition, and finally add the solute and make solvent–solute bonds. This is an unrealistic mechanism, but because the initial and final states are the same then the energy balance will work out correctly. We will use this approach to calculate phase diagrams in Chap. 4.

These calculations are based on considerable approximations both in the form of the interaction potentials and in deciding the number of bonds involved. Even if they were completely reliable, we would be hard pressed to use them to describe the complex interactions in highly complex mixtures of food molecules. An alternative is to use the interactions as part of a computer simulation. The approach here is to use a computer to generate a virtual box containing a collection of particles (i.e., molecules, ions, polymers, colloidal particles) (Fig. 2.21). If we can simulate the interactions of the components in this box, and if this box is representative of the overall system, then we should be able to predict measurable changes in bulk properties and connect the microscopic and macroscopic worlds.

Due to computational limitations, the boxes are necessarily small ($\sim nm^3$) but the effects of the small size are mitigated by using periodic walls—a particle leaving the box on one side will be replaced by another identical particle entering the box through the opposite wall. Similarly, interaction forces reaching the edge of the box will be felt by particles close to the opposite wall[1]. In this approach, it is important that the size of the box be no smaller than twice the maximum range of the interaction forces to prevent one molecule being attracted to itself. The particles interact according to the intermolecular forces selected and will provide a contribution to the overall internal energy of the system depending on their arrangement. The distribution of the particles is allowed to change until the free energy reaches a minimum at which point
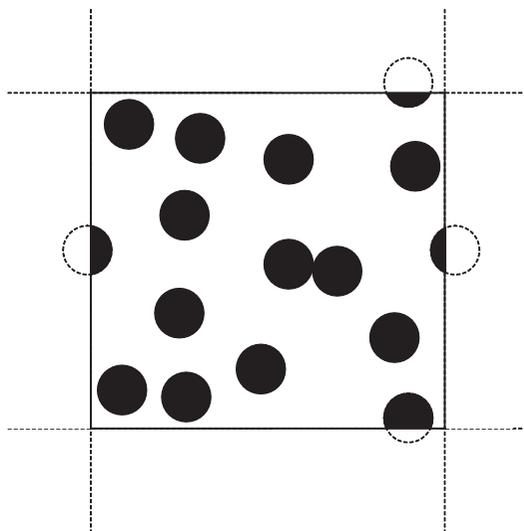

Fig. 2.21 Representation of particles in a box used in a simulation of molecular properties. Note that the walls are periodic so a particle leaving the box will be replaced by an identical particle entering through the opposite wall

the system is said to have reached equilibrium. There are two important approaches to moving particles in a computer simulation: Monte Carlo and molecular dynamics.

In a Monte Carlo simulation, new configurations are selected at random and the free energy of the rearranged system recalculated. If the change in free energy is negative, the new configuration is accepted but even if it results in an increase in energy, it is given a probability of being accepted comparing using the Boltzmann distribution. This last criterion is important as it stops the system getting stuck in a false energy minima. Importantly in Monte Carlo simulation, the changes do not correspond to real dynamics of the system. The final steady state should correspond to an equilibrium state but the pathway taken is determining by the random selection of particles. Monte Carlo simulations can therefore only be used to model equilibrium properties of a system.

The second group of simulation methods is molecular dynamics. In this approach, the particles are each randomly assigned a velocity depending on the temperature and are allowed to move for a short time step on a trajectory governed by the forces acting on them due to the intermolecular interactions. It is important

---

[1] Readers of my generation may find the computer game "Asteroids" a helpful way to visualize periodic walls.

the time step is small so the forces do not alter significantly during movement; in a molecular systems femtosecond time steps are often used. After movement, transfers of momentum due to any collisions are calculated and then the intermolecular forces are recalculated for each particle before allowing them to move forward another time step. This process is repeated until a satisfactory amount of experimental time is generated. Molecular dynamics reveal a reasonable dynamic pathway by which the molecules approach equilibrium and so can reveal dynamic properties as well as equilibrium (e.g., diffusion as well as equilibrium partitioning).

The properties of the virtual molecules can be averaged to calculate virtual bulk properties for the system (e.g., density, concentration) that can then in turn be compared to real experimental data. Assuming both the real and virtual experiments were conducted correctly and represent the same phenomena, any differences in the results must correspond to errors in the assumptions used to generate the model. Adjusting the properties of the model to better describe reality sheds light on the interactions that govern the phenomena of interest.

2.14   Summary

For many purposes, molecules can be treated as moving particles repelling one another at short separations and attracting one another at longer separations. The chemical structure affects the magnitude of the intermolecular interactions, with polar, and particularly charged, molecules interacting more strongly over longer ranges than nonpolar molecules. Water is an important special case, not just because of its prevalence in foods but because its highly hydrogen-bonded structure in the liquid state results in hydrophobic interactions between nonpolar molecules.

This simple mechanical model can be used in simulations to relate the microscopic proper-

ties of molecules to the bulk properties of materials. In later chapters we will use molecular interactions in other ways to understand the enthalpic changes that are important in controlling different types of food structure. Looking back to Richard Feynman's statement quoted at the start of this chapter, we will use these properties of molecules to explain the properties of food.

2.15   Bibliography

The basic properties of molecules are covered in sufficient depth in most general chemistry texts and in Chap. 30 of "Molecular Driving Forces" (Dill et al. 2003). The properties of molecules in solids, liquids, and gasses is described in Chap. 21 of "Atkins' Physical Chemistry" (Atkins and De Paula 2006). Students may find value in causing phase transitions and diffusion in one of the "toy" molecular dynamics simulations available online, for example "Democritus" (Côté et al. 2014).

Israelachvilli's (1991) "Intermolecular and Surface Forces" remains the most widely used book on the forces between molecules and their consequences. However, Dill and Bromberg (Dill et al. 2003), McClements (McClements 2004), and Walstra (Walstra 2003) make similar material more accessible.

The properties of water are described in wonderful depth by Talbot (2014) on his website "Water Structure and Science" and Chap. 30 of "Molecular Driving Forces" (Dill et al. 2003). The importance of water structure in foods is described in Chap. 2 of "Fennema's Food Chemistry" (Reid and Fennema 2008).

Chap. 1 of "Computer Simulation of Liquids" (Allen & Tildesley 1987) describes the basic principles of simulation and most importantly why simulation provides a powerful intellectual companion to theory and experiment (see especially Fig. 1.2 in that work).

Kinetics

3.1 Introduction

The equilibrium properties of a food depend on the ingredients selected and the conditions, so for example bread left in the oven will reach its equilibrium as a blackened crisp while bread left on the shelf will stale. However, by controlling the time and temperature during cooking and storage, the bread never reaches either of these undesirable states but is maintained in the correct non-equilibrium state. In fact most foods are in non-equilibrium states and we must understand both the thermodynamics, to understand if change is possible, and the kinetics, to understand how far it can proceed.

We will use kinetics in two main ways: first to answer the practical questions about processing and storage—if I can pasteurize milk in 30 min at 63 °C, how long must I process at 73 °C? What shelf life can I claim for my cookies? I know the vitamins advertised as healthy supplements in my granola bar will oxidize during storage—how much do I have to add to make sure that the claims on the label will still be valid if the product is stored for a year before it is sold? The second group of questions concerns the mechanism of reactions. The way the concentration of an ingredient changes with time and the way the rate is affected by the presence of other ingredients can be used to test proposed mechanisms for the reaction. In this chapter, we will start by contrasting kinetic theory with thermodynamics. We will then look at how rates of change can be modeled in terms of the rate equations and briefly

consider the relationships between kinetics and mechanism. We will show how the temperature dependence of rate can be used to infer some of the properties of the transition state and finally consider catalysis.

3.2 Kinetics and Thermodynamics

Thermodynamics is helpful in predicting the properties of a system at equilibrium. For a general reversible reaction: $A \rightleftharpoons B$, if the energy difference between A and B is ΔE, then the equilibrium state is given by the Boltzmann distribution:

$$\frac{[A]}{[B]} = \exp\left(\frac{-\Delta E}{k_B T}\right) \quad (3.1)$$

where the square brackets indicate concentration in whatever units are appropriate. In this chapter, we will use k_B for Boltzmann's constant as k will be widely used as a rate constant.

We can use a similar approach to understand the rate of reaction by postulating the existence of a high-energy intermediate state in the pathway between A and B (Fig. 3.1). Now only those reagent molecules with energies greater than $\Delta E_f^\ddagger$ would be able to form the high-energy intermediate state which could in turn break down to form the product (the $\ddagger$ symbol refers to the properties of the intermediate state). Similarly, only those product molecules with energies greater than $\Delta E_r^\ddagger$ would be able to perform the reverse reaction. The fraction of the molecules with sufficient energy to perform either the forward or reverse
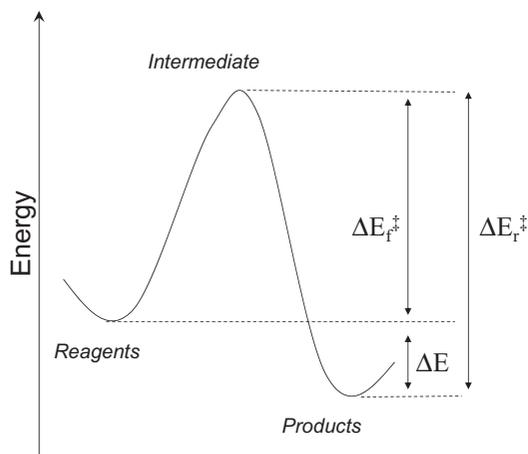
Fig. 3.1 Reaction surface illustrating the energy differences important in the thermodynamics and kinetics of a reaction

reaction in each case is given by the Boltzmann distribution. The greater the height of the energy barrier compared to the thermal energy of the molecules (i.e., $k_B T$), the fewer molecules that can form the intermediate and react and the slower will be the reaction. If the energy barrier is too high then the reaction cannot proceed regardless of the energy difference between reagents and products.

A physical analogy for the chemical system would be to imagine the molecules as marbles on a curved surface with a shape similar to the energy surface shown in Fig. 3.1. In this thought experiment, we are using gravity in place of chemical bonds as a source of potential energy and we will simulate the randomizing effects of heat by gently shaking the surface. We will start with all the marbles in the "reagent" trough and as we shake the marbles will move with a range of speeds. Occasionally, the random knocks will give a marble enough energy to hop over the barrier and roll down into the "products" trough and start bouncing around there. The number crossing the barrier in unit time depends both on the total number of "reagent" marbles and on the height of the barrier relative to the intensity of the shaking. If the height of the barrier is greater for the return journey $(\Delta E_r^{\ddagger} > \Delta E_f^{\ddagger})$, a smaller proportion of "product" marbles will be able to clear that bar-

rier and return to the "reagent" trough. At steady state, molecules are making the forward journey at the same rate as the return journey but because the former is easier than the latter, there are more "product" than "reagent" molecules.

The intermediate state itself is hard to study because it represents an energy maximum rather than a minimum. The intermediate does not accumulate and it is difficult to measure the exact molecular configuration it represents. Sometimes we can make a good guess, while in other cases, the nature of the intermediate must remain as speculation but its implied presence is a helpful tool for us to understand the observed kinetics of the reaction. However we imagine it, it is important to remember that the intermediate state is important only for the kinetics of the reaction. The thermodynamic equilibrium depends only on the energy difference between stable reagents and products.

3.3 Rate Equations

The rate of the general reaction, A→B, is proportional to the concentration of reagents raised to a power, that is:

$$\frac{-d[A]}{dt} = \frac{d[B]}{dt} = k[A]^{n_a} \qquad (3.2)$$

where t is time and n_a is the order of the reaction (often zero or an integer) and k a proportionality constant (the rate constant of the reaction). Note that because one molecule of A reacted to form one molecule of B, the rate of formation of the latter is equal to the (negative) rate of formation of the former. If the stoichiometry of the reaction were different, for example $aA \rightarrow bB$ then: $-\frac{1}{a}\frac{d[A]}{dt} = \frac{1}{b}\frac{d[B]}{dt}$.

In practice, the rate of a process is less useful than knowing the concentration as a function of time and this requires the integration of Eq. 3.2. Some standard results of the integration are given for $n_a = 0$, 1, and 2 are given in Table 3.1. In zeroth order reactions, the concentration of A

Table 3.1 Rate equations for various orders of reaction. Equations are for a general reaction A→B with the initial concentration of A being A_0. The unit of the rate constant is (concentration)$^{order-1} \cdot$ time^{-1}

	Zeroth order	First order	Second order
Rate equation, $-\dfrac{d[A]}{dt} =$	k	$k[A]^1$	$k[A]^2$
Integrated rate equation, $[A]=$	$[A_0]-kt$	$[A_0]\exp(-kt)$	$\dfrac{[A_0]}{1+kt[A_0]}$
Linearized plot	$[A]$ vs. t, slope is $-k$, intercept is A_0	$\ln([A]/[A_0])$ vs. t slope is $-k$, intercept is zero	$1/[A]$ vs. t slope is k, intercept is $1/[A_0]$

changes linearly with time but for other orders the relationship is nonlinear. However by plotting the logarithm of concentration or the reciprocal of concentration, it is possible to generate linearized plots with time for first- and second-order reactions, respectively.

Example: Thermal Processing of Peas

Ryan-Stoneham and Tong (2000) measured the concentration of chlorophyll in pea puree heated to different temperatures as a function of time. Chlorophyll is responsible for the green color of peas and in this work could be used to optimize the cooking to minimize the color loss. Sample data are shown for measurements at 100 °C at pH 5.5 (Fig. 3.2, note that the chlorophyll concentration at any time is shown as a fraction of the initial concentration). The concentration decreases with time as the chlorophyll molecules break down via an unknown mechanism and the rate of the reaction (slope of the concentration–time plot) also decreases with time (Eq. 3.1). To calculate the value of the rate constant we need to know the order of the reaction so we can choose the right equation from Table 3.1. One approach is to try all of them and see which gives the expected straight-line fit to the linearized plot (Fig. 3.3). In this case, the first order plot gives the best fit so we can conclude that the reaction proceeds via
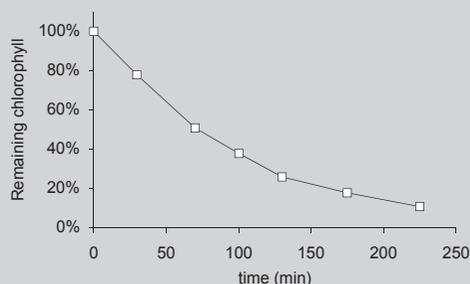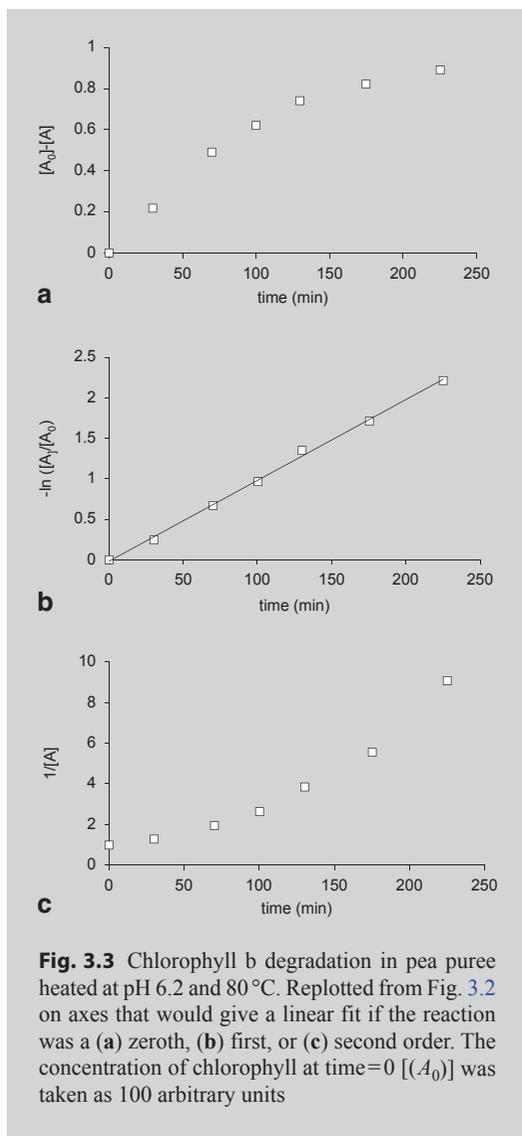

Fig. 3.2 Chlorophyll b degradation in pea puree heated at pH 6.2 and 80 °C. Chlorophyll concentration is reported relative to the concentration present at the start of the experiment. (Replotted from Ryan-Stoneham and Tong 2001)

an apparently first order mechanism and the rate constant can be calculated from the negative slope of the line. If none of the functions fitted the data well, we could conclude that either error in the measurements prevents an adequate fit or that the models we are testing are appropriate (e.g., n is not an integer). It is important to remember that determining order and rate by this method is a curve-fitting exercise and depends on the quality and quantity of the data available. For example in Fig. 3.2 and 3.3 if there were only 50 min of data available, it would be almost impossible to say which order of reaction was most appropriate.

Fig. 3.3 Chlorophyll b degradation in pea puree heated at pH 6.2 and 80 °C. Replotted from Fig. 3.2 on axes that would give a linear fit if the reaction was a (**a**) zeroth, (**b**) first, or (**c**) second order. The concentration of chlorophyll at time=0 [(A_0)] was taken as 100 arbitrary units

It may be impossible to know the molecules responsible for important changes in foods. For example, a strawberry packer might measure the proportion of fruit with mold defects as a function of time, a commercial baker might use an online color meter to measure the brown crust development in bread in the oven, or a coffee roaster might use a sensory panel to assess the affect of roasting time and temperature on the development of a "charred" flavor note of the final beverage. While none of these are measurements of molecular concentration, they could all be useful for kinetic analysis. In the last example, it might

be helpful to imagine the reaction being modeled as coffee with no charred flavor → coffee with charred flavor and by measuring the taste intensity as a function of time, calculate an apparent rate constant.

3.4 Kinetics and Mechanism

Equation 3.2 and its integral forms are empirical in nature and are practically useful in describing change. The observed kinetics are also a reflection of the underlying molecular mechanism of the reaction. We shall use a few simple examples to see the nature of this relationship but first a word of caution: Just because a reaction fits well with a mathematical model does not prove anything about the molecular mechanism. It is possible, although complex, to use kinetic experiments to disprove proposed mechanisms but not to positively prove them. We will return to this question after considering some very simple relationships between mechanism and kinetics.

First-order Reaction: Bacterial Growth Bacteria multiply by dividing in two. If a bacterium divides once per hour then how will the number of bacteria change with time? If there were one bacterium at the start of the experiment, there would be two after an hour, four an hour later, and so on. The rate of the reaction (i.e., number of bacteria formed in an hour) is equal to the number of bacteria present at the start of that hour and we can write an expression for rate as a first-order reaction:

$$Rate = \frac{d[bacteria]}{dt} = k \cdot [bacteria]^1$$

Second-order Reaction: Foam Collapse Assume the rate at which foam breaks down depends on the rate two small bubbles merge to form a single, larger bubble (Fig. 3.4). How would the number of bubbles change with time? The rate depends on the number of bubble–bubble encounters in a given time. The chance of an individual bubble colliding with and merging with another bubble in a given period of time
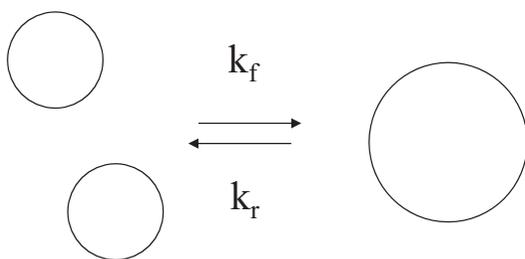
Fig. 3.4 A possible mechanism for the collapse or formation of a foam. In the forward reaction, two bubbles merge to form a larger one and in the reverse reaction, one large bubble is broken into two small ones

depends on the number of bubbles present, that is, [bubbles]. However, the total number of collisions would be this number summed for all of the many bubbles present, that is, $[bubbles]^2$. The rate of the reaction (i.e., number of bubbles lost per unit time) is proportional to the square of the number of bubbles present and we can write an expression for rate as a second-order reaction:

$$Rate = \frac{-d[bubbles]}{dt} = k \cdot [bubbles]^2$$

Zeroth-order Reaction: Enzymatic Reactions Many catalyzed reactions are zeroth order if the catalyst is "saturated", that is, all its active sites in use at all times. Under these conditions, adding more substrate will not further increase the reaction rate because the availability of catalyst concentration, not the substrate concentration, is the limiting factor. At lower concentrations of course, the enzyme is not saturated, so adding more substrate will allow the reaction to go faster. (A nonchemical analogy here would be an industrial production line. The rate of product formation depends on the slowest machine on the line, not the amount of raw material in the warehouse.) This topic is discussed in greater depth in Sect. 3.6 below.

In all these simplest examples, it was possible to intuitively relate a proposed mechanism with the measured kinetics. In essence, the kinetic model provides a hypothesis to be tested against the observed kinetics. However, most real chemical reactions involve a complex multistep mechanism. For example the oxidation of glucose has

the following stoichiometry: $C_6H_{12}O_6 + 6\,O_2 \rightarrow 6\,CO_2 + 6\,H_2O$. But this does mean that six oxygen molecules simultaneously collide with one glucose molecule, form an active intermediate, then break down to form carbon dioxide and water. The reaction is not sixth order with respect to oxygen. Similarly a reaction with forward and backward processes occurring simultaneously might give an unexpected order (Fig. 3.4). A real effort to develop a kinetic test for a chemical mechanism is an involved process requiring many series of measurements at different concentrations of each of the reagents and catalyst (see several of the books in the bibliography for more details). For most physical changes though the kinetics are used in a more descriptive manner and only very rarely to understand the mechanism.

3.5   Effect of Temperature on Reaction Rate

The rate of a reaction may change with time as the concentration of reagents change, but the rate constant will remain constant as long as the reaction mechanism and the conditions (e.g., temperature and pH) do not alter. The effect of temperature on the rate constant is often modeled using the empirical Arrhenius equation:

$$k = k_0 \exp\left(-\frac{E_a}{RT}\right) \tag{3.3}$$

where k is the rate at absolute temperature T, E_a is the activation energy of the reaction R is the gas constant ($= 8.314\ \mathrm{J \cdot K^{-1}\ mol^{-1}}$) and k_0 the frequency factor. By measuring rate as a function of temperature, the constants in Eq. 3.3 can be experimentally determined and used to calculate rates at other temperatures. Some of the strengths and weaknesses of this approach are illustrated in the following examples.

Example: Effect of Temperature on Color Loss Kinetics in Peas
Ryan-Stoneham and Tong (2000) repeated their measurements of chlorophyll degradation in peas over a series of different
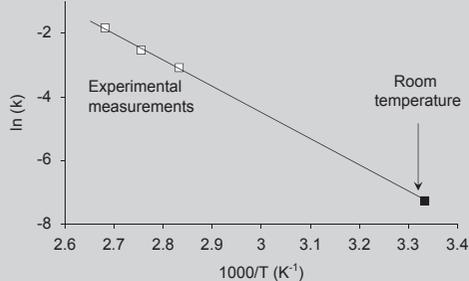
Fig. 3.5 Arrhenius plot showing the effect of temperature on the rate of chlorophyll degradation in peas. The line shown is a best fit to the experimental (*open*) points and is extrapolated to estimate the rate constant at room temperature (*filled point*). (Data from Ryan-Stoneham and Tong (2001))

temperatures. The rate of loss increased with temperature but the general shape of the concentration–time curves was similar, suggesting the same first-order mechanism was responsible in each case and a the rate constant was calculated at each temperature. Equation 3.3 was used in its linearized form ($\ln k = \ln k_0 - E_a/R \cdot 1/T$) to calculate the activation energy of the reaction from a plot of the logarithm of the measured rate constants against reciprocal absolute temperature (Fig. 3.5; note in the figure, the x-axis is reciprocal temperature so the hottest sample, 100 °C is to the left). The slope of the line is $-E_a/R$ and can be used to calculate the activation energy ($= 68.1$ kJ mol^{-1}) and the intercept is $\ln k_0$, the frequency factor ($= 20.1$, $k_0 = 5.4 \times 10^8$ min^{-1}).

Knowing E_a and k_0, it is possible to calculate the rate constant at any temperature. This is particularly important in shelf-life testing of products. Often the desired shelf-life may be several years and it is not practical to test formulations over that period before they are brought to market. An alternative is to measure the kinetics of decay at several higher temperatures and use an Arrhenius approach to calculate the rate constant at lower temperatures. So using the data in Fig. 3.5, we could estimate the rate constant at room temperature (300 K) as 0.0007 min^{-1} and use this in the first-order rate equation from Table 3.1 to

estimate the time for, for example 50 % of the chlorophyll to be lost at room temperature (16 h). This type of approach is fraught with difficulties associated with extrapolation. First, there is considerable error associated with extrapolating a best-fit line over a wide temperature range. A small uncertainty in the values of slope and intercept for the line will lead to larger errors in the extrapolated $ln(k)$. Second, the reactions important at one temperature may not be relevant at another. As a simple example, fresh-cut fruit will brown by an enzymatic mechanism and baked apple will also go brown via the Maillard reaction. Trying to use measurements of browning kinetics at one temperature to predict the browning rate at another, radically different temperature would be fruitless, as the mechanism responsible has changed. One indication of a changing mechanism is a change of slope in the Arrhenius plot as illustrated in the following example.

Example: Kinetics of Milk Protein Denaturation

α-lactalbumin is an important protein in the whey fraction of milk. It can be denatured (i.e., unfolded from its physiological structure—see Chap. 7 for more details) during thermal processing. Anema and McKenna (1996) used gel chromatography to separate the proteins in heated reconstituted whole milk and measured the concentration of each by staining the gel with a protein-sensitive dye. The rate of α-lactalbumin loss due to denaturation was shown to follow first-order kinetics and the effect of temperature on the rate constant was plotted Fig. 3.6) as an Arrhenius relationship but there was no simple linear relation between $ln(k)$ and $1/T$. The authors interpreted their data to suggest that there are two mechanisms important for protein denaturation. Both mechanisms have the same measured consequences (i.e., loss of native protein) but the mechanism important at high temperatures has a lower activation energy (slope of the best-fit line) than the mechanism important at
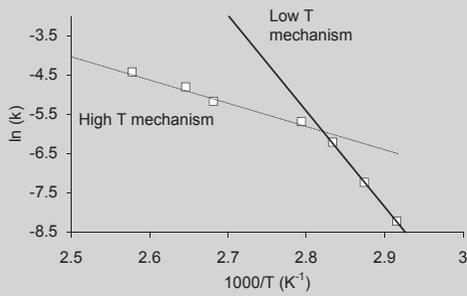
Fig. 3.6 Arrhenius plot showing the effect of temperature on the rate of α-lactalbumin degradation in heated reconstituted whole milk. There are two mechanisms responsible for the degradation, each with a different activation energy (slope of the regression lines shown). (Adapted with permission from Anema and Mckenna (1996). Copyright 1996 American Chemical Society).

low temperatures. If this type of data were to be used in an accelerated test, then the measurements of high temperature rates constants would overestimate of the rate constant at lower temperatures if the low temperature line were simply extrapolated.

The Arrhenius approach is empirical, but it bears a close relation to our understanding of equilibrium energy distributions described by the Boltzmann distribution (Eq. 1.4). Returning to our thermodynamic basis for kinetics (Fig. 3.1), we could argue that the rate of the forward reagents → products reaction is proportional to the fraction of the reagent molecules with energies greater than $\Delta E_f^{\ddagger}$. In effect, the Arrhenius relationship is telling us that this fraction, and thus the rate of the reaction, increases with temperature according to the Boltzmann distribution. The frequency factor is the rate at infinite temperature where all of the molecules have sufficient energy to clear the barrier and the rate is limited only by the rate at which they encounter one another. A more theoretically based approach to the problem was developed by Eyring and others who argued that there was a transition state between reagents and products that exists very briefly: A few molecular oscillations, before breaking down to product or returning to the reagents. The measured rate of the reaction

is related to the Gibbs free energy of forming the transition state from the reagents ($\Delta G^{\ddagger}$) as:

$$k = \frac{k_B T}{h} \exp\left(\frac{-\Delta G^{\ddagger}}{RT}\right) \qquad (3.4)$$

where h is Planck's constant ($= 6.62 \times 10^{-34}$ Js^{-1}). The Gibbs free energy can be split into enthalpic ($\Delta H^{\ddagger}$) and entropic ($\Delta S^{\ddagger}$) contributions as $\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}$, so:

$$k = \frac{k_B T}{h} \exp\left(\frac{-\Delta H^{\ddagger}}{RT}\right) \exp\left(\frac{\Delta S^{\ddagger}}{R}\right) \quad (3.5)$$

and the values of $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ can be estimated from the activation energy and frequency factor of an Arrhenius plot as:

$$\Delta H^{\ddagger} \approx E_a \text{ and } \Delta S^{\ddagger} \approx R.\ln\left(\frac{k_0 h}{k_B T}\right) \qquad (3.6)$$

Note that $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ all refer to the formation of the transition state from the reagents and not the overall changes in Gibbs free energy, enthalpy, and entropy of the reaction. They are useful parameters as knowing them allows some assessment of the nature of the unobserved intermediate state and thus the pathway of the reaction.

Example: Formation of a Transition State During α-lactalbumin Denaturation
Anema and McKenna (1996) used their measurements of the temperature dependence of whey protein denaturation rates (see Fig. 3.6) to calculate the free energy, enthalpy, and entropy changes associated with forming the transition state from native α-lactalbumin. For both the high and low temperature mechanisms, the Gibbs free energy to form the transition state was about the same 105–110 kJ·mol^{-1}. However for the low-temperature mechanism, $\Delta H^{\ddagger} = 192$ kJ·mol^{-1} and $\Delta S^{\ddagger} = 0.24$ kJ·mol^{-1} K^{-1} while for the high temperature mechanism, $\Delta H^{\ddagger} = 54.5$ kJ·mol^{-1} and

$\Delta S^{\ddagger} = -0.14$ kJ·mol^{-1} K^{-1}. That is to say at low temperatures there was a bigger increase in enthalpy to form the transition state than at high temperatures suggesting more chemical bonds needed to be broken. Similarly, the change in entropy was positive for transition state formation by the low temperature mechanism, suggesting the transition was more disordered than the native protein while the entropy change was negative for the high temperature mechanism, suggesting that the transition state was more ordered than the native proteins. They argued that chain unfolding should have a high $\Delta H^{\ddagger}$ as many bonds are broken and a positive $\Delta S^{\ddagger}$ as the unfolded product is more disordered. An aggregation reaction on the other hand has a lower $\Delta H^{\ddagger}$ as fewer bonds are broken and a negative $\Delta S^{\ddagger}$ the product is more ordered. Therefore, that while the protein is denatured at both low and high temperatures, unfolding represents the rate-limiting step at low temperatures and aggregation at high temperatures.
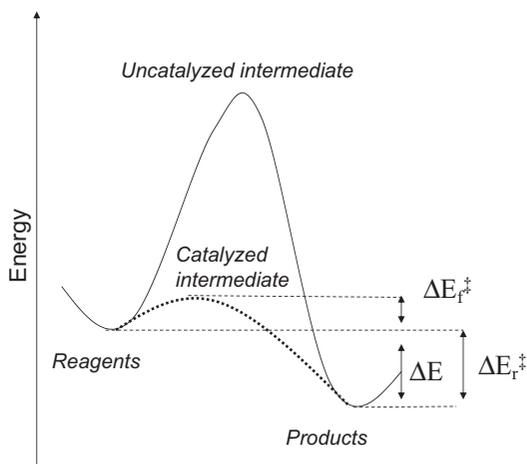
Fig. 3.7 Effect of a catalyst on a reaction surface: The rate of the forward and reverse reaction is increased while the equilibrium distribution is unaffected
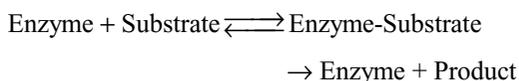
3.6 Catalysis

A catalyst is a substance that accelerates the rate of a reaction but is not consumed by it. The catalyst has the effect of stabilizing the intermediate and lowering its energy (Fig. 3.7). A higher proportion of molecules are able to form the lower energy intermediate and so the rate of the forward and reverse reaction is accelerated. Importantly the catalyst does not affect the energies of the reagents or products, so the equilibrium position of the reaction is unaffected.

An enzyme is a protein that can very specifically bind certain compounds and catalyze their transformation. Enzymes are essential to life as they enable cells to regulate their chemistry by controlling the kinetices. For example, high-energy chemicals (e.g., fats, starch) can be stored for long periods and then broken down when re-

quired by controlling the release of an enzymatic catalyst. Enzymes can remain active in fresh foods but as the plant or animal that made the tissue is dead and has lost control of the cellular processes, they can sometimes react in unexpected and undesirable ways. For example, lipase enzymes are important in lipid digestion, where they catalyze the removal of fatty acids from the glycerol backbone of a triacylglycerol. They are also present in unpasteurized milk, and if the milk is mixed vigorously, the membrane surrounding the fat globules can be damaged, allowing the enzyme access to the lipid and very quickly producing a characteristic off-flavor due to the release of butyric acid. Enzymes are also used in food processing (e.g., amylase enzymes used to hydrolyze starch to form corn syrup, chymosin used to cleaving κ-casein to form a yogurt gel from milk) and of course in the digestion of foods.

The mode of action of most enzymes involves a two stage process (i) binding the substrate and (ii) facilitating the chemical transformation and releasing the product:
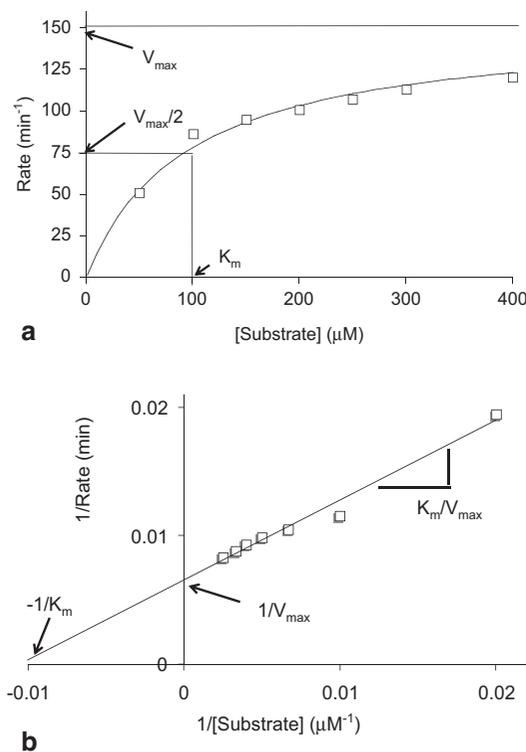
$$\text{Enzyme} + \text{Substrate} \rightleftharpoons \text{Enzyme-Substrate}$$
$$\rightarrow \text{Enzyme} + \text{Product}$$

Fig. 3.8 (**a**) Rate of lipase catalyzed hydrolysis of a triacylglycerol as a function of substrate concentration. (**b**) Lineweaver–Burk plot of the same data. (Unpublished data courtesy of Dr. Josh Lambert (Penn State University))

To analyze this mechanism, the rate of the reaction (v) is measured as a function of substrate concentration ([S], Fig. 3.8a):

- When there is no substrate, the reaction does not take place so rate is zero.
- At low substrate concentrations, the enzyme is able to catalyze the chemical change faster than it can bind more substrate. The overall rate is limited by the first step of mechanism and the increases with increasing substrate concentration. The capacity of the enzyme to bind limited substrate is the binding coefficient (K_m).
- At high substrate concentrations, the enzyme is saturated. The overall rate of the reaction depends on the second step of the mechanism and adding more substrate makes little difference. The maximum rate possible (V_{max}) is seen when the enzyme is saturated with enzyme.

The overall dependency of rate on substrate concentration is given by the Michaelis–Menten equation:

$$v = \frac{V_{max}[S]}{K_m + [S]} \tag{3.7}$$

Equation 3.6 is more commonly seen in its linear form:

$$\frac{1}{v} = \frac{K_m}{V_{max}} \cdot \frac{1}{[S]} + \frac{1}{V_{max}} \tag{3.8}$$

and plots of reciprocal rate against reciprocal concentration can be used to calculate K_m and V_{max} (i.e., a Lineweaver–Burk plot, Fig. 3.8b).

Nonenzymatic catalysts are important in industrial chemistry; for example liquid oils are hydrogenated to form solid fats by reaction with hydrogen under high pressures in the presence of a Raney nickel catalyst is very slow. Raney nickel is treated to produce a very high surface area per unit mass (~ 100 m^2 g^{-1}) that can adsorb the reagents and lower the free energy required to form the intermediate and thus catalyze the reaction. Purely physical transformations can also be catalyzed. For example, a glass of a carbonated drink will bubble slowly because the chemical potential of carbon dioxide is lower in the atmosphere than in the solution. The process is slow because of a free energy barrier associated with forming a small bubble. The reaction can be catalyzed with a spoonful of sugar or salt that will cause the drink to fizz spectacularly as the formation of small bubbles is catalyzed at the solid surface. (It is easy to show that the solid surface is important, as a spoonful of sugar solution has no comparable effect.)

3.7 Summary

Kinetics is distinct but complementary subject to thermodynamics. It postulates the existence of a high-energy intermediate state that reacting molecules must overcome to form products. The energy barrier slows down change and even stops the system reaching equilibrium, but has no affect on the final composition at equilibrium. Reactions in living systems are very tightly ki-

netically regulated and nonequilibrium states are the rule rather than the exception. Similarly few foods are consumed in their equilibrium states.

The empirical rates of reaction are proportional to the concentrations of reagents raised to a power and the integral form of this relationship can be used to model changes in concentration with time. The proportionality constant (the rate constant) usually increases with temperature according to the Arrhenius equation, which itself is similar to the Boltzmann distribution in that it provides a measure of the proportional of molecules with sufficient energy to react at a given temperature. Catalysts, particularly enzymes, can be used to increase the rate of biological reactions by lowering the activation energy needed to produce the intermediate.

In the next chapter, we will bring together the laws of thermodynamics, the properties of molecules, and the kinetic theory studied here to investigate the general phenomena of why some food ingredients mix and others will more or less quickly separate into two phases.

3.8 Bibliography

This very introductory description of kinetics is readily extended by most physical chemistry textbooks (Atkins and De Paula 2006; Tinoco et al. 2002) and I have found "Reaction Kinetics and Mechanism" (Avery 1974) particularly helpful. Dill et al. (2003) cover much of the same material in Chap. 18 and 19 of "Molecular Driving Forces" but use a more intuitive, molecular approach. Walstra considers similar material with a focus on foods in Chap. 4 of "Food Physical Chemistry" (Walstra 2003).

Phase Behavior

4.1 Introduction

Some ingredients will dissolve in one another while others do not, or at least not completely. For example, ethanol and water can be blended at all proportions, as can olive oil and canola oil. On the other hand, while oil and water can be combined to make a salad dressing, they do not actually dissolve in one another. Often the tendency to dissolve depends on the conditions; if you add sugar to iced tea, the first few crystals will dissolve, sweetening the drink, but sugar added beyond a certain limit will sink to the bottom of the pitcher and not be tasted. It is possible to make much sweeter hot tea as the solubility limit of sucrose in water increases with temperature.

If ingredients do not dissolve, they must form separate phases within the food. In the language of physical chemistry, a phase is a region where at least some of the properties (e.g., chemical composition, density, viscosity) change abruptly at the boundary. The change in properties must persist over several molecular dimensions; while the dissolved lactose molecules in cream are not each treated as separate phases but as part of an aqueous solution phase, the oil in micrometer-sized droplets is an independent liquid phase (Fig. 4.1). While cream appears to the naked eye to be a homogeneous fluid, it is microscopically a two-phase system. The presence of fine oil droplets affects the texture of the product, for example higher fat creams are more viscous than lower fat ones (Chapter 8) The phase behavior of real foods can be very complex. For example, if

cream is sweetened, whipped, and frozen to form ice cream, the final product we eat has a crystalline ice phase, a concentrated sugar solution phase, a lipid phase (in droplets), and some bubbles of an entrained gas phase (Fig. 4.2). Once more, the properties of the product depend on the proportions of each phase present, for example, if too much of the water is frozen into ice or too little air whipped in, then the product would be hard and difficult to scoop. The balance of phases present depends on the conditions (i.e., temperature), and the ingredients used (e.g., an ice cream formulated with high sugar content will tend to be softer and have less ice).

The goal of this chapter is to first understand how to use phase behavior to understand and predict the properties of a food and second to understand phase behavior in terms of molecular interactions and the laws of thermodynamics. It is usually productive to start with a simple system, understand it thoroughly, and then add complexity later. With that in mind, we will consider the phase behavior of the simplest of foods—water.

4.2 Single-Component Phase Diagrams

Water is a unique material as it occurs on our planet and in our food in abundant quantities in its solid, liquid, and gaseous phases. We could make experimental measurements of the freezing point and boiling point of water and plot them on a line to show the conditions at which the dif-
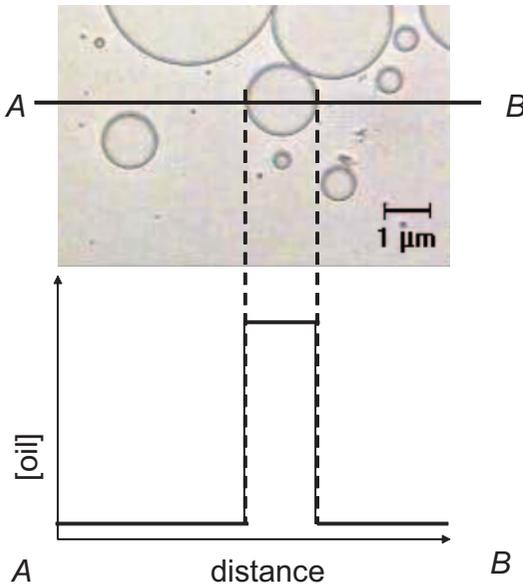
Fig. 4.1 Optical micrograph showing the *oil* and *water* present as separate phases in a model cream. The concentration of oil changes abruptly at the oil–water interface, justifying us describing it as a separate phase
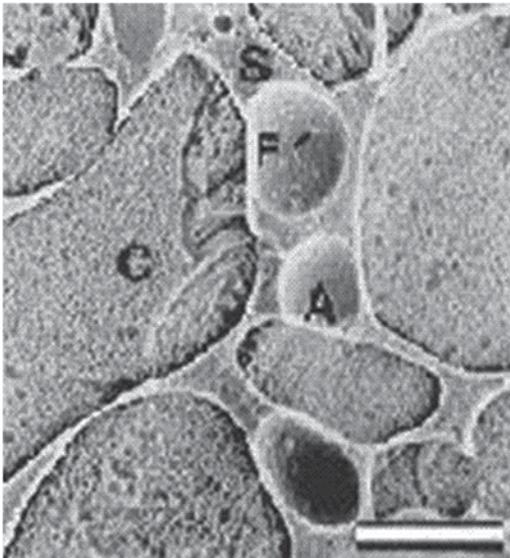

Fig. 4.2 Electron micrograph showing the different phases present in ice cream: ice crystals (*C*), air bubbles (*A*), unfrozen serum phase (*S*), partly crystalline fat droplets (*F*). (Image courtesy Dr. Douglas Goff, University of Guelph)
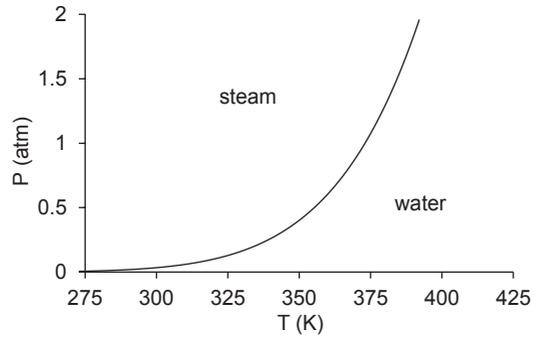

Fig. 4.3 Phase diagram of pure water (NIST 2014)

ferent phases are seen: Above 100 °C steam is the stable form of water and the liquid will boil, below 0 °C ice is the stable form and water will crystallize. Precisely at 100 °C and 0 °C, we can say water is in equilibrium with steam and ice respectively and we can see two phases together. Alternatively, if we have a glass of iced water or a pan of boiling water, we know the temperature of the water, precisely because there are two phases present simultaneously.

The boiling point of water is also a function of pressure (water boils at only about 69 °C at the top of Mount Everest). We can show the effects of temperature and pressure on the phase behavior of water by using them as the axes of a phase diagram (Fig. 4.3). The labeled regions show the conditions where the different phases of water (ice, water, and steam) are seen. For example, under ambient conditions ($p = 1$ atm, $T = 25$ °C) water is liquid but at -10 °C it is solid. We can freely vary temperature and pressure within one of these regions so long as we do not cross one of the lines that represent the conditions for a phase transition (i.e., boiling, freezing, or sublimation). Along the lines, two phases are at equilibrium with one another and if we want to maintain this equilibrium, we are not free to independently change temperature and pressure. For example, if we have boiling water and we want to raise the temperature to 121 °C, the pressure must increase to 2 atm to stay on the line. If water is boiling at 69 °C, the pressure must be 0.28 atm because of the phase line. The point where the boiling and freezing lines intersect is the triple point of water (0.01 °C, 0.006 atm) and these are the
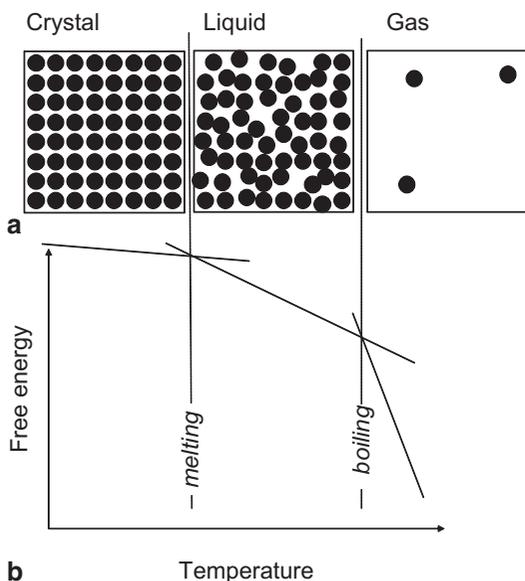
Crystal   Liquid   Gas

a

b   Temperature

Fig. 4.4 **a** A cartoon representation of the arrangement of molecules in a crystalline solid, a liquid, and a gas phase. **b** The free energy of the different phases as a function of temperature

only conditions where it is possible to see steam, water, and ice are at equilibrium.

A practical application of the water phase diagram is seen in operation of the retorts used to cook canned foods. A retort is an industrial pressure cooker; the cans of food are sealed into a container partly filled with water and heated. In an open container, the water would boil at a constant temperature until all the water had evaporated before the pan itself would begin to heat further. However, inside a sealed retort, the water would initially boil at 100 °C but as soon as some steam was generated, the pressure would increase and the boiling would stop. The retort would then have to be heated a little further to reach the boiling point of water at the higher pressure but as soon as it boiled again, the pressure and hence the temperature required for further boiling would increase. In practice, the temperature and pressure inside the retort would increase together—so long as there was water and steam present together. We are used to cooking times in domestic recipes being specified in terms of time and oven temperature but retorting processes are usually specified in terms of a cooking time and a retort pressure which fixes the temperature according

to the phase diagram (and has the additional advantage of being easier to measure inside a sealed container).

What is it about water molecules that cause them to behave in this way? We can go some way to understanding this using a simple cartoon representation of the crystalline, liquid, and gaseous states (Fig. 4.4a). In a crystalline solid, the molecules are regularly packed and strongly associated with their neighbors. Each molecule may vibrate about its position but not diffuse away from it. In a liquid, the molecules are still closely packed, but now free to diffuse from their starting positions and there is no persistent regular arrangement of the molecules. In a gas, the molecules are no longer closely packed, and can diffuse freely. The overall free energy of a phase is a combination of enthalpy terms (bonding) and entropy terms ("disorganization") i.e.: $G = H - TS$ (Eq. 1.10). In a crystal, there are many strong bonds so the energy term is large but as the structure is regular, the entropy is small. In a gas, there are few bonds so bonding energy is low and but the entropy high. The properties of a liquid lie between these extremes, but are usually closer to those of a crystal than those of a gas.

The effect of temperature on the free energy of a gaseous, liquid, and crystal phase is shown schematically in Fig. 4.4b. At low temperatures, the entropy term is relatively unimportant so the free energy is dominated by the bonds present and the strong bonds in the crystalline phase means it has a lower free energy than the gas (with the liquid as an intermediate state). The stable phase will be the one with the lowest free energy so materials tend to crystallize when cooled. Free energy decreases with temperature for all materials (again see Eq. 1.10, $G = H - TS$), but as the entropy of the gas phase is greater than the crystal phase, the rate of decrease is greater (with the liquid somewhere between the two). The temperatures where the free energy curve for the liquid phase intersects with those of the solid and gaseous phases correspond to the melting and boiling points respectively.

Another way of looking at boiling and freezing would be to say that higher temperatures give the molecules sufficient thermal energy to break

the bonds holding them together. Similar thinking can help understand the effects of pressure: Compression forces the molecules together into more condensed phases. Because increasing temperature and pressure have opposing effects on phase behavior, we are free to use changes in one to compensate for changes in the other. In effect, that is what is happening in the retort example earlier: Water boils at 100 °C at atmospheric pressure but we can make it boil at a higher temperature by using increased pressure to hold the higher energy molecules together in a dense phase. Applying too much pressure and the steam would condense into the water phase—too little pressure and all the water would boil off into steam.

The one-component system was helpful to understand phase transitions in terms of molecular interactions and entropy but to begin to understand foods we must now add more components and look at the phase behavior of mixtures. In this case, we will encounter additional complications due to the entropy effect of mixing (favoring miscibility) and the different types of possible molecular interactions. We will begin by examining phase diagrams as practical descriptions of the behavior of mixtures and then try to provide some sort of theoretical framework.

4.3　Multicomponent Phase Diagrams

The first challenge in drawing a multicomponent phase diagram is to represent composition as an axis on the plot. We can succinctly express the composition of a two-component mixture on a phase diagram as a mass or mole fraction. For example dissolving 2 g of sugar in 10 g of water would produce a 16.7 wt% (i.e., $2/(2+10)\times 100\%$) solution. Knowing the fraction of one component fixes the other (i.e., 83.3 wt% water) and so we can describe the overall composition of a mixture using only this one number. If we use one axis for composition, then the remaining one can be either temperature or pressure but we cannot show the effects of both parameters simultaneously. Temperature is usually more relevant to foods so a typical two-component phase

diagram represents experimental measurements of the phase(s) present in a mixture as a function of temperature and composition. It is possible to draw phase diagrams for more complex mixture of ingredients but in practice it is unusual to see data for more than a ternary mixture. For a three-component mixture A + B + solvent, one axis represents the mass fraction of A and the second the mass fraction of B.

A phase diagram shows the phases at equilibrium with one another under given conditions. How the data is generated depends on the nature of the system but as an example consider how we might form a sugar–water phase diagram. We could just add sugar to the water until no more dissolves, and mark that concentration as the solubility limit. In practice, the process of dissolution can be prohibitively slow and it works better to just add an excess of sucrose, let the undissolved crystals sediment out, and measure the concentration of the saturated sucrose solution (Fig. 4.5a). Repeating the experiment over a range of temperatures, we could generate a phase boundary in the sucrose–water phase diagram (Fig. 4.5b). Any point on the phase line shows the composition of the sucrose solution at equilibrium with sucrose crystals. Mixtures more dilute than the solubility line are solutions, while mixtures that are more concentrated will phase separate as sucrose crystals and a saturated sucrose solution.

Real mixtures show many types of phase transition, for example sucrose–water will boil (i.e., phase separate into sucrose solution and water vapor) if heated or freeze (i.e., phase separate into sucrose solution and ice) if cooled. Measuring these transitions might be as simple as measuring the boiling point and melting point of different sucrose solutions. Plotting all the transition lines on a single figure yields the complete phase diagram (Fig. 4.6).

The main value of a multicomponent phase diagram is to calculate the composition of a system at equilibrium. This is done by drawing a tie line on the phase diagram which connects the composition of phases at equilibrium with one another. In two-component phase diagrams, the tie line is always parallel to the x-axis. For ex-
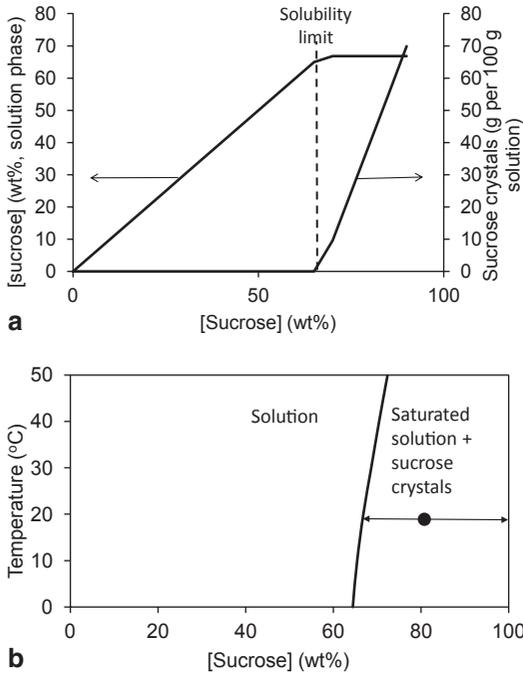
Fig. 4.5 **a** Concentration of the solution and mass of any crystal present as a function of added sucrose concentration at 20 °C. The solubility limit of sucrose in water is 66.8 %. **b** Phase line showing the solubility of sucrose in water as a function of temperature. The *arrows* show the tie line connecting the phases at equilibrium for an 80 % sugar water mixture at 20°C
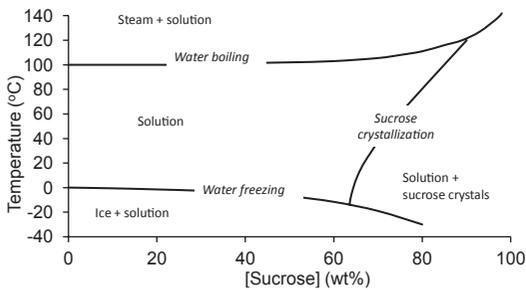

Fig. 4.6 The sucrose–water phase diagram. (Data from Peres and Macedo 1996 and Young and Jones 1949 as well as measurements in the author's laboratory)

ample in Fig. 4.5b, an 80 % sugar–water mixture phase separates into sucrose crystals and a 66.8 % sucrose solution (compare to Fig. 4.5a). This approach can be readily extended to calculate the amounts of each phase by a mass-balance as illustrated in the following example.
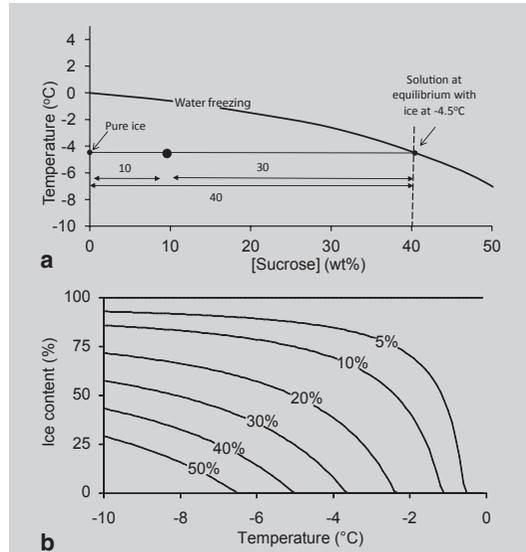
Fig. 4.7 **a** Freezing point of sucrose–water solutions. *Construction lines* show the application of the phase rule to a 10 % sucrose solution cooled to −4.5 °C. **b** Ice content of various sucrose solutions as a function of temperature calculated from the phase diagram

Example: Freezing Sorbet

Pure ice melts completely and suddenly at 0 °C but, in a frozen solution, the product begins to melt progressively from much lower temperatures. Sorbet is hard in the freezer but, slowly softens as it warms up and the amount of ice slowly decreases. If we model sorbet as a sucrose–water mixture, then we can use the freezing line from the sucrose–water phase diagram (Fig. 4.7a) to calculate the composition of the solution phase in equilibrium with ice as a function of temperature. For example, if the sorbet mix was 10 % sucrose and it was cooled to −4.5 °C, there is a phase separation as ice forms and comes to equilibrium with the concentrated sucrose phase. The composition of the two phases is calculated by drawing a horizontal tie line (i.e., pure ice and 40 % sucrose solution) and the relative amounts of each phase is calculated as a mass balance. Starting with 100 g of sorbet mix (i.e., 90 g water and 10 g sucrose), in the frozen state all of the sucrose was in

the solution phase plus enough of the water to make the solution concentration 40 %, i.e.: $0.4 = 10/(x + 10)$ where x is the amount of water in the solution phase. Solving the mass of water in the solution $x = 15$ g, so the mass of the solution phase is 25 g and, by difference the mass of the ice phase is 75 g. The mass balance can also be handled more simply using the lever rule; the fraction of each phase is equal to the distance along the phase line from the initial composition to the phase line divided by the total length of the phase line (see construction lines in Fig. 4.7b):

- Fraction of ice = 30/40 = 75 %
- Fraction of solution = 10/40 = 25 %

Repeating these calculations over a range of temperatures yields a plot of ice content as a function of temperature (Fig. 4.7b). A softer, lower ice product could be made by increasing the sugar content of the mix.

Example: Polymer Mixtures

High molecular weight polysaccharides (glucans) are believed to contribute to the cholesterol-reducing effects of oats in the diet. Kontogiorgos et al. (2009) were interested in ways these polysaccharides could be combined with other food polymers. While the individual polymer solutions were clear, some of their mixtures were cloudy. Turbidity in liquids is an indication that there are small particles of a dispersed phase scattering the light. Centrifuging the cloudy liquids gave a pellet of precipitate in a clear liquid (Fig. 4.8). By measuring the concentration of each polymer in the liquid phase, they could measure the composition of the liquid in equilibrium with the solid and plot a phase diagram. We will return to the ways polymer phase diagrams differ from those of small molecules in Sect. 7.6.
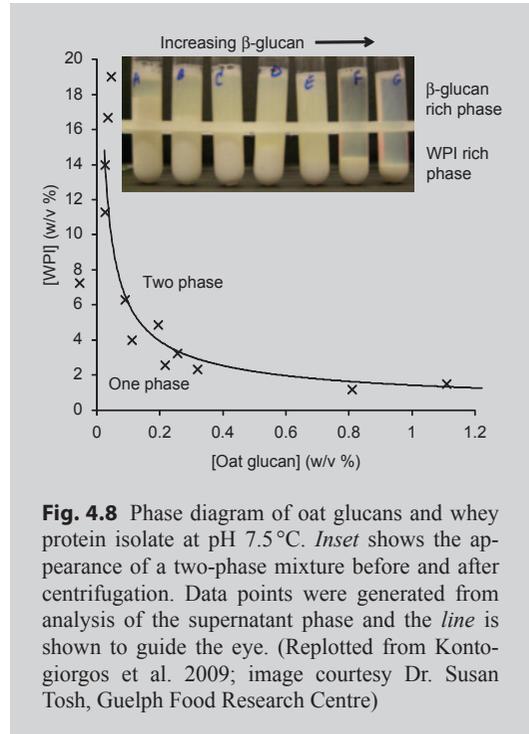

Fig. 4.8 Phase diagram of oat glucans and whey protein isolate at pH 7.5 °C. *Inset* shows the appearance of a two-phase mixture before and after centrifugation. Data points were generated from analysis of the supernatant phase and the *line* is shown to guide the eye. (Replotted from Kontogiorgos et al. 2009; image courtesy Dr. Susan Tosh, Guelph Food Research Centre)

Having seen the practical use of phase diagrams, it is helpful to consider their theoretical basis, at least under some idealized conditions. We will consider two special cases, the colligative properties case where the solute affects only the entropy of the solvent, and the regular solution model where enthalpic interactions play a limited role.

4.4 Calculation of Phase Lines—Colligative Properties

Ice has a lower chemical potential than water at −10 °C because the lower-entropy, higher-enthalpy crystal phase is more stable at low temperatures. On the other hand, a concentrated (56 %) sucrose solution is at equilibrium with ice at the same temperature (Fig. 4.6). To remain liquid, the chemical potential of water in the solution must be lowered by the presence of the solute.[1] One

---

[1] Remember that in Sect. 1.9, water activity, a measure of the chemical potential of water in the food, was lowered by the presence of a solute.

Crystal   Solution   Gas
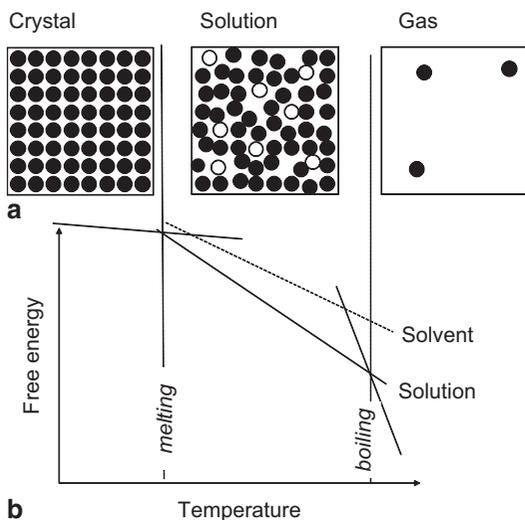
a

b Temperature

Fig. 4.9 **a** A cartoon representation of the arrangement of molecules in a crystalline solid, a solution, and a gas phase. *Filled circles* represent solvent molecules, *open circles* solute molecules. **b** The free energy of the different phases as a function of temperature. The properties of the pure solvent are included as a *broken line* for comparison

reason is that removing water from a solution to form ice leaves the remaining solution more concentrated and the crystallization is opposed by the entropy of mixing. There may also be enthalpic effects as the type of intermolecular bonds change as the solution becomes more concentrated but, to start with, we will treat the solution as ideal and thus governed by entropy changes. The decrease in solvent chemical potential with solute concentration is given by a modification of Eq. 1.18:

$$\mu_{solvent} = \mu_{solvent}^{o} + RT \ln(1 - x_{solute}) \quad (4.1)$$

Thus the chemical potential of the 56% sucrose solution (i.e., mole fraction $x_{solute} = 0.06$) is 161 J mol^{-1} less than pure water. The chemical potential of the solution is lowered by the entropic effects of solute while the properties of the ice phase are not affected. Thus, the freezing point is decreased (Fig. 4.9). Similarly, the properties of the vapor phase are not affected by the presence of the (nonvolatile) solute so the boiling point is increased by the same mechanism.

Assuming the solution behaves ideally, the effect of solution concentration on freezing point can be calculated as:

$$T_{0f} - T_f = \frac{RT_0^2}{\Delta H_{fus}} x_{solute} = K_f \cdot m_{solute} \quad (4.2)$$

The difference between the boiling point of the solution (T_f) and the freezing point of the pure solvent (T_{0f}) is proportional to the mole fraction of the solute (x_{solute}) and inversely proportional to is the enthalpy of vaporization of the solvent (ΔH_{fus}). If the solution is dilute, mole fraction is proportional to concentration in units of molality (m_{solute}, i.e., moles of solute per kilogram of solvent). The constants are often combined as

$$K_f \left(= \frac{RT_0^2}{\Delta H_{fus}} \frac{1000}{M_{solvent}} \right);$$ a constant for a solvent,

e.g., for water $K_f = 0.51$ K kg mol^{-1}.

The parallel equation for boiling point elevation is:

$$T_{0b} - T_b = \frac{RT_0^2}{\Delta H_{vap}} x_{solute} = K_b \cdot m_{solute} \quad (4.3)$$

The difference between the boiling point of the solution (T_b) and the boiling point of the pure solvent (T_{0b}) is proportional to the mole fraction of the solute (x_{solute}) and inversely proportional to the enthalpy of vaporization of the solvent (ΔH_{vap}). If the solution is dilute, mole fraction is proportional to concentration in molality and the constants are often combined as

$$K_b \left(= \frac{RT_0^2}{\Delta H_{vap}} \frac{1000}{M_{solvent}} \right).$$ The parameter K_b is a

constant for a solvent, e.g., for water $K_f = 1.86$ K kg mol^{-1}, i.e., the same amount of solvent increases the boiling point more than it depresses the freezing point.

Equations. 4.2 and 4.3 have no terms for the chemical properties of the solute. The number of solute molecules is important in calculating entropy of mixing but as chemical interactions are neglected for ideal solutions, the chemical properties of the solute are unimportant. Both equations give good predictions of the properties of dilute solutions but at higher concentrations,
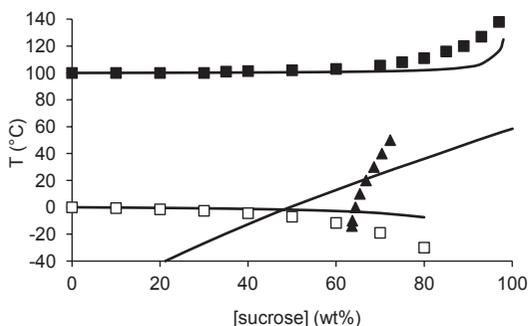
Fig. 4.10 The phase diagram of sucrose showing measured data (*points*) and the predictions based on colligative properties (Eqs. 4.2, 4.3, and 4.4)
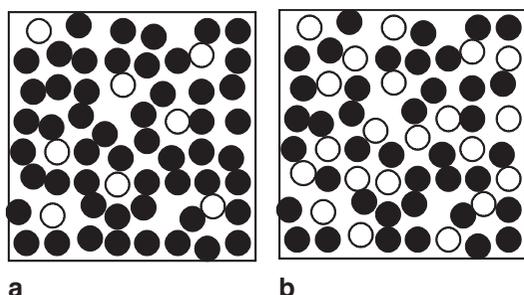

Fig. 4.11 Interactions in dilute (**a**) and concentrated (**b**) solutions. Solvent molecules are represented by *filled circles* and solute by *open circles*. In a dilute solution, the chemical interactions of a given solvent molecule are mainly with other solvent molecules while in a concentrated solution there are many more solvent–solute interactions

they are less reliable (see for example the phase diagram of sucrose, Fig. 4.10). In the dilute solution, most of the water molecules are not close to solute molecules so moving a water molecule from the solution phase to the ice phase does not make much difference to the number or types of intermolecular bond (Fig. 4.11a). At higher concentrations, more water molecules are close to solute molecules so moving a water molecule from the solution phase to the water phase means decreasing the number of solute–water bonds and increasing the number of water–water bonds. The colligative properties theory presented here does not account for this change in bonding interactions so is not reliable.

Finally, we can try to use the same theoretical framework to understand the solubility of a solute in a solvent. A solute lowers the freezing point of the solvent but the words "solvent" and "solute" are just labels we attach to different types of molecules and have no special meaning with regards to the theory. We could equally say a solvent lowers the freezing point of a solute. So while pure sucrose crystals melts at 186 °C, it "melts," or more precisely dissolves, at lower temperatures in the presence of water. The solvent lowers the chemical potential of the liquid phase so it takes less energy to melt the crystals. Using the same logic that led to the equations for boiling point elevation and freezing point depression the solubility limit is given by:

$$\ln x_s = -\frac{\Delta H_{\text{fus}}}{R}\left(\frac{1}{T} - \frac{1}{T_f}\right) \qquad (4.4)$$

where x_s is the solubility limit (in mole fraction units) at temperature T, temperature T_f is the melting temperature of the pure solute, and ΔH_{fus} is the molar enthalpy of fusion of the solute. Note that in this case, the only enthalpy is that of the solute because in this case only the solute can crystallize. The Hildebrand equation becomes less reliable in dilute solutions (see for example the phase diagram of sucrose, Fig. 4.10). This is because the enthalpic effects for the solute become more important rather than less as the solution becomes more dilute. In a highly concentrated solution, most of the sucrose molecules surrounded by similar molecules so moving from the solution phase to the crystalline phase does not make much difference to the number or types of intermolecular bond. In a dilute solution, most of the sucrose molecules are surrounded by water molecules, so moving from the solution phase to the crystalline phase means decreasing the number of solute–water bonds and increasing the number of solute–solute–water bonds. The colligative properties approach does not account for this change in bonding interactions, so is not reliable. The Hildebrand equation is most useful in circumstances where the molecular interactions are less important, e.g., between triglycerides with similar structures.
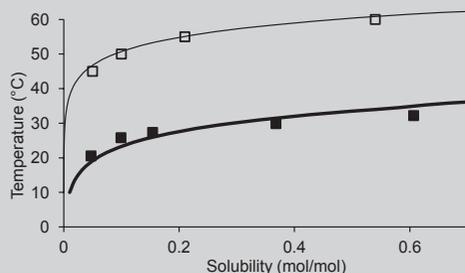
Fig. 4.12 Solubility (expressed as a mole fraction) of tripalmitin (*open points*) and cocoa butter (*filled points*) in liquid palm oil as a function of temperature. *Lines* are predictions from the Hildebrand equation using measured values of the melting point and enthalpy of fusion of the pure solid fat samples. (Calculated from measurements by Zhou and Hartel 2006)

Example: Solubility of Hard Fats in Liquid Oils

Soft, spreadable fats like margarine and shortenings can be made by blending a hard crystalline fat with liquid oil. Some of the fat crystals in the hard fat dissolve in the liquid oil to form a product with an intermediate solid fat content and a desirable soft texture. Most of the molecules in both fat components are triacylglycerols; so although there are some differences in their chemical structure (i.e., typically lower molecular weight and more double bonds in the liquid oil), the interactions between solute (hard fat) and solvent (liquid oil) molecules are similar and therefore the Hildebrand equation may be appropriate.

Zhou and Hartel (2006) mixed solid tripalmitin or solid cocoa butter with liquid palm oil and allowed the mixtures to crystallize and come to equilibrium at different temperatures, before measuring the concentration of the hard fat in the liquid phase. The solubility increased with temperature for both fats and cocoa butter, the lower melting fat, always had lower solids content than corresponding blends of tripalmitin and palm oil (Fig. 4.12). The authors used differential scanning calorimetry to

measure the melting point and enthalpy of fusion of pure tripalmitin crystals (63.4 °C and 150.5 kJ mol^{-1}) and cocoa butter crystals (36.5 °C and 119.4 kJ mol^{-1}). They then used to Hildebrand equation to calculate a good theoretical prediction for the phase behavior of the tripalmitin–palm oil mixture and a rather poorer one for the cocoa butter–palm oil mixture. They argued that the deviation was due to the nonideal nature of the cocoa butter in palm oil solution.

The simple entropy-only effects described in the colligative properties are useful in explaining some, but not all, phase behavior. Importantly the ideal solution assumption meant that molecular interactions were completely neglected. An alternative approach is Hildebrand's regular solution model that allows more complex mixed phases and also different enthalpic interactions between different types of molecules.

4.5   Calculation of Phase Lines—the Regular Solution Model

The colligative properties theory provided a good basis for some phase behavior. In this section, we will attempt another, more general approach to calculating phase lines. This method is rather more involved algebraically, but the process of developing the model provides a link to the molecular properties studied in Chap. 2.

The Hildebrand model describes the phase behavior partially miscible phases, e.g., butanol and water (Fig. 4.13). Butanol dissolves in water up to a certain point then forms a second separate phase of butanol saturated with water. Increasing the ratio of butanol increases the volume of the butanol-rich phase (butanol saturated with water) at the expense of the water-rich phase (water saturated with butanol) until there is none of the latter left. If more butanol is added it will decrease the water concentration in the water-in-butanol solution. There is a concentration range when
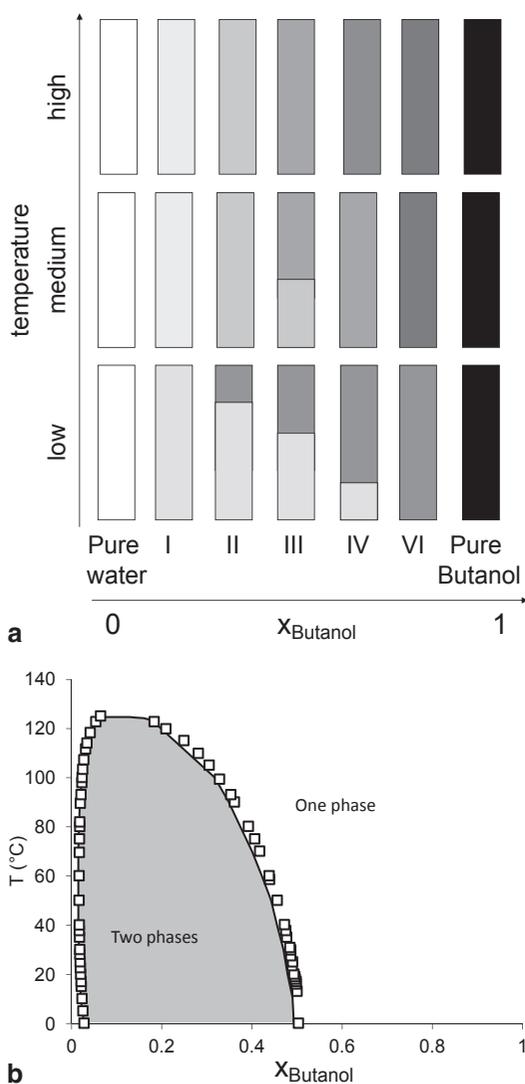
Fig. 4.13 **a** Schematic representation of the phase behavior of water–butanol mixtures at different temperatures. *Depth of shading* indicates butanol concentration. **b** Butanol–water phase diagram from Góral et al. (2006)

butanol water mixtures will form phase-separated regions, and repeating the experiment over a range of temperatures allows us to map this as a phase diagram (Fig. 4.13b). Our goal is to understand the molecular and thermodynamic basis for phase behavior. We will approach this problem using a lattice model.

In a lattice model, the system is divided into a grid of spaces that can be occupied by one and only one "molecule." Figure 4.14 shows a
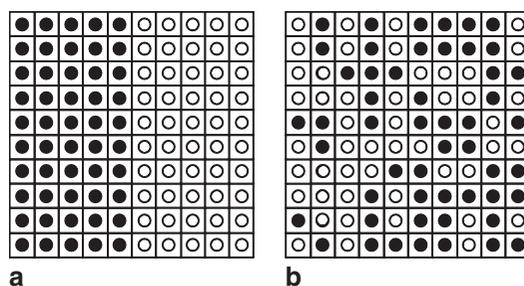
Fig. 4.14 Lattice model for the mixing of two types of molecule. There is **a** one unmixed configuration but many mixed configurations of which **b** is an example

10×10, two-dimensional square lattice containing 50 "black molecules" and 50 "white molecules" in an unmixed state and an example mixed state. The molecules interact with adjacent molecules with characteristic bond energies (w_{BB}, w_{WW}, and w_{BW} for bonds between a pair of black molecules, a pair of white molecules, and a black and a white molecule respectively) but do not interact with molecules more than one square away.

At first glance, this may seem an arbitrary and even preposterous simplification of reality. Real liquid molecules do not sit on a grid and certainly not a two-dimensional one, they move and fluctuate randomly and interact with one another in all of the complex ways described in the previous chapter—how can this checkerboard model hope to teach us anything useful? All these criticisms (and many more) are valid, but their significance is debatable. First, although liquids are free to fluctuate, they will have an average number of nearest neighbors that they most closely interact with (i.e., the coordination number). Perhaps this number is something other than the four implied on our square lattice, but we could easily have drawn other lattices with other shapes or in three dimensionalities and this simple case may serve to illustrate our point. Indeed, we can even leave the coordination number as a variable in our model to take any value we think is most reasonable. Second, although real intermolecular bonds have complex dependencies on molecular separation, the simple lattice rules (no molecules may share the same space, characteristic attraction/repulsion at short ranges, no interaction at

long ranges) bear some similarities to the Lennard–Jones potential in Fig. 2.19. In any case, the only test for any scientific theory is empirical evidence. We will put our doubts to one side and push on with the model; only after comparing its predictions with the behavior of real mixtures of ingredients will we know if our assumptions are reasonable.

Our goal is to compare the free energy[2] of the mixed and unmixed systems to see if we can expect a one- or two-phase system. First, we will look at the contributions due to entropy.

Entropy of Mixing Using a lattice model simplifies this calculation considerably; entropy is due solely to the positions of the molecules on the lattice. The entropy is related to the number of microstates available (Eq. 1.3, $S = k \ln \Omega$), so we need to calculate the number of possible configurations on the lattice for the mixed and unmixed systems. The unmixed case is easy—one microstate and, as $\ln(1) = 0$, the entropy of the unmixed state is zero. Any exchange of a black molecule with a white molecule would count as mixing and not be allowed. Any exchange of a black for a black (or a white for a white) would not count as a new microstate as it would not be distinguishable from the outside (see Chap. 1 for a longer discussion of microstates). The mixed state is more complex.

We will start by calculating the number of ways we could fill up the lattice. If there are N spaces, then we have N choices for the first molecule added. The second molecule has one fewer slot available, so we only have $(N-1)$ choices. The total number of possibilities for the first two choices together are the product of these probabilities, i.e., $N(N-1)$. Continuing this argument until the rest of the slots on the lattice are filled, the total number of ways we could possibly do it is: $N(N-1)(N-2)\ldots 1 = N!$. For example, there

---

[2] Because the lattice has a fixed number of spaces, i.e., fixed volume, we will calculate a Helmholtz free energy in terms of entropy and internal energy rather than the more familiar Gibbs free energy used at constant pressure and calculated in terms of entropy and enthalpy. The differences are relatively minor and need not concern us here.

are 100! $(= 10^{158})$ ways we could fill up our 10×10 example lattice. Be careful though; this is not the final answer to our problem, as not every one of these configurations is an independent microstate. Look again at the sample mixed configuration shown in Fig. 4.14b; this would be one of the $N!$ configurations and represents a microstate of the system. If we swapped two black (or two white) molecules, we would have another of our $N!$ configurations but not another microstate as it would not be distinguishable from the state before we made the swap. To calculate the proportion of the $N!$ configurations that are distinguishable microstates, we need to subtract out all the black–black and white–white swaps. If there are n_B black molecules on our grid, there are $n_B!$ ways they could be configured without moving any of the white molecules. Similarly, there are $(N-n_B)$ white molecules that could be arranged $(N-n_B)!$ ways without moving a black molecule. Neither of these sets of combinations counts as a separate microstate, so we must move them from the total number of configurations to get the number of mixed microstates. Remembering that in order to exclude combinations, we divide:

$$\Omega_{\text{mixed}} = \frac{N!}{n_B!(N-n_B)!} \quad (4.5)$$

So, in Fig. 4.14b with 50 molecules of each type we have $100!/(50!50!) = 10^{29}$ distinguishable microstates. As the entropy of the unmixed state was zero, the change in entropy on mixing is given by:

$$\Delta S_{\text{mix}} = S_{\text{mixed}} - S_{\text{unmixed}} = k \ln\left[\frac{N!}{n_B!(N-n_B)!}\right] \quad (4.6)$$

In our example the entropy change for mixing is $\Delta S_{\text{mix}} = k \ln(10^{93}) = 66.8k$. Factorials are hard to deal with, but we can remove them from the equation using Sterling's approximation: $x! = x \ln x - x$. Also concentrations are easier to work with than number of molecules so defining the mole fraction of black molecules as $x_B = n_B/N$ and the mole fraction of white molecules as $x_W = (N-n_B)/N$. Substituting these expressions

into Eq. 4.6, we get a more useful expression for the entropy of mixing on our lattice:

$$\Delta S_{mix} = -Nk(x_B \ln x_B + x_W \ln x_W) \quad (4.7)$$

Checking once more with the example in Fig. 4.14, $x_B = x_W = 0.5$ so from Eq. 4.7, $\Delta S_{mix} = 69.3k$—close to what we got from Eq. 4.6. The approximations in Eq. 4.7 become more valid for larger lattices and the convenience of using concentrations rather than molecular numbers makes it a much more useful expression. As mole fractions are by definition less than one, both $\ln x_W$ and $\ln x_B$ are negative and so, as we would expect, the entropy change for mixing is always positive. However, many combinations of ingredients do not spontaneously mix and the resistance must come from the energy of interaction.

Energy of Mixing The internal energy of the system depends on the number and type of bonds present. On our simple lattice, there are only three types of bond: black–black, white–white, and black–white, each with characteristic energies (w_{BB}, w_{WW}, w_{BW}) so the internal energy of the system depends on the numbers of each (m_{BB}, m_{WW}, m_{BW}):

$$U = m_{BB} w_{BB} + m_{WW} w_{WW} + m_{BW} w_{BW} \quad (4.8)$$

Eq. 4.8 would be more useful if we could replace the numbers of each type of bond, a difficult number to estimate, with parameters related to the composition of the system. Each molecule can form a characteristic number of bonds (i.e., the coordination number, z) and these can be either with similar or with dissimilar molecules. Therefore, the total number of bonds formed from the black molecules is twice the total number of black–black bonds plus the total number of black–white bonds (i.e., $zn_B = 2m_{BB} + m_{BW}$). A similar expression can be written for the number of bonds from the white molecules, zn_W. Rearranging and substituting into Eq. 4.8, we can eliminate the w_{BB} and w_{WW} terms and express the total energy of the system in terms of the number of black–white contacts.

$$U = \frac{zw_{BB}}{2} n_B + \frac{zw_{WW}}{2} n_W + \left(w_{BW} - \frac{w_{BB} + w_{BB}}{2} \right) m_{BW} \quad (4.9)$$

where n_B and n_W are the number of each molecule. We still need an estimate of the number of black–white contacts, and we can get that by assuming the two components are mixed randomly, and the chances of finding a black molecule in a space adjacent to a white molecule is given by the fraction of black molecules in the system (i.e., the mean field approximation). While useful from the point of view of getting the calculation to work, this is a significant assumption and one that will be inevitably violated when the two phases that form have a radically different composition from the initial state. Nevertheless, the mean field approximation yields $m_{BW} \approx zNx_B x_W$. Substituting into Eq. 4.9 gives the internal energy of the system but now exclusively in terms of its composition:

$$U = \frac{zw_{BB}}{2} n_B + \frac{zw_{WW}}{2} n_W + \frac{kT}{N} \chi_{BW} x_B x_W \quad (4.10)$$

where χ_{BW} is the exchange parameter:

$$\chi_{BW} = \frac{z}{kT} \left(w_{BW} - \frac{w_{BB} + w_{WW}}{2} \right) \quad (4.11)$$

The change in internal energy due to mixing can be calculated by subtracting from Eq. 4.10 the internal energies of the unmixed phases (the first two terms on the right-hand side of the equation) to give:

$$\Delta U_{mix} = kT \chi_{BW} \frac{n_B n_W}{N} \quad (4.12)$$

The exchange parameter (χ_{BW}) is the variable that relates bond strengths with the energy of mixing. We could imagine the process of mixing in two steps, first we need to pull a molecule out of each pure phase breaking the intermolecular bonds in that phase to do so (Fig. 4.15a). Next, we swap the molecules over and push them into the dissimilar phase and reform the bonds between adjacent molecules (Fig. 4.15b). Doing the energy accounting for this process we must:

Break the black molecule out of its pure phase	$-\frac{1}{2}.z.w_{BB}$
Break the white molecule out of its pure phase	$-\frac{1}{2}.z.w_{WW}$
Put the black molecule into the white phase	$+\frac{1}{2}.z.w_{BW}$
Put the white molecule into the black phase	$+\frac{1}{2}.z.w_{BW}$
Total =	$z.(w_{BW}-\frac{1}{2}w_{WW}-\frac{1}{2}w_{BB})$

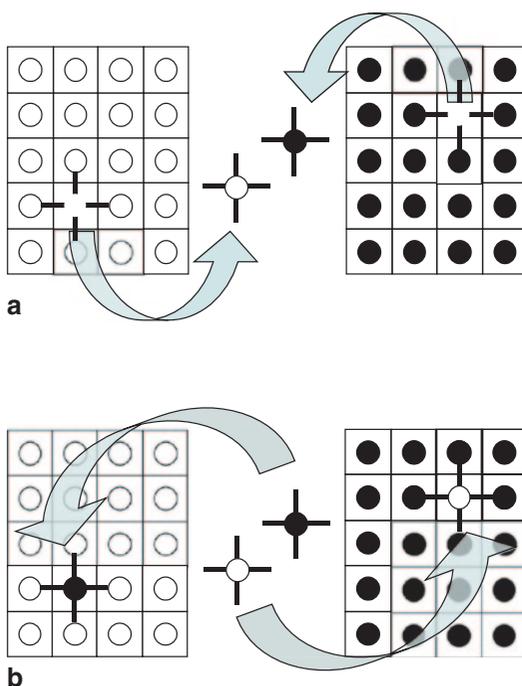(The ½ terms are to avoid double counting bonds; two molecules share one bond)

a

b

Fig. 4.15 Lattice model of mixing a solute (filled circles) with a solvent (open circles). **a** A molecule of each is pulled out of their respective pure phase by breaking like-like bonds. **b** The molecules removed are put back into the other phase making like-unlike bonds

Expressing the total cost of breaking and making bonds in terms of the thermal energy (kT), yields the exchange parameter in Eq. 4.11. If χ is negative, the energy of mixing is negative, and heat is absorbed in the reaction (i.e., endothermic); if χ positive heat is released during the reaction (i.e., exothermic). Importantly, the energy of mixing depends on both the interactions between the solvent and solute but also the interactions within the solute and within the solvent. Very strong bonds

between solvent and solute favor mixing but can be overcome if the bonds within the solute phase or within the solvent phase are stronger. This is important in understanding why salt can dissolve in water. The electrostatic interactions between sodium and chloride ions in the crystal lattice are strong and hard to overcome but the ion–water interactions in the solution are also strong and the two terms almost cancel one another out.

Free Energy of Mixing We are now in a position to calculate the total free energy of mixing; combining the entropy and energy terms in Eqs. 4.7 and 4.12:

$$\Delta F_{mix} = \Delta U_{mix} - T\Delta S_{mix} = kT\chi_{BW}\frac{n_B n_W}{N} \tag{4.13}$$
$$+ kTN(X_B \ln x_B + X_W \ln x_W)$$

We can tidy this expression up by first expressing energy in units of kT, and then doing the calculations per mole of lattice spaces (i.e., $N=N_{av}$, Avagadro's number):

$$\frac{\Delta F_{mix}}{RT} = \chi_{BW}x_B x_W + x_B \ln x_B + x_W \ln x_W \tag{4.14}$$

where the first term on the right-hand side of the equation represents the energetics of mixing, and the second and third terms represent the entropy of mixing. If $\chi_{BW}=0$, i.e., the interactions between like molecules are the same as the interactions between unlike molecules, the free energy of the mixture is given only by the entropy terms.[3] Equation 4.14, Hildebrand's regular solution model, is the ultimate goal of our modeling efforts; from the composition of the system and the chemical interactions expressed through the

---

[3] It may be reassuring to some readers that we could have calculated the free energy in this case directly from Eq. 1.18, $\mu_A = \mu_A^o + RT\ln x_A$. As chemical potential is additive on a molar basis (Eq. 1.13) the free energy of one mole of mixture is $x_B\mu_B + x_W\mu_W$ or
$F_{mixed} = \left[x_B\left(\mu_B^o + RT\ln x_B\right) + x_W\left(\mu_W^o + RT\ln x_W\right)\right]$
The free energy of the same ingredients if they were not mixed is $x_B\mu_B^o + x_W\mu_B^o$ (taking the standard state as the pure substance). Subtracting gives the free energy change of preparing one mole of an ideal mixture from the pure components: $\dfrac{F_{mixed}}{RT} = x_B \ln x_B + x_W \ln x_W$.
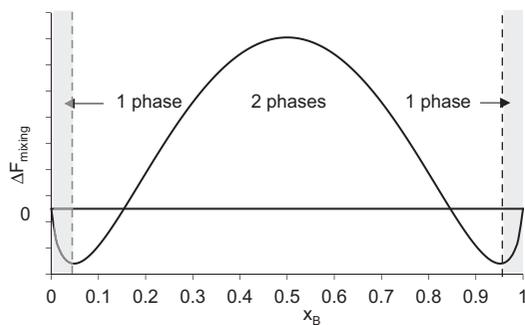
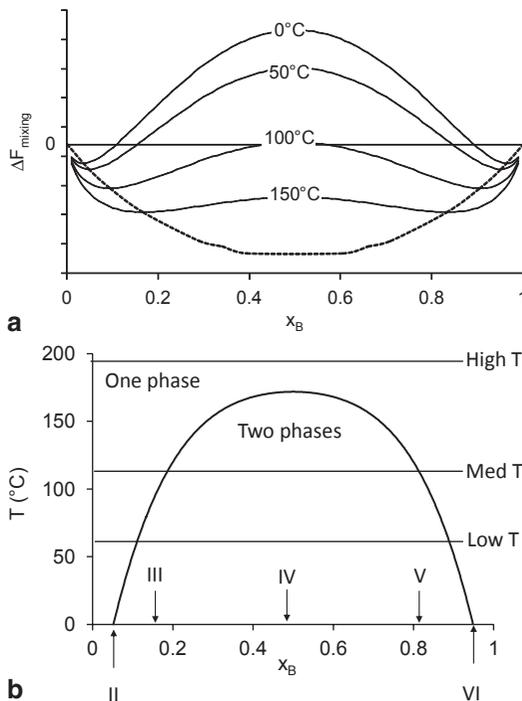**Fig. 4.16** Free energy of mixing as a function of solute mole fraction calculated using a lattice model for $\chi=3$

exchange parameter, we can calculate the free energy of mixing. As a first test of our model, we will look at the effect of the exchange parameter on the free energy of mixing at a fixed temperature (Fig. 4.16).

If a small amount of solute was added to the solvent, the change in free energy on mixing is negative, and the ingredients will mix. However, adding more solute (beyond the free energy minima in Fig. 4.16, i.e., $x_B>0.05$) leads to an increase in free energy, and is thermodynamically opposed. The curve is symmetrical so we could equally start by adding small amounts of solvent to the solute. A small amount of solvent would dissolve in the solute but only up to a critical limit (the second minima in Fig. 4.16, i.e., $x_B>0.95$). Mixtures prepared with compositions between the free energy minima will tend to separate into two phases with compositions given by the energy minima. The behavior shown by the model in Fig. 4.16 bears some qualitative resemblances to the experimental results for butanol–water mixtures at low temperatures (Fig. 4.13b). In the experimental butanol data, the components became more mutually soluble at higher temperature, and we can use the model to make similar predictions from theory (Fig. 4.17).

Temperature comes into our model as part of the exchange parameter. Taking $\chi=3$ at 300 K, we can calculate the phase behavior of our model as a function of temperature (Fig. 4.17a). The position of the minima shift away from the extremes of concentration as the temperature increases, suggesting that at higher tempera-

Fig. 4.17 **a** Free energy of mixing as a function of solute mole fraction and temperature calculated using a lattice model for $\chi=3$. The *broken line* connects the compositions of minimum free energy. Plotting these compositions as a function of temperature yields a phase diagram for this model (**b**). The *temperatures* and *Roman numerals* correspond approximately to the experimental data shown in Fig. 4.13a

tures more solute can dissolve in solvent before becoming saturated. We can calculate the maximum concentration of the one-phase region by connecting the minima[4] on the curves (broken line in Fig. 4.17a). Plotting these concentrations against temperature provides a phase diagram for our model (Fig. 4.17b). The range of concentrations over which the ingredients phase separate gets narrower at higher concentrations, and

---

[4] More formally, we should be looking for two points on the curve with similar slopes. Fig. 4.16 is a plot of free energy against molar composition and the slope of the curve is the chemical potential. By finding compositions with similar slopes, we are finding phases with equal chemical potential that can be at equilibrium with one another. However, this curve is symmetrical and so these points are coincident with the minima and we will not pursue the point further.
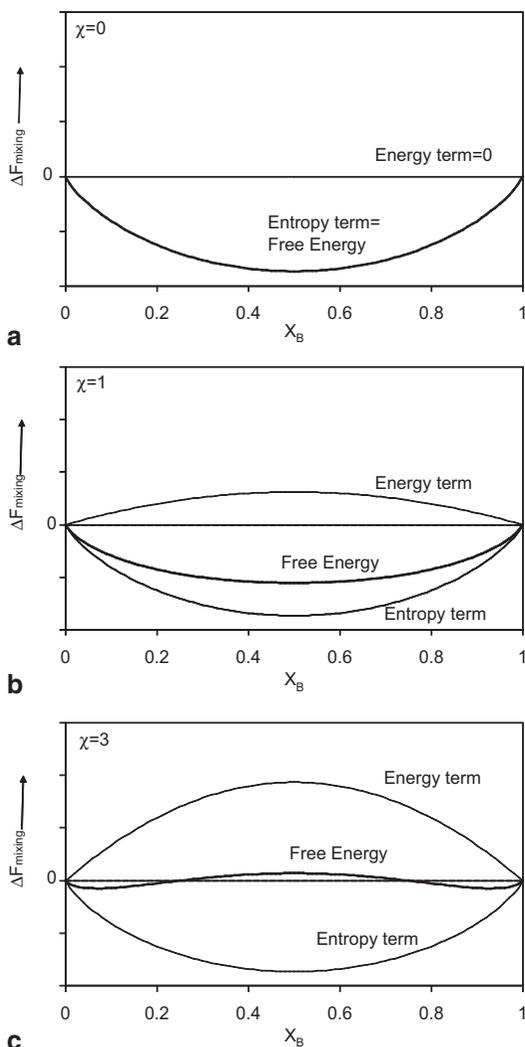
Fig. 4.18 Free energy of mixing as a function of solute mole fraction calculated using a lattice model for **a** $\chi=0$, **b** $\chi=1$, **c** $\chi=3$

dissimilar from one another (i.e., increasing χ). In Fig. 4.18a, the χ term is zero, so the first term of Eq. 4.14, the energetics of mixing, is also zero (i.e., an ideal mixture). The entropy gain associated with mixing means the overall free energy of mixing is negative over the whole concentration range. An example of this would be a mixture of liquid triglyceride molecules, such as is commonly found in commercial cooking oil. Although there are some structural differences between the molecules (number of double bonds, chain length, etc.) the interaction between similar and dissimilar molecules is about the same, and the oil will never spontaneously separate into multiple phases. In Fig. 4.18b, χ is set to a moderately positive value. The bond energies oppose mixing, but not enough to overcome the entropy effects and the ingredients will still mix. The solution will not behave ideally, but can be prepared over the whole range of compositions. An example of this would be a mixture of ethanol and water. The molecular interactions between like molecules and between unlike molecules are different from one another, but not sufficiently different to overcome the entropic drive to mixing. The final case (Fig. 4.18c, $\chi=3$), corresponds to phase separation.

When we launched into this modeling effort, we did so with the proviso that of any model could only be tested by comparison with experimental results—so have we succeeded or failed? The strength of the regular solution model is it provides a general framework to relate the chemistry of ingredients to macroscopic phase behavior. It does not provide precise details of every phase diagram. We enjoyed moderate success with the butanol–water system based on a guessed value of $\chi=3$. The real interactions within a water–butanol system are far more complex than implied by that single parameter, and the mean field approach we used to calculate the energy of mixing could never account for the clustering of water molecules. Yet, despite that, the general form of the prediction was reasonable. We could probably make further improvements to the theory to account for some of these factors and get a better fit, but this would be involved and lose the generality and simplicity of Eq. 4.14.
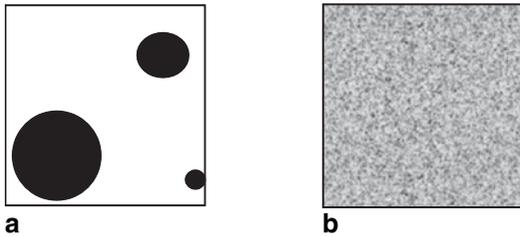
above a critical temperature, the ingredients are miscible over all concentrations. We can qualitatively compare some of the features of the experimental data (Roman numerals in Fig. 4.13a) with calculations from the model (Roman numerals in Fig. 4.17b) and see common features: first mutual solubility up to a certain point then separation into phases of fixed composition, and second increased solubility at higher temperatures.

The final adjustable parameter in the model is the exchange parameter, χ, and we see the effects of making the solute and solvent more

Fig. 4.19 Microstructures typical of phase separation via a binodal (**a**) or spinodal (**b**) mechanism

4.6   Kinetics of Phase Separation

Our consideration of phase transitions to date has been purely thermodynamic, and dealt with the amounts and compositions of phases at equilibrium. However, many foods reach phase equilibrium very slowly. For example, we can make a sweetened iced tea by adding sugar to the hot drink and then cooling it but sugar added to cold tea would just sink to the bottom of the glass and dissolve slowly.

The kinetics of phase separation follows one of two mechanisms:

- Binodal phase separation. The new phase forms from localized points which then grow to form larger spheres or "blobs" as the phase separation proceeds (Fig. 4.19a). There is often a lag time before phase separation.
- Spinodal phase separation. The new phase forms uniformly throughout the material and to form a very intimately mixed microstructure (Fig. 4.19b). There is usually no lag time before phase separation. A good example here is the formation of a cloud when ouzo is diluted with water; the aniseed oil instantaneously phase separates from the diluted ethanolic solution.

We can go some way to understanding reasons for these different mechanisms by looking more carefully at the free energy composition curves (Fig. 4.20). As we learnt in the previous section, if mixtures of composition x_A or x_B were prepared they would both phase separate into different amounts of the two phases with compositions at the free energy minima. Now imagine the process of phase separation. On their way to their final equilibrium compositions, the two new
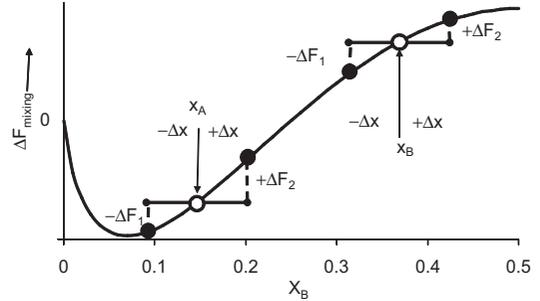
Fig. 4.20 Free energy composition plot calculated from the regular solution model. *Construction lines* show that the partial phase separation of mixture x into two, non-equilibrium, phases $x+\Delta x$ and $x-\Delta x$ leads to a decrease in free energy in one phase and an increase in the other. For composition x_A, the net change in energy is positive ($\Delta F_2 - \Delta F_1 > 0$) so the beginnings of the phase separation is not spontaneous. For composition x_B, the net change in energy is negative ($\Delta F_2 - \Delta F_1 > 0$) so the beginnings of the phase separation is spontaneous. Both x_A and x_B can lose energy by complete phase separation but the process will be binodal in the first case and spinodal in the second

phases must pass through the intervening compositions where one region is somewhat enriched in component x and the other depleted ($x+\Delta x$ and $x-\Delta x$). One of these states has a higher free energy than the starting mixture but the other is lower. In the figure, the mixture with initial composition x_A has gained energy ΔF_1 and lost energy ΔF_2 but as $\Delta F_2 > \Delta F_1$ the net energy change is negative and the reaction can proceed spontaneously. In contrast, the mixture with initial composition $x_B \Delta F_2 < \Delta F_1$ the net energy change is positive and the reaction cannot proceed spontaneously. While complete phase separation is thermodynamically favored, the intermediate state has a higher energy so the reaction is kinetically slowed (see the general discussion around Fig. 2.1 on the effects of a high-energy intermediate on the kinetics of a reaction).

The kinetic barrier to phase separation occurs only when the local free energy-composition curve is concave-down (e.g., composition x_A in Fig. 4.20). If it is convex-up, any partial phase separation will lower the free energy of the system and proceed spontaneously (e.g., composition x_B in Fig. 4.19). The boundary between the conditions for kinetically delayed phase separations and spontaneous phase separations occurs
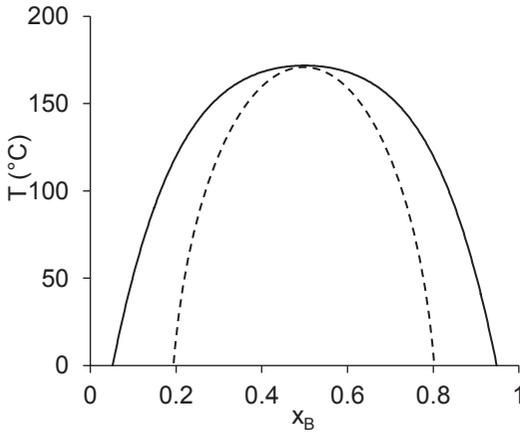
Fig. 4.21 Model phase diagram showing the binodal (*solid line*) and spinodal (*broken line*) phase boundaries

when the free energy curve flattens out (i.e., second derivative of Eq. 4.14 equals zero) and is known as the spinodal curve. Figure 4.21 shows the binodal and spinodal curves for a mixture of ingredients. The solid line (binodal) is the same as that shown in Fig. 4.20b and shows the compositions that will thermodynamically be driven to phase separate. A smaller set of compositions lie inside the broken line (spinodal). The region between the binodal and spinodal lines is thermodynamically driven to phase separate but is kinetically hindered because of the higher energy of the partially separated states. Consequently it may remain a single phase for a considerable period of time. The region inside the spinodal line has no kinetic barrier to phase separation and the phase transitions occurs very quickly throughout the material at diffusion-limited rates giving rise to the finely divided microstructure seen in Fig. 4.19b.

To explain the phase transition in the binodal region, we must assume there are local, transient fluctuations in molecular distribution. The overall composition of the mixture is fixed but there will be regions that are briefly enriched with respect to one component. If the fluctuations are large enough, the nucleus will grow and lead to a macroscopic phase separation. A binodal phase separation starts after waiting for a local composition fluctuation that is large enough to pass the spinodal. Once this nucleus of the new phase is

formed, it will grow by diffusion of other molecules into it or away from it. Under the microscope, the binodally phase separated system will appear as a series of "blobs" of the new phase (Fig. 4.19a). Each blob corresponds to a nucleation event where a local fluctuation initiated the phase separation. The size of each blob is determined by how much additional material diffused into it so the larger particles in Fig. 4.19 presumably nucleated before the smaller ones and had more time to grow. Eventually the blobs will start to impinge upon one another and it is no longer possible to differentiate them. The final phase volumes are given by the equilibrium phase diagram but the sizes of the structures depend on the number of nucleation events and hence are inversely related to the magnitude of the energy barrier slowing phase separations. We will return to this in the specific case of crystallization in Chap. 6, but first we must deal with the consequences of the observations that properties of the surfaces between phases are not the same as the phases themselves.

4.7   Summary

Most foods are mixture of molecules, and most of them are at least partly phase separated. How these phases are subdivided is the basis for much of the microstructure present in food. The limits of solubility are given by a phase diagram which can be used to calculate the composition and amount of the phases at equilibrium. The shape of the phase diagram can, in principle, be calculated from the intermolecular interactions so the study of phase behavior is an essential step in relating molecular properties (Chap. 2) to bulk properties.

Study of phase separation and the properties of phase separated systems is a major theme of the remainder of this book; in the next chapter, we will look at the distinct properties of the interface between phases and then see practical examples of phase separation in the context of crystallization (Chap. 6), polymer systems (Chap. 7), and lastly examine the properties of finely divided phase mixtures (Chap. 8).

4.8   Bibliography

Phase diagrams, particularly multicomponent phase diagrams, are important but often not treated well in introductory physical chemistry texts. "Atkins' Physical Chemistry" (Atkins and De Paula 2006) is very complete (Chaps. 4–6) but builds its arguments around how chemical potential depends on concentration and so may be rather abstract for some readers. A more physically intuitive approach is taken by Dill et al. (2003) who describe and calculate phase diagrams starting from the interactions between molecules (Chaps. 14, 15, and 25). The lattice argument used here is adapted from that work.

Hartel (2001) describes the properties of solutions and phase diagrams in Chap. 4 of "Crystallization in Foods." This book is particularly good on the practical use of phase diagrams in calculating the composition of phases at equilibrium.

Timms (1984) gives a comprehensive and practical overview of phase behavior of food lipids while Wesdorp et al. (2005) offer a more complete thermodynamic rationale. Surfactants have rich phase diagrams important in foods but largely neglected here. Chaps. 1 and 10 of "The Colloidal Domain" by Evans and Wennerström (1994) takes a similar theoretical approach to phase diagrams as used here but with a focus on surfactants.

Surfaces

5.1 Introduction

If you shake oil and vinegar together to make a vinaigrette dressing, the oil will disperse as fine droplets in the vinegar but will not dissolve. Thermodynamically, at least as phase diagrams were introduced in Chap. 4, shaking changes nothing; the total volumes and overall compositions of the oil and vinegar phases are the same. However there must be some missing factor in this analysis as if it were true, then there would be no free energy difference between the dispersed and separated states and we would expect the product to be stable indefinitely. Inevitably though, the intimate mixture of ingredients spontaneously separates back to the initial state. Moreover, although liquids are free to take on any shape, all of the oil droplets in the vinaigrette are spherical (Fig. 4.1). The common feature that differentiates small from large droplets and nonspherical from spherical droplets is spontaneous change always takes place in the direction of the lower area of contact between immiscible phases.

Before shaking, the area of contact between the oil phase and aqueous phase is just the area of the container—a few square centimeters. Assuming shaking led to the formation of 1 mm oil droplets then each droplet would have:

radius, $r = 1$ mm

volume, $V = 4/3 . \pi . r^3 = 4.2 \, \mu l$ per droplet

interfacial area, $A = 4 . \pi . r^2 = 12.5$ mm^2 per droplet

Or in terms of a liter of oil, we would have in total

number of droplets, $n = 1000/V = 238,732$

interfacial area $= A . n = 3$ m^2

Even this modest reduction in the size of the oil phase led to a huge increase in oil–water interfacial area. Further decreases in the size (r) of the droplets would lead to a decrease in the volume ($\sim r^{-3}$) and interfacial area ($\sim r^{-2}$) of each droplet but an increase in the number of droplets ($\sim r^3$) and total interfacial area ($\sim r$) (Fig. 5.1). Spontaneous change is from many, small droplets to fewer, large droplets to minimize the area of contact between phases. Similarly, the droplets are spherical because any other shape would have a larger interfacial area (e.g., cubic droplets would have twice the interfacial area as spherical droplets of the same volume).

We can describe the preference for small interfaces as a free energy term proportional to area of contact between phases:

$$dG = \gamma \cdot dA \qquad (5.1)$$

where γ is the proportionality constant linking changes in area (dA) to changes in surface excess free energy (dG) of the system. The proportionality constant γ is the surface or interfacial tension[1], a characteristic of the two phases in

---

[1] Surface tension is sometimes used to describe interfaces with a gas phase while interfacial tension is sometimes
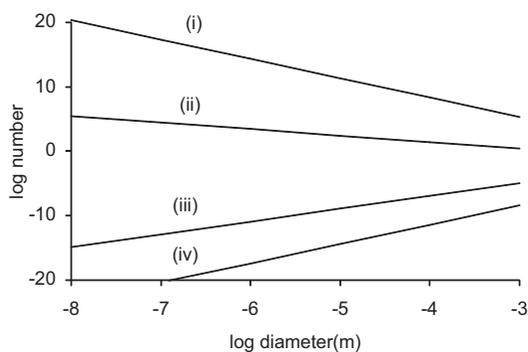
Fig. 5.1 Effect of droplet diameter on the (*i*) total number of droplets or (*ii*) total interfacial area (m^2) per liter of dispersed phase and (*iii*) surface area (m^2) or (*iv*) volume (m^3) per droplet

Table 5.1 Selected interfacial tensions (at 20 °C unless otherwise stated). Surface tension measurements of pure materials is very sensitive to the presence of impurities so reported values can vary considerably

	Surface tension/mNm^{-1}	
	Against water	Against air
Commercial vegetable oil	19.5–23.5[a]	34[b]
Purified vegetable oil	30–31.5[a]	30[c]
Vegetable oil + 1% lecithin	3[d]	
Tetradecane	45[e]	27[f]
Ethanol	–	22[g]
Octanol	8[g]	27[g]
Octane	51[g]	22[g]
Water	–	72.8[g]
NaCl solution (1 M)	–	74.5[h]
Sucrose solution (55%)	–	76.5[i]
Tween 20 (surfactant) solution (16.9 μM, i.e., >CMC)	–	35[j]
Bovine serum albumin (protein) solution (5%)	–	52[j]
Mercury	375[f]	485[f]

[a] Gaonkar (1989)
[b] Esteban et al. (2012)
[c] Chumpitaz et al. (1999)
[d] Johanssen and Bergenståhl(1995)
[e] Inaba and Sato (1996) (at 30 °C)
[f] Jasper and Kring (1955)
[g] Shaw (1992)
[h] Jungwirth and Tobias (2006)
[i] From gPhysics.net
[j] Nino and Patino (1998)

contact (see some examples in Table 5.1). Interfacial free energy is important as it helps connect thermodynamics to the structure of foods. Thus the free energy of the vinaigrette would be about 30 mJ (i.e., 3 m^2 × 10 mJ.m^{-2}) higher than the phase-separated ingredients so the emulsion will spontaneously phase separate to the initial state. The energy to generate the surface came from the shaking (although vastly more energy was wasted as heat) and more intense shaking or mechanical homogenization would yield smaller droplets.

Surface energy becomes increasingly important as the size of phases decreases and the surface area gets large (Fig. 5.1), so we can expect surface properties to be most significant in powders and fine dispersions. However, identifying the surface properties of real foods is a challenge. First, because foods typically contain many phases, there will be many types of interfaces. For example, ice cream has at least four phases (Fig. 4.2, unfrozen solution, fat, ice and air phases), so it will have potentially six types of interfaces (i.e., air–ice, air–unfrozen solution, unfrozen solution–ice, etc.) each with its own interfacial tension. Second, the interfacial tension depends on the chemical properties of the phases in contact at any interface, and this will change over time in important and unexpected ways. For example, the surface free energy driving phase

separation in the vinaigrette will be reduced as any proteins present in the aqueous phase and monoglycerides present in the oil diffuse and adsorb at the interface. However, the challenges of applying principles to real systems is a common one in food science and should not discourage us from proceeding with caution. A good place to start is with the measurement of surface tension and to do that we need to reconceive the energy as a force.

5.2 Surface Tension

Picture a soap film trapped in a loop. Blow a little as if you were about to blow a bubble but just let the film stretch a little then stop blowing

---

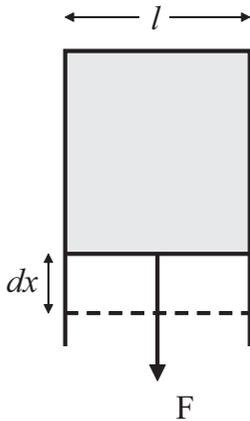used to describe interfaces between condensed phases. We will use the terms interchangeably.

Fig. 5.2 A soap film is trapped in a wire loop and a force *F* is applied along one face to increase the area by *l*.d*x*. The force per unit length on the surface and the change in free energy per unit change in area are both equal to the surface tension
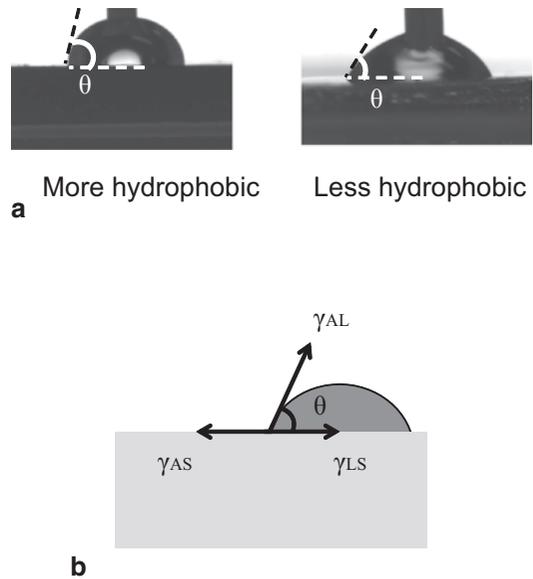

Fig. 5.3 **a** A water droplet spreads on two surfaces of different hydrophobicity. **b** The balance of surface forces resulting in a contact angle. (Source: Images courtesy of Dr. David Beattie, University of South Australia)

and let the film relax back into its original shape. An imaginary line on the surface of the film would be stretched like a piece of elastic and the restoring (elastic) force was the interfacial tension. We could equivalently have said that blowing increased the air–film interfacial area and the tendency to recoil was to reduce this energy term. Both views are the same but the new, force-based, statement provides a mechanical picture of how surfaces behave while the former energy-based approach fits more obviously into thermodynamics. It is easier to show that both approaches are the same using a simpler bubble.

Imagine a soap bubble caught in a rectangular wire loop with one freely sliding edge of length *l* (Fig. 5.2). Apply a force (*F*) at right angles to the sliding edge to stretch the film (d*x*). The amount of work done (*w*) is force times distance, or in terms of infinitesimal changes: d*w* = *F*.d*x*. The area created by the work is 2 *l*.d*x* (the two because the film has two sides) so from Eq. 5.1 the change in free energy is *γ*.2*l*.d*x*. The work done on the system is equivalent to the change in free energy, so we can set the two terms equivalent to one another. Then, *F*.dx = 2*l*.*γ*.dx or:

$$\gamma = F / 2 \, l. \tag{5.2}$$

The surface tension defined energetically in Eq. 5.1 is equivalent to the force acting at right angles to line of unit length drawn on the surface of the film. (Once more, the two is because the soap bubble has two surfaces). Surface tension is usually expressed in units of energy per unit area or force per unit length but dimensional analysis quickly shows these units are the same. The surface force is measurable, but the planar film with a frictionless wire is experimentally impractical and we must look for its consequences elsewhere. One approach is to look at the contact not between two phases but between three.

Water forms beads on hydrophobic surfaces (e.g., many plastics) and spreads out as a film on more hydrophilic ones (e.g., clean glass) (Fig. 5.3a). The droplet shape is readily measured as the contact angle at the surface of the droplet, θ. Qualitatively, a low contact angle corresponds to a spread-out droplet or a solid surface that would "prefer" to be in contact with the liquid than with the air. We can quantitatively relate wetting angle to interfacial tension by looking at the forces acting on the point of contact between the three phases (Fig. 5.3b):

- The air–solid interfacial tension (γ_{AS}) pulls the contact point to the left to reduce the air–solid interfacial area.
- The liquid–solid interfacial tension (γ_{LS}) pulls the contact point to the right to reduce the liquid–solid interfacial area.
- The air–liquid interfacial tension (γ_{AL}) pulls the line at angle θ to the surface and reduces the air–liquid surface area. The horizontal component of this force acts to pull the contact point to the right while the vertical component pulls the solid surface upwards. The vertical force is matched by the elastic properties of the solid and causes no significant movement.

At equilibrium, these three forces are in balance so we can set the horizontal components equal to one another as Young's equation:

$$\gamma_{SL} + \gamma_{AL} \cos\theta = \gamma_{AS} \qquad (5.3)$$

The right-hand side of the equation is the pull from right to left to minimize air–solid contact, which is equal to the sum of the pull from left to right to minimize solid–liquid contact and the horizontal component of the pull to minimize air–liquid contact. Measurements of contact angle are a function of all three interfacial tensions acting and without additional information we cannot calculate any one of them. However, contact angle measurements are among the only ways available to measure the energetics of solid surfaces typically by reporting changes in wetting behavior in response to changes in either solid or liquid phase composition.

Example: Contact Angle on Zein Edible Films
Edible films are formed from food biopolymers as packaging material or diffusion barriers within the food sttructure. Ghanbarzadeh et al. (2007) formed edible films from zein, the highly hydrophobic and insoluble storage protein in corn, with glucose added as a plasticizer to maintain flexibility. They hoped that the hydrophobic protein would make a good barrier to water diffusion and used contact angle measurements to get a measurement of the film–
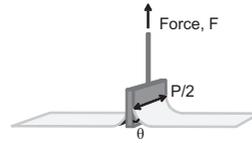
Fig. 5.4 A Wilhelmy plate at point of detachment from a liquid surface. *P* is the perimeter of the plate. (Source: Image courtesy Dr. Claire Berton-Carabin, Wageningen University)

water interface. The unplasticized zein film had a water contact angle of 62.3° and the film plasticized with glucose had a contact angle of 52.3°. Adding more hydrophilic sugar lowered the contact angle so the film was more readily wet with water. However, it is worth considering whether this was truly an equilibrium measurement because the glucose in the film would dissolve into the water and at least to some extent the water would soak into the film matrix changing the properties of the phases and the tension between them.

Most liquids will rise up to form a meniscus at the container walls according to the balance of forces described in Young's equation. A meniscus causes a bulk deformation in a fluid due to surface forces and provides a route to measure interfacial tension. For example, a common type of surface tensiometer works by carefully measuring the weight of a small plate of an inert metal hanging from a fine wire (i.e., a Wilhelmy plate, Fig. 5.4). The plate is brought into contact with the liquid surface of interest. A meniscus forms around the perimeter of the plate, and the weight of the liquid lifted is measured as an increase in the weight of the plate. According to Young's equation, the force needed to do this is $\gamma_{SL}\cos\theta$ per unit length of contact. If the perimeter of the plate in contact with the liquid is P then the measured force is:

$$F = P\,\gamma_{SL}\,\cos\theta \qquad (5.4)$$

So now, if the equilibrium contact angle is known, then a force measurement can be used to calcu-
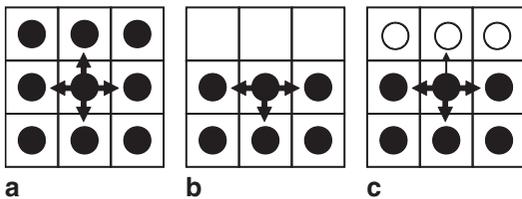
Fig. 5.5 Lattice models of bonding in a molecule in **a** bulk liquid, **b** liquid–gas interface, and **c** liquid–liquid interface. *Arrows* represent intermolecular attraction; the thickness of the *arrow* is an indication of the strength of the interaction

late surface tension. Very often the plate material is chosen so that the wetting is effectively perfect (e.g., platinum), and the contact angle is taken as zero (i.e., $\cos\theta = 1$ and the force per unit length of the perimeter is the interfacial tension).

Having defined and measured surface tension, the next challenge is to relate it to the molecular properties of the materials in contact. We used lattice models in Chap. 4 to provide a thermodynamic explanation of phase behavior in terms of the interactions between molecules and we will use a similar approach here to understand surface properties.

5.3   Molecular Basis of Surface Tension

A molecule in the bulk (i.e., away from the interface, Fig. 5.5a) has equal intermolecular bonds on all sides and no net attraction in any direction. Figure 5.5b shows a similar molecule now at the interface with a gas phase (because the density of the gas is so low, the lattice sites are left empty). The surface molecule can form one fewer intermolecular bond, so it has a higher energy and the molecule experiences an attractive force away from the surface. Increasing the surface area means moving more molecules against this force from the bulk where they have strong bonds (low energy) to the surface where they have weaker bonds (higher energy). The imbalance of interactions at the surface is the molecular basis for interfacial tension.

We can go further and use the lattice to calculate surface properties in a similar way as we did for the bulk phase partitioning in Chap. 4. Imag-

ine a pure condensed phase with N molecules of which n are at the surface. The $(N−n)$ molecules in the bulk have a coordination number of z and the n molecules at the surface have one less bond and a coordination number of $z−1$. If the energy of each bond is w_{BB} then the total internal (bonding) energy of the system (U) is the sum of the bond energies.

$$U = \frac{zw_{BB}}{2} \cdot (N-n) + \frac{(z-1)w_{BB}}{2} \cdot n = F \quad (5.5)$$

The first term is the energy due to bonding in the bulk of the liquid and the second is bonding at the surface. The ½ factors are to avoid double counting bonds. As the liquid is a pure phase, there is no mixing entropy to worry about so $S=0$ and Eq. 5.5 gives the free energy for the system (i.e., $F=U$).

To calculate surface tension we need to know the rate of change of free energy with surface area, Eq. 5.1. As the surface area (A) increases, the number of molecules at the surface (n) will also increase so we can first express the differential as a product then attack the parts:

$$\gamma = \frac{dF}{dA} = \frac{dF}{dn} \cdot \frac{dn}{dA} \quad (5.6)$$

The first part of the product is the differential of Eq. 5.5 with respect to n ($=-\frac{1}{2}w_{BB}$), and the second can be calculated by expressing the area in terms of the number of molecules at the surface and the area taken up by each molecule at the surface (a), i.e., $A=na$ so $dn/dA=a^{-1}$. Substituting into Eq. 5.6 gives an expression for surface tension in terms of intermolecular bond strength and molecular area:

$$\gamma = \frac{-w_{BB}}{2a} \quad (5.7)$$

Surface tension is related exclusively to intermolecular forces in the liquid phase and the size of the molecules. Real liquids are unlikely to be pure phases, so the concentration of one phase at the interface of a solution would tend to decrease entropy as would the alignment of asymmetric molecules at the surface. However, in most cases
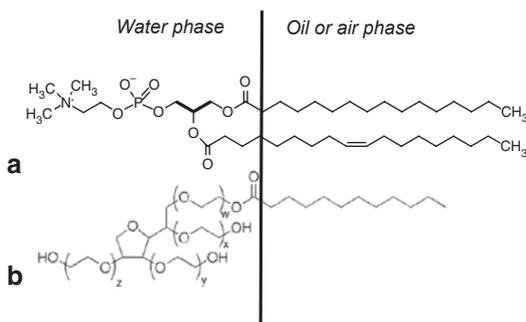
Fig. 5.6 Molecular structure of **a** example phosphatidylcholine (an important phospholipid in lecithin, different fatty acid residues may be present as the lipophilic portion of the molecule) and **b** *polyoxyethylene (20) sorbitan monolaurate (Tween 20, x+y+z=20)* at an interface. (From Wikipedia.org)

Eq. 5.7 is reasonable in treating surface tension is largely as an internal bonding effect in the liquid. Surface tension increases with the strengths of the internal bonds holding the liquid molecules together (w_{BB}), so mercury with strong metallic bonds has higher surface tension than hydrogen-bonded water which in turn is higher than octane with just has weak intermolecular Van der Waals forces (Table 5.1).

A lattice model for liquid–liquid interfacial tension is shown in Fig. 5.5c paralleling the model for surface tension at the gas–liquid surface in Fig. 5.5b. The important difference is the weaker like–unlike bond at the surface partly compensates for the loss of a like–like bond. We can use the exchange parameter χ (see Eq. 4.10) in place of w_{BB} in Eq. 5.7 to get a measure of the interfacial tension between two condensed phases. A high value of χ corresponds to poor mixing between the phases and also to a high interfacial tension (e.g., octanol is more soluble in water than octane and the interfacial tension is lower, Table 5.1). At lower values of χ, the phases will become increasingly mutually miscible, for example, the water–butanol phases at higher temperatures in Fig. 4.12. In that case, the appropriate interfacial tension is not between pure phases but between saturated solutions of each in the other and can be very low. If two liquids are completely miscible, then there is no interface and no surface tension.

Some solutes cause only small changes to surface tension at relatively high concentrations (e.g., salt, sugar, Table 5.1). On the other hand, only very small amounts of proteins or small molecule surfactants are needed to cause large changes in surface tension. For example, food grade vegetable oils typically contain a percentage of free fatty acids, mono- and di-glycerides, and removing them causes a large change in surface tension. The reason for this is certain solute molecules can have a preference for the interface over either of the bulk phases and will accumulate there to high concentrations and shield the incompatible phases from one another to some extent.

5.4 Emulsifiers

Given the opportunity to move between two phases, solutes accumulate at higher concentrations in the phase where their molecular interactions with the solvent are better. For example, in an oil-in-water mixture, Vitamin C would accumulate in the aqueous phase and Vitamin E in the oil phase. We would describe Vitamin C and E as hydrophilic and hydrophobic respectively because of their different noncovalent interactions with water. Other molecules have a part of their structure that is hydrophilic and a part that is hydrophobic; they are amphiphiles. For example, phosphatidylcholine is a phospholipid and an important component of lecithin. It has a hydrophilic, ionic phosphate group and two hydrophobic, hydrocarbon fatty acids coupled to a glycerol backbone (Fig. 5.6a, glycerol backbone shown bold). Phosphatidylcholine is very poorly soluble in both oil and in water because moving into either pure phase is "good" for part of its structure and "bad" for the other part. Instead it will tend to accumulate at the interface between the phases.

Phosphatidylcholine is one example of the many small molecule surfactants naturally occurring in or added to foods. Other examples include the neutrally charged polysorbates e.g., polyoxyethlyene (20) sorbitan monolaurate, (Tween 20, Fig. 5.6b), fatty acids, and negatively charged diacetyl esters of tartaric acid (DATEM). All of these are small molecules (MW<~1000)
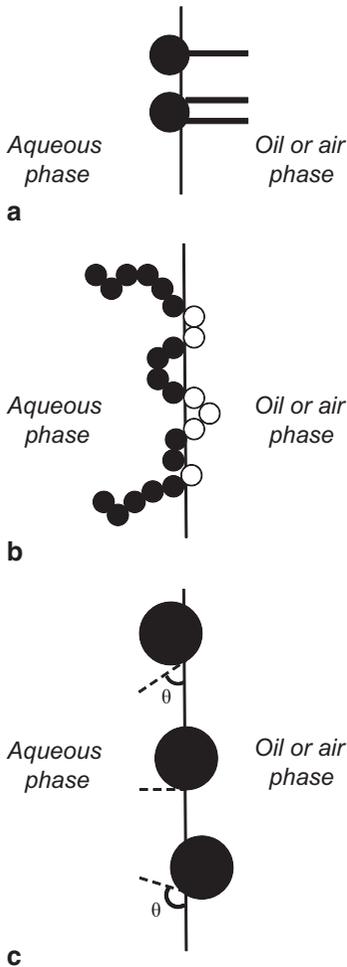
Fig. 5.7 Schematic representation (not to scale) of different emulsifiers adsorbed at a surface. **a** Surfactants. The hydrophobic head group is represented by a *circle* and the hydrophobic tail(s) with *lines*. **b** Protein. The *filled* (*open*) *circles* are representations of hydrophobic (hydrophilic) amino acid residues. **c** Pickering particles. From *top* to *bottom*—hydrophilic, neutral, and hydrophobic particles with increasing contact angle (θ)

the structure is expressed as a hydrophile–lipophile balance (HLB) number on a 0–20 scale corresponding to completely hydrophobic to completely hydrophilic. All that holds the surfactant to the interface is the energy of transferring the hydrophobic group into the aqueous phase (~a few kT) so surfactants can rapidly detach and re-adsorb at a surface.

Polymers, importantly proteins, can also be surface active. Some important examples include the whey proteins and caseins from milk as well as soy proteins and gelatin. On the other hand, relatively few food polysaccharides have sufficient hydrophobic groups to be usefully surface active, but some exceptions include gum Arabic and certain types of pectin. Proteins are described in more detail in Chap. 7, but here it will suffice to treat them as hundreds or thousands of small molecules (amino acid residues) linked in a chain. Each amino acid residue has a molecular weight of approximately 100–200 Da so the typical protein is several orders of magnitude larger than the typical surfactant. Different amino acid residues have different degrees of hydrophobicity. In water, many proteins form dense, highly ordered coils with the hydrophobic amino acid residues in the dry core of the molecule. At a surface, the protein can unfold to some extent and find a new conformation with the hydrophobic amino acid residues removed from the aqueous environment but anchored to the aqueous phase by its hydrophilic amino acids (Fig. 5.7b). Because a protein is attached through multiple anchor points to a surface, it tends not to spontaneously desorb.

While less important in foods than surfactants or proteins, fine particles can also be surface active. For example, fat crystals and starch granules are believed to play a role as Pickering stabilizers for some food emulsions (Dickinson 2012). Particles will tend to adsorb at surfaces to achieve optimum wetting by each phase (Fig. 5.7c). Once adsorbed, the energy for detachment is very high (~hundreds of kT) and the particle is effectively permanently attached to the surface.
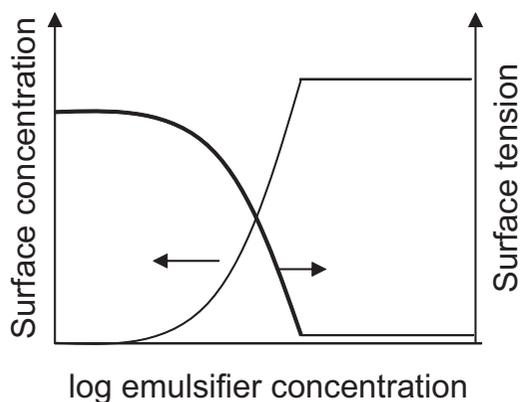
with a water-soluble head group (i.e., a charged or polar group) and water insoluble tail (i.e., a hydrocarbon chain). For example, models of two-tailed phosphatidyl choline and single-tailed Tween are shown adsorbed to an interface using this model in Fig. 5.7a. The relative importance of the hydrophobic and hydrophilic portions of

Fig. 5.8 The changes in surface concentration and surface tension as a function of added emulsifier concentration

5.5   Sorption

Together small molecule surfactants, proteins, and Pickering particles are described as emulsifiers. For an emulsifier, the interface provides a third "phase" preferred over either of the bulk phases[2]. The surface concentration, expressed as amount per unit area, can be calculated by subtracting the amount dissolved in each bulk phase from the total present and normalizing to the surface area. Surface concentration increases with added emulsifier concentration (Fig. 5.8) from zero to a plateau value (i.e., the monolayer concentration, typically a few milligrams per square meter).

A molecular basis for the sorption isotherm is shown in Fig. 5.9 for a small molecule surfactant partitioning between an aqueous phase and the interface with a gas phase. Figure 5.9a shows a cross-section across the interface while Fig. 5.9b shows the same process looking at the surface from the aqueous phase. The arrow in part a of the figure shows the position of the observer in part b, and the numbers (i)–(iv) indicate increasing emulsifier concentration. The more emulsifier added to the aqueous phase, the more that will adsorb at the surface (i)–(ii), until the surface is full or no more will dissolve in the aque-
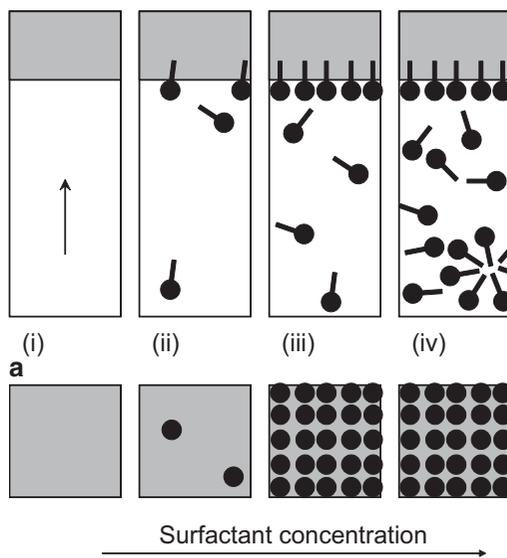
Fig. 5.9 The effect of surfactant bulk concentration on surface concentration at an air–water interface. **a** Cross-sectional view and **b** view from the aqueous phase towards the air phase. The direction of observation in **b** is shown as an *arrow* in the first *panel* of **a**

ous phase (iii). Any more added emulsifier accumulates in the aqueous phase (iv), and does not change the properties of the surface. Different molecules have different characteristic affinities for the interface but typically, proteins adsorb at lower concentrations than small molecule surfactants. An important consequence of this is when an emulsion (e.g., cream) is diluted in water; although the protein concentration in the aqueous phase is reduced, the proteins at the surface do not desorb and the emulsion remains stable.

Adsorption of an emulsifier to a surface lowers the surface tension (Fig. 5.8). The adsorbed surfactant decreases the area of unfavorable contacts between water and air (Fig. 5.9b). Beyond the monolayer concentration, no further material can adsorb so there are no further changes in interfacial tension. Different molecules have different capacities to lower the interfacial tension but typically, small molecule surfactants lower the surface tension more than proteins. Consequently, small molecule surfactants will tend to dominate at a surface in a mixed protein/surfactant system and may even displace pre-adsorbed protein from an interface.

---

[2] How the surface can really be treated as a phase according to the definitions provided in Chap. 4 is the subject of the Appendix to this chapter: The Gibbs Surface.

The changes in surface tension (γ) and surface concentration (Γ) with the total concentration of emulsifier added (c) are related through the Gibbs sorption isotherm. The full form of the Gibbs isotherm is rigorously derived and generally correct, while the following widely used formulation is only valid in dilute systems where concentration can be used in place of activity:

$$\Gamma = \frac{-1}{RT}\left(\frac{d\gamma}{d\ln c}\right) \qquad (5.8)$$

The Gibbs sorption isotherm is particularly useful in calculating surface load or tension when one parameter is known but the other cannot be measured. For example, the surface tension at a solid interface cannot be measured but it can be calculated if surface load is calculated as a function of added emulsifier concentration.

Example: Adsorption of Tween 20 at an Air–Water Interface
Niño and Patino (1998) measured the air–water interfacial tension for a variety of solutions of polyoxyethylene sorbitan monolaurate (Tween 20). Surface tension decreased linearly with log concentration then reached a constant value of about 35 mN/m (Fig. 5.10). The break point between the two regimens is the CMC ($=16.9$ μM at 20 °C) and represents the maximum concentration of monomeric Tween 20. Numerical differentiation of the decreasing portion of the sorption isotherm
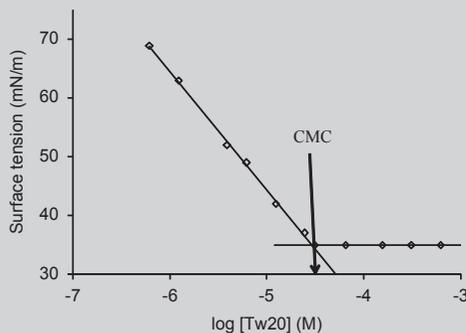
Fig. 5.10 Sorption isotherm of Tween 20 at an air–water interface. (Adapted from Niño and Patino 1998)

gave $d\gamma/d(\log[\text{Tween 20}])$ which reached a maximum at the CMC and was substituted into Eq. 5.8 to give the surface excess concentration ($=3.56$ μmoles/m^2, 4.3 mg/m^2). Alternatively if there are 3.56 μmoles of Tween 20 per square meter of surface, then each molecule at the surface occupies 0.466 nm^2.

This is a thermodynamic treatment of sorption but the kinetics can be equally important. When a new surface is generated, for example, when a large oil droplet is broken up in a homogenizer to form an emulsion, it is "bare" until emulsifiers can adsorb at the interface. In general, the rate of accumulation at the surface increases with emulsifier concentration and decreases with molecular weight (i.e., small molecule surfactant >proteins >Pickering particles). Very often in foods, a mixture of proteins and surfactants are present and the balance found at the surface depends on which get there first. In some cases, an existing interfacial layer can be displaced by another emulsifier added later (i.e., competitive adsorption). Small molecule surfactants are added to ice cream mix to partly displace dairy proteins from the oil–water interface and facilitate some partial coalescence (see Chap. 9).

Example: Lecithin at Fat Crystal Surfaces
Lecithin is widely used to modify the texture of fatty foods and is believed to adsorb at the surface of fat and sugar crystals. Johansson and Bergenståhl (1995) investigated the surface properties of lecithin through measurements of the contact angle between water, oil, and crystalline fat as a function of time for different concentrations of lecithin in the lipid phase (Fig. 5.11). Note that because oil floats on water the instrument to measure contact angle (inset in Fig. 5.11) was inverted compared to the one shown in Fig. 5.3. The contact angle increased over time as the lecithin diffused to the surface and adsorbed to the interface. An increase in the contact angle meant the
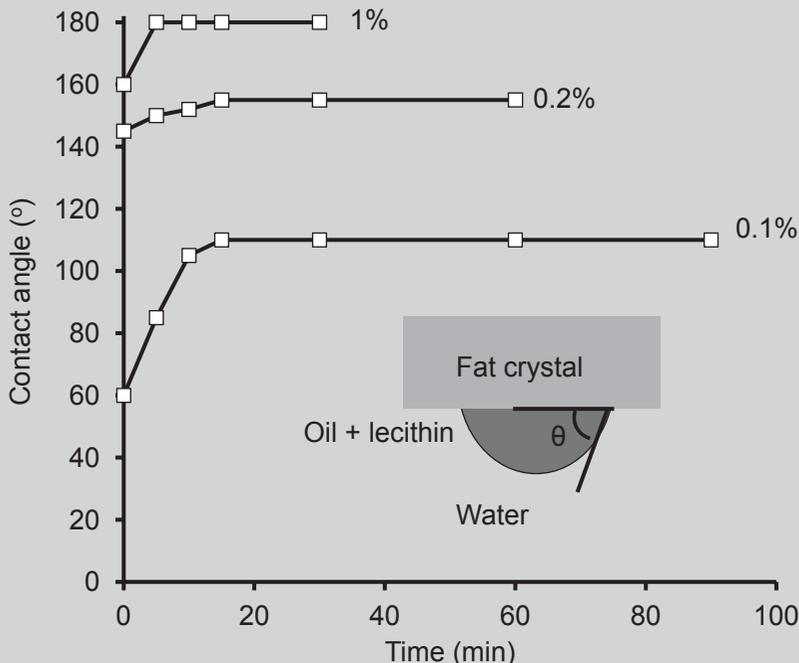
Fig. 5.11 Contact angle in an oil droplet at the air–fat crystal interface for three different concentrations of phosphatidyl choline in the oil. *Inset:* Diagram showing the measurement. (Adapted from Johansson and Bergenståhl 1995)

fat crystal surface became more polar so water–fat contacts were preferred over oil–fat contacts. The equilibrium contact angle was larger for higher concentrations of lecithin.
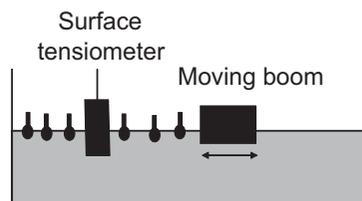

Fig. 5.12 The Langmuir trough. A moving boom varies the area available to the emulsifier (shown schematically) while a Wilhelmy plate measures the resultant interfacial tension

5.6   Properties of Surface Layers

The physical and chemical properties of the surface film are essential to the overall properties of foods with large interfacial areas, i.e., fine dispersions. However, the most practical way to assess film structure is to manufacture a large (~square centimeters) flat interface that has the same properties as the small, curved surface of the fine particles of interest but be much easier to measure. Characterization of a planar surface usually involves use of a Langmuir trough; a container of liquid with a defined surface whose area can be changed by dragging a boom across it (Fig. 5.12). A typical experiment involves fill-

ing the trough with water then injecting a known amount of emulsifier and allowing it to accumulate at the air–water interface (if necessary, oil can be layered on top of the water to study an oil–water interface). The surface area is then reduced by moving the boom, and any material at the surface is compressed into a smaller area and surface tension is measured at each step. Typically surface tension decreases as the fixed amount of material at the surface is confined to progressively smaller areas. This is typically plotted as an increasing surface pressure, π ($=\gamma-\gamma_0$, where γ_0 is the interfacial tension in the absence of
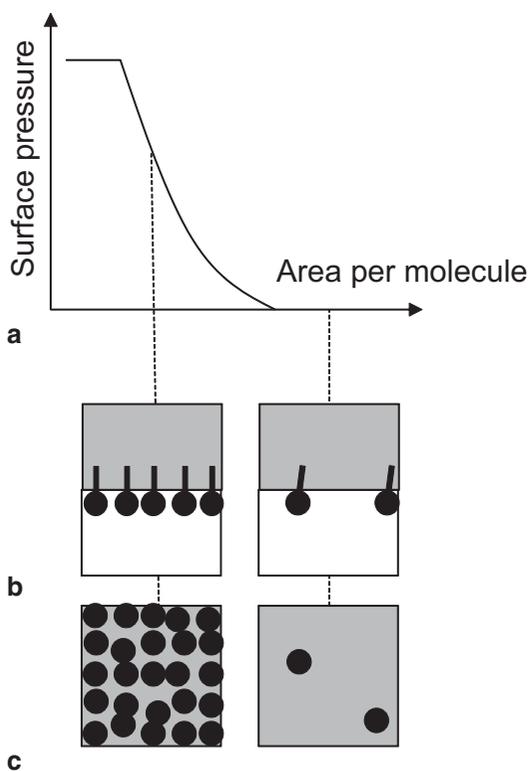
Fig. 5.13 a Typical π-A isotherm for a small molecule surfactant and diagrams of the structures of the interfacial film of small molecule surfactant **b** in cross-section and **c** viewed from the aqueous phase

emulsifier). Surface pressure is the reduction in surface tension of the "bare" interface due to the presence of the emulsifier. We will interpret the surface pressure vs. area relationships in terms of the interactions of emulsifier molecules on a two-dimensional surface; this is similar to interpreting the pressure vs. volume relations in terms of the interactions between molecules in bulk solids, liquids, and gasses in three-dimensional space.

Typical results are shown in Fig. 5.13a along with schematic illustrations of the structure of a small molecule surfactant at the surface viewed in cross section (Fig. 5.13b) and at the surface from the aqueous phase (Fig. 5.13c). Initially the surfactant is well dispersed on the surface but as it is compressed, it collapses into more condensed structures. Figure 5.13c looks remarkably like the cartoon representations of three-dimensional solids, liquids, and gasses presented in Fig. 4.4a and in many respects, the surface film

is just a two-dimensional reflection of the same phenomena where area takes the place of volume and surface pressure takes the place of pressure. The uncompressed film is a two-dimensional gas that suddenly collapses at a critical volume to a much smaller area per unit molecule and forms a two-dimensional liquid phase then eventually a two-dimensional solid phase.

The shape of a π-A isotherm can be used to calculate the properties and interactions of emulsifiers at the surface just as a pressure–volume curve can be used to investigate the properties of bulk phases. For example, the space taken up by each molecule at the surface (useful in the Gibbs isotherm, Eq. 5.8) is calculated by extrapolating a line through the steep portion of the curve to the x-axis and the elasticity of the surface (specifically the dilational modulus, the force required to stretch the surface by unit amount) is calculated as the slope of the π-A curve.

Example: Properties of β-casein at a Surface
β-casein is an important dairy protein used to stabilize emulsions and foam. Rodríguez Patino, Sanchez and Niño (1999) dissolved the protein in buffer at pH 5 and 7 then injected a small volume into buffer in a Langmuir trough and compressed it (3.3 cm.min^{-1}) to get a pressure–area isotherm (Fig. 5.14). At large areas there was very little surface pressure; the interfacial protein was acting as a two-dimensional gas. As the area available for the proteins decreased, the proteins began to interact with one another and form a two-dimensional liquid monolayer. The increase in surface area was seen at pH 7 before pH 5, suggesting the caseinate takes up a larger area at that pH, and so the molecules interact with one another sooner. This is presumably because at pH 7, the protein molecules have a negative charge so they repel one another at the surface while pH 5 is close to the isoelectric point. Further decreases in area require more and more pressure and the slope of the line is the surface dilational modulus. The authors identify two differ-
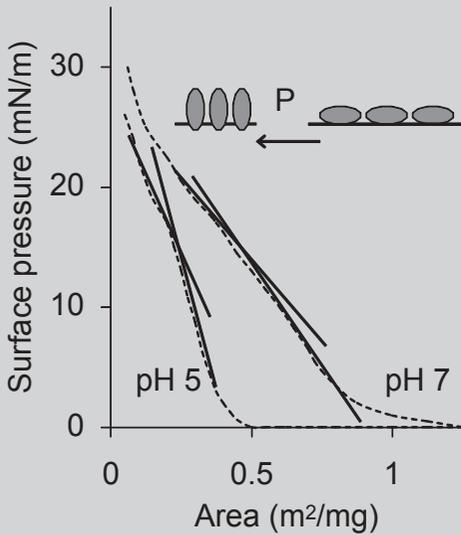
Fig. 5.14 Surface pressure isotherm for β-casein at pH 7 and 5. *Lines* show the elasticity of the different regimes in the monolayer and the *inset* is a schematic diagram showing a phase transition between these regimes as the proteins move from one configuration to another. (Adapted from Rodríguez Patino et al. 1999)

ent forms of the surface layer corresponding to different slopes to the isotherm with a transition between them. The transition probably corresponds to a change in conformation of the protein molecules at the surface as they are pushed against one another. Finally, the change in slope at the smallest interfacial areas corresponds to the collapse of the interfacial film.

A variety of complementary techniques are available to characterize the planar film present on a Langmuir trough at different surface pressures. A classic method is to transfer the film to a smooth solid surface then image it using electron or atomic force microscopy but other methods (e.g., ellipsometry, Brewster angle microscopy) can be used to image the film directly on the fluid–fluid interface. It is also possible to measure the rheology of a surface film by lowering a thin disk on a wire into the plane of the surface, and measuring how it rotates when the surface itself is rotated. This is a direct two-dimensional

equivalent of the three-dimensional rheological methods for the characterization of liquids and solids that will be described in Chaps. 7 and 9 and similar properties (e.g., surface viscosity and surface shear modulus) can be measured. Using these approaches and others, researchers have found many of the behaviors of three-dimensional liquids occurring in two dimensions at surfaces (e.g., phase separation, gelation, and chemical reactions). The most significant practical challenge in all of these experiments is maintaining acceptable levels of cleanliness. The surface to volume ratio in the planar surface is much smaller than in the real food being modeled. Minute quantities of surface-active materials that would be diluted to the point of irrelevance across the large surface area of a real emulsion in the real case will accumulate at the small planar surface area in the experiment and hopelessly skew the results.

While planar surfaces are widely used to study the properties of emulsifiers, most of the important surface properties occur on fine and thus highly curved surfaces. Curvature has consequences for surface properties that cannot be captured using a Langmuir trough.

5.7 Curved Surfaces

Surface tension is an elastic pull at right angles to a line drawn in the surface. When a surface is curved, the elasticity serves to compress the material inside the curvature and raise the internal pressure relative to the external pressure. For example, the equilibrium size of a bubble is reached when the contracting force of the surface tension is matched by the expanding force of the pressurized gas inside (Fig. 5.15). The pressure difference (ΔP) across a curved surface is the Laplace pressure which, for a spherical[3] surface, is given by:

---

[3] More generally the curvature at any point on any surface can be defined by the radii of two circles (R_1 and R_2) drawn at right angles to one another and at a tangent to that point. The internal pressure at that point can then be calculated by replacing $2/r$ in Eq. 5.8 by ($1/R_1 + 1/R_2$). For a spherical surface $r = R_1 = R_2$ so $1/R_1 + 1/R_2 = 2/r$. A simi-
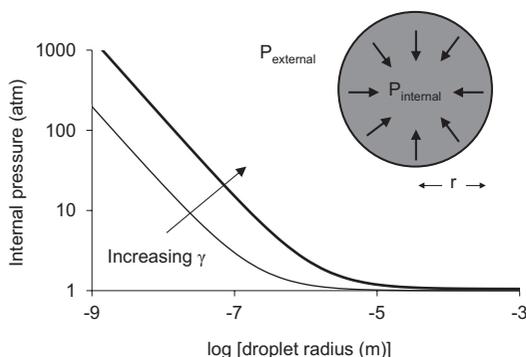
Fig. 5.15 Effect of droplet radius and interfacial tension on the internal pressure of a droplet or bubble. The *bold line* shows typical values for a water droplet in air and the *fine line* a food oil droplet in water. The *inset* diagram shows the interfacial basis of the internal pressure

$$\Delta P = \frac{2\gamma}{r} \qquad (5.9)$$

The internal pressure increases with surface tension (γ, i.e., the tendency of molecules to move away from the surface) and decreasing particle radius (r, i.e., smaller particles have more surface to pressurize a volume) (Fig. 5.15). Small bubbles tend to behave as hard spheres and resist deformation under most circumstances while larger ones are more easily stretched. For example, the small bubbles in a fizzy drink remain spherical as they rise to the surface while the large gas cells in bread dough are deformed by the kneading and baking process. One important feature of Fig. 5.15 is how sharply the pressure increases for very small particles and so its consequence are most acute at the micro- to nanoscale.

We can understand the thermodynamic effects of surface curvature starting from a molecular picture. Imagine a molecule at a flat gas–liquid interface held into the liquid phase by intermolecular bonds with its neighbors (Fig. 5.16a, i.e., the same model we used to understand the molecular basis of interfacial tension in Fig. 5.5). When the same liquid is present in a small droplet with a
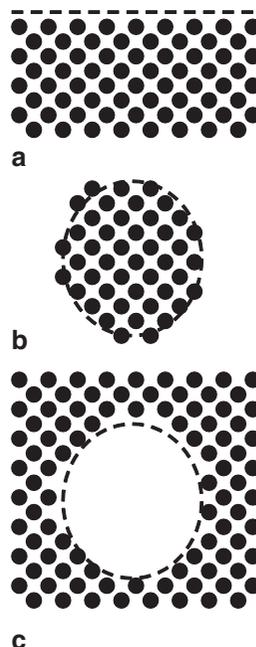
Fig. 5.16 Highly schematic diagram for molecules **a** at a planar surface, **b** in a fine droplet, and **c** surrounding a bubble. The curvature of the surface is vastly exaggerated relative to the scale of molecules to show the change in coordination number at the surface more clearly. The position of the surface is shown as a *dashed line*

highly curved (convex) surface, the coordination number at the surface is reduced so it is easier for a molecule to escape into the gas (Fig. 5.16b). When the curved surface is due to a bubble in the liquid (i.e., a concave), the coordination number at the surface is increased and it is harder for a molecule to escape into the gas (Fig. 5.16c).

The tendency of a molecule to escape is measured as a partial pressure (i.e., the fraction of the total pressure due to the molecule of interest, see Chap. 1.7) and is either reduced for a concave surface or increased for a convex surface. In more explicit thermodynamic terms, the chemical potential in the vapor phase (i.e., the chemical potential of the molecules from the liquid that have escaped into the gas phase, μ_A) decreases with partial pressure (p_A):

$$\mu_A = \mu_A^0 + RT \ln p_A (\text{atm}) \qquad (5.10)$$

lar approach can be taken for any of the equations derived from the Laplace equation.
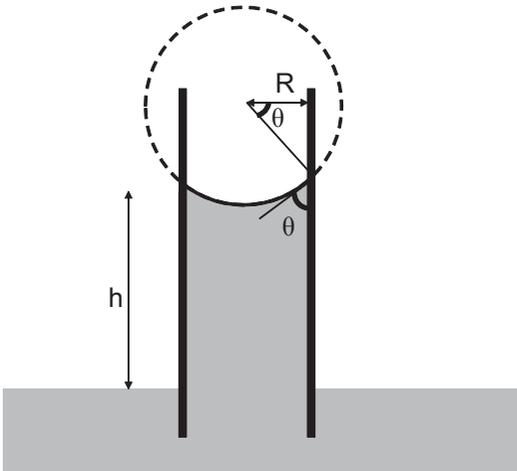
Fig. 5.17 Schematic illustration of capillary rise

Thus lowering the partial pressure lowers the chemical potential in the gas phase and, at equilibrium, the chemical potential of similar molecules across the curved interface in the liquid phase. The chemical potential of a liquid at a convex (Fig. 5.16b) or concave (Fig. 5.16c) is lower or higher respectively than the standard value at a planar surface.

Having identified the equivalent mechanical (in terms of a pressure difference) and thermodynamic (in terms of chemical potential) consequences of surface curvature, we can now look at some examples of their importance in food systems. Remember that the magnitude of the effects increases with surface tension and becomes much greater as the radius of curvature gets very small.

Capillary Rise Water tends to wet glass effectively so it will tend to climb up glass walls against the force of gravity to achieve a contact angle (θ) defined by Young's equation (Eq. 5.3). At the edge of a large container, surface wetting effects cause a meniscus but in a fine capillary surface effects cause the liquid to rise up the tube to a characteristic height (Fig. 5.17). In effect, the weight of the liquid in the capillary hangs from its meniscus. The fluid inside the tube has a curved interface with an effective radius of curvature

of $R.\cos\theta$, where R is the radius of the tube and the curvature reduces the pressure immediately under the meniscus by $2\gamma.\cos\theta/R$ (Eq. 5.9). The low pressure caused by the curved surface must match the hydrostatic pressure of the column of liquid in the tube ($=\rho gh$, where ρ is the fluid density, g the acceleration due to gravity, and h the height of the column in the tube). Rearranging gives the interfacial tension of the fluid in terms of the height of the capillary rise:

$$\gamma = \frac{1}{2}\,\rho gh.R.\cos\theta \qquad (5.11)$$

Capillary measurements remain a useful and precise method to measure surface tension, provided the liquid wets the glass effectively. Capillary pressure is also important as it tends to "suck" liquids into pores and cracks in solid food materials, and capillaries provide a lower energy environment for condensation.

Homogeneous Nucleation If a liquid is cooled to the point that the chemical potential of a crystal phase is just lower than the chemical potential of the liquid then thermodynamics will favor crystallization (Chap. 4). However, the initial crystal formed (the crystal embryo) is very small with a highly curved surface and a consequently higher than expected chemical potential. Therefore, while a large crystal with effectively flat surfaces would have lower chemical potential than the liquid and be stable, a small one would be unstable because of the effects of surface curvature. The pathway for the formation of crystal is blocked by a high surface energy small-crystal intermediate and the kinetics of crystallization are slowed (see further discussion of nucleation mechanisms in Chap. 6).

Ostwald Ripening The surfaces of small particles are highly curved and so their contents have a higher chemical potential than similar molecules at the flatter surfaces of large particles. Therefore, the smaller the droplet, the more readily its contents will dissolve in the surrounding liquid or gas phase. For example, in a foam there
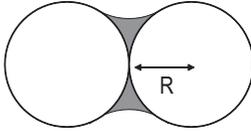
Fig. 5.18 The gap between two spherical particles (radius *R*) acts as a capillary of radius zero at the closed end and 2*R* at the open end. Liquid absorbed into the gap provides a capillary force holding the particles together

is a distribution of bubble sizes and the gas will diffuse from small to large to minimize the surface curvature (see further discussion of Ostwald ripening of crystals and the related phenomenon of crystal accretion in Chap. 6).

Capillary Condensation   A vapor will condense on a surface (e.g., steam on a cold window) when the chemical potential of the liquid phase is lower than the vapor phase. On a planar surface, this occurs at a characteristic dew point (cf. the measurement of water activity, Fig. 1.9). If the surface contains small cracks, then liquid can condense in them at lower vapor pressures because the concave curvature of the surface means the chemical potential of the liquid phase is lowered.

The gaps between grains of a powder function as capillaries of varying width (Fig. 5.18). Not only will vapor condense in these capillaries below the dew point of the liquid, but also the presence of a liquid in the gap will generate a force holding the particles together and agglomerating the powder. Imagine the ring of liquid in the gap between two spherical particles. The concave surfaces mean the liquid is compressed which provides a force to hold the particles together. Consequently, many powders are free flowing if stored below a critical humidity but clump if the moisture content gets too high and adsorbs in the gaps. For spherical particles, the magnitude of the adhesive force depends, to a good approximation, solely on the radius of the particles and the surface tension of the liquid. The magnitude of the forces tends to be much lower for particles with rough surfaces as there are smaller regions of contact.
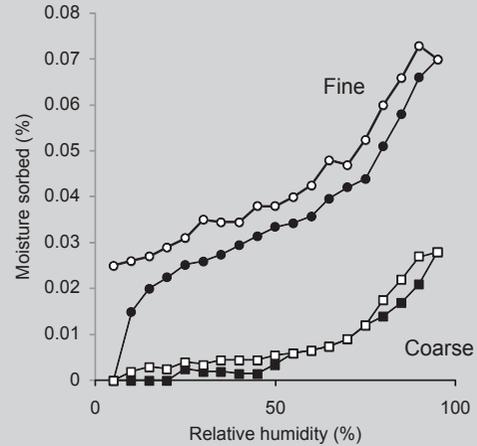
Fig. 5.19   Moisture sorption (*filled points*) and desorption (*open points*) isotherms of fine (*d* = 50 µm) and coarse (*d* = 500 µm) glass beads. (Adapted from Stoklosa et al. 2012)

Example: Powder Caking

Dry crystalline powders (e.g., sucrose, citric acid) are often free flowing, but cake if stored above a critical humidity. In the dry state, the particles adsorb very little moisture, but above a critical level liquid water begins to form and partly dissolves the crystals. The liquid itself provides some adhesion between particles which is reinforced if the material is again dried and the dissolved material recrystallizes to form a solid link. Stoklosa and co-workers (2012) studied the effect of particle size and humidity on the flowability or a range of crystals and also glass beads as a model system of water-insoluble particles. The moisture sorption and desorption isotherm of fine (*d* = 50 µm) and coarse (*d* = 500 µm) glass beads is shown in Fig. 5.19. Notably finer particles adsorbed more moisture than coarse particles at the same humidity, and when the particles were subsequently dried (i.e., desorption), some moisture was still remaining in the fine particles. When the particles were stored at 85 % relative humidity for a week, their flowability was

measured as the "avalanche angle"—the angle at which a powder bed begins to flow. The fine glass beads could be tilted to a greater angle before flowing than the coarse glass beads (37° vs. 25°). The smaller particles were more affected by moisture because of the stronger effects of capillary condensation.

In each of these examples, there is a mathematical expression for the magnitude of the effect in terms of the radius of curvature of the surface and the interfacial tension of the liquid (see Bibliography). These are thermodynamic statements and are correct within the approximations and assumptions used to generate them. One frequently unstated assumption, often violated in foods, is that the surface is unreactive with any adsorbed liquid. For example, the clumping of clay particles could be modeled using the approaches described above because the clay particles are effectively insoluble. The same surface chemistry would favor the adsorption of water onto sugar crystals but the basic theories would not account for the dissolution of the surface sugar into the desorbed liquid. The strength of the aggregates formed would depend on capillary forces but also on the mixing of saturated sugar solutions coating each individual crystals and possible crystal accretion (for more detail see Chap. 6). The theories developed are valid within their assumptions, but their application to foods requires careful consideration.

5.8   Summary

Because most foods are phase separated, there are many types of interface present. Although only a small proportion of the molecules in a food are at the surface, the properties of interfaces are distinct and can determine the properties of the food.

The surface tension (or surface free energy) between two immiscible phases is a consequence of immiscibility and can be understood in terms of the difference in interactions between molecules. When the surface is curved, the elasticity of the surface tends to compress the internal phase which is particularly important for fine particles. We will see its importance in the nucleation and perfection of crystals (Chap. 5). Surface free energy can be reduced by adsorbing amphiphilic materials to the surface or by changing the microstructure to reduce the interfacial area (e.g., dispersed phases tend to be spherical and their average size tends to increase over time). Surface energy effects are most important for fine particles and the properties of interfaces will be seen to be critical in affecting the stability of fine dispersions (Chap. 9).

5.9   Appendix 5.1: The Gibbs Surface

In Sect. 5.3 we looked at the sorption of amphiphilic molecules at an interface as a type of partitioning between the bulk phases and a two-dimensional interfacial "phase." This picture of a surface is physically unrealistic but as we shall see in this appendix, it provides an acceptable and thermodynamically valid model for reality.

The example of two adjacent phases given in Fig. 4.1 shows the concentration changing abruptly as a step function at the surface. In reality, molecules are not so well behaved, and at the molecular scale, a fluid interphase is more like a churning ocean than a placid mill pond. Individual molecules of one phase may briefly penetrate a short distance, perhaps a few molecular diameters, into the adjacent phase before collapsing back into a more hospitable chemical environment (Fig. 5.20a). It is more realistic to draw the properties of the material changing smoothly (though not necessarily linearly) over a thin region of space from one constant value in one phase to another constant value in the second phase. The surface then is not the plane implied by our thinking to date but rather a thin region of space between the phases whose properties change with position.

In thermodynamics, it is helpful to be able to divide a system into parts, each with defined
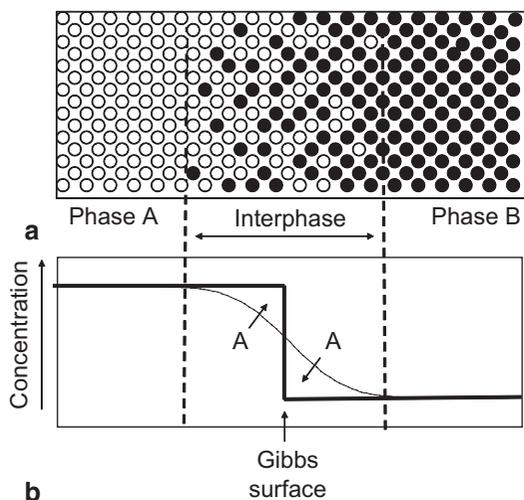
Fig. 5.20 a Distribution of molecules at an interface. The region of varying concentration may extend several molecular dimensions before the bulk concentrations in each phase are established. The changing concentration of one of the phases is shown as a *fine line* in **b**. In the Gibbs approach (*bold line*), the concentration of both phases is assumed to be constant and change abruptly at the interface. The position of the Gibbs surface is selected so the overestimate of solvent concentration by the model on one side of the surface is matched by the underestimate on the other (i.e., areas marked *A* are equal)

Fig. 5.21 a Distribution of an emulsifier (*shaded circles*) across an interface. The emulsifier is more soluble in *Phase B* than in *Phase A* but accumulates to higher concentrations in the interfacial region. The changing concentration of emulsifier is shown as a *fine line* in **b**. In the Gibbs approach (*bold line*) the emulsifier is taken as constant up to the interface and all the excess material (i.e., the sum of areas marked *B*) are assumed to be in the two-dimensional surface phase

and uniform properties (e.g., the oil phase has volume *x,* internal energy *y,* density *z,* and so on) but this way of thinking is hard to accommodate to the interfacial region where composition changes gradually. However, we can follow Josiah Gibbs and sidestep this impasse with the logical trick of mentally replacing the real system with a thermodynamically equivalent but imaginary system consisting of two bulk phases intersecting at a two-dimensional planar interfacial phase (Fig. 5.20b). The imaginary system has the same thermodynamic properties and overall composition as the real one but unlike the real system, the properties and composition of the imaginary phases do not vary with position and can therefore be given unique values.

On one side of the imaginary surface, the concentration of solvent in the model system is

higher than the real system and on the other side, the model system concentration is higher than the real system. Gibbs' convention is to place the imaginary surface so that, for the solvent, the excess concentration on one side of the interface exactly matches the deficit on the other side. This is not the case for solutes and especially important for surface-active solutes. Figure 5.21 shows a typical distribution of a surfactant across the interfacial region; there is a higher concentration in one bulk phase than the other and an accumulation in the interfacial region. Once more constructing an equivalent imaginary system with uniform bulk concentrations up to the planar interface, we can see an excess surfactant concentration on both sides of the imaginary surface (i.e., real concentration is higher than the model). All of this excess emulsifier is assumed to be "in" the two dimensional surface phase—the surface excess concentration.

5.10   Bibliography

Dill and co-authors develop their lattice model of phase separation used in the previous chapter to describe the thermodynamics of surfaces in terms of molecular interactions (Dill, Bromberg and Stigter 2003, Chap. 15) and again that approach is used here. Chapter 2 of "The Colloidal Domain" (Evans and Wennerström 1994) describes the properties of surfaces very clearly.

Similar material is covered in more detail by Heimenz and Rajagapolan (1997) in "Principles of Surface Chemistry" (especially Chap. 6) and by Berg (2010) in "An Introduction to Interfaces and Colloids" (Chaps. 2 and 3).

The properties of surfaces as applied to foods are covered in more depth by Walstra (2003, Chap. 10), McClements (2004, Chap. 5), and Dickinson (1992, Chaps. 2 and 5).

Crystals

6.1 Introduction

A few foods—salt, table sugar, and especially rock candy—are obviously crystals. They look like the crystals familiar from science classes and science fiction films: angular solids with shiny faces and clear geometric shapes (Fig. 6.1). However, there are many other smaller and less immediately obvious crystals in foods. The hardness of butter and ice cream depends on the proportion of the oil or water respectively that is crystalline, even though the individual fat and ice crystals are too small to be seen by the naked eye. Likewise in a starch granule, parts of the amylopectin molecules are present in tiny crystallites that melt upon gelatinization (see Fig. 1.6). All these crystals contribute to the properties of food and share some common features that make them worth considering as a class in their own right.

Crystallization is also interesting as a special case of a phase transition (Chap. 4) and a good illustration of the ways to combine thermodynamics and kinetics to generate structure in foods. A liquid cooled below the freezing point of the solvent is supercooled and a solution concentrated above the solubility limit of the solute is supersaturated. Supercooling or supersaturation means the system is in a two-phase region of the phase diagram which is a thermodynamic precondition for crystallization (i.e., formation of solvent or solute crystals respectively). However, the process of crystallization involves three distinct and kinetically limited steps that may occur sequentially but often overlap:

- Nucleation—initial formation of crystals from the liquid. The number of crystals increases but only a very small mass of material is crystallized.
- Growth—change in total crystal mass without change in number of crystals.
- Perfection—change in crystal structure while the total mass and number of crystals remain constant.

We will start our discussion of crystals by using a simple model of molecular interactions to examine crystal structures, and will then look at the processes of crystallization.

6.2 Crystal Structure

Whether crystals are formed from molecules (e.g., sugars, fats, and water), parts of molecules (e.g., segments of amylopectin in a starch granule), or ions (e.g., sodium chloride), their common distinguishing feature is that they are ordered at the microscopic level, and consequently in their bulk forms. Crystals are low entropy phases that minimize their free energy with strong bonding enthalpy between their elements. A stable crystal is one where a regular structure—defined by strong, fixed intermolecular bonds—is stable against the disordering effects of random thermal motion.

Intermolecular bonds have a characteristic optimum length (σ^*) where they have minimal energy (Chap. 2). A stable crystal is a regular
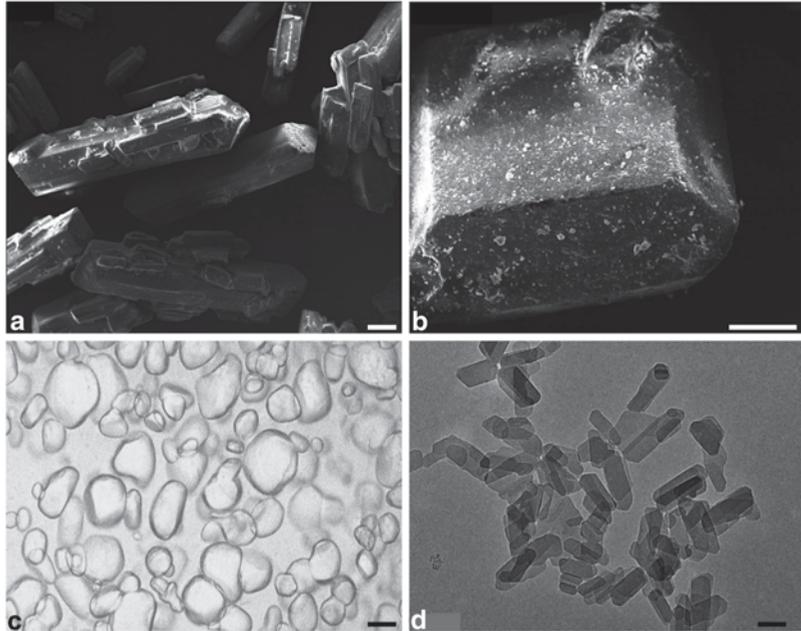
arrangement of molecules (or ions etc.)[1] such
that as many bonds are as close to this optimum
as possible. A good way to find a suitable ar-
rangement is to mentally replace the molecules
with spheres of radius σ^* (Fig. 6.2). This simpli-
fication allows us to forget about the details of
structure, and just work out ways that molecules
would have to be packed to maintain their opti-
mum separation. Certainly, real molecules with
asymmetric shapes and preferred bond angles
would be better represented as shapes other than
spheres, but this simple model suffices to illus-
trate the core relationship between molecular
packing and the properties of crystals. Our prob-
lem is to generate a regular packing arrangement
for these spheres.

> That old man with spotted hands invited me to
> think of the several ways in which cannonballs
> might be stacked on a courthouse lawn, of the
> several ways that oranges might be arranged in a
> crate. "So it is with atoms in crystals, too; and two
> different crystals of the same substance can have
> quite different physical properties." (Kurt Vonne-
> gut 1963, Cat's Cradle, p. 46)

Fig. 6.2 The optimum packing of molecules interacting,
so that their optimum separation (lowest bond energy) is
σ^*, can be found by treating the molecules as hard spheres
of radius σ^*

The first row of cannon balls, oranges, or mol-
ecules is easy and arises simply from the fact that
the spheres are all the same: a central unit sur-
rounded by six neighbors (Fig. 6.3a). Any other
arrangement would be less favorable because
there would be fewer optimum length bonds (i.e.,
contact points between the spheres). The second

---

[1] It gets clumsy to constantly restate that any material
packing regularly can be thought of as a crystal. We will
subsequently use "molecules" as a general term to de-
scribe the elements in a crystal unless we are explicitly
talking about something else.
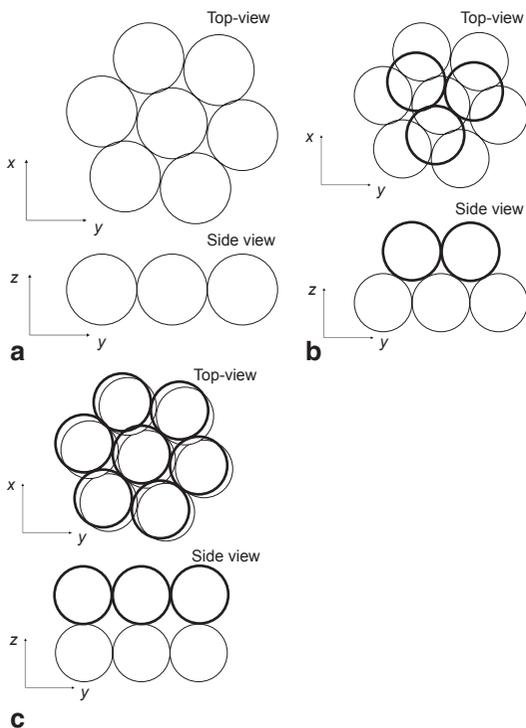
Fig. 6.3 Sample arrangements of *spherical molecules* on a crystal lattice (**a**), the first row of molecules is packed in a regular hexagonal pattern but two alternative structures are formed based on the relative arrangement of the next layer (**b** and **c**)
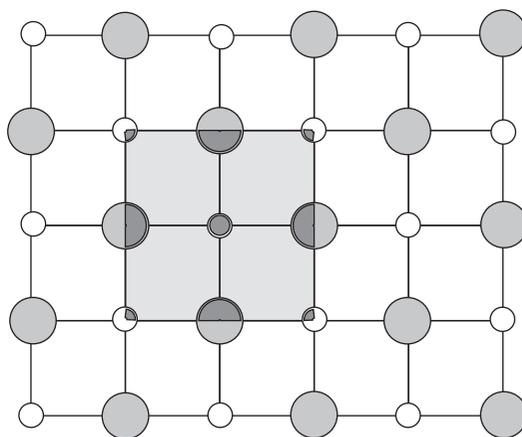

Fig. 6.4 Sodium (*small open circles*) and chloride (*large shaded circles*) packed into a crystal lattice. Layers of the lattice above and below this one are offset by one Na–Cl bond length, so sodium ions are arranged above chloride ions. The *lines* represent the crystal lattice drawn through sodium ions and the *highlighted region* shows the sodium chloride unit cell. (Note the boundaries are drawn through certain ions and only the fraction of the ions inside the *cube* is considered part of this unit cell)

row offers more choices: we could build either by putting the second row spheres into the triangular gaps between the first row molecules (Fig. 6.3b) or just stack them directly above the first row molecules (Fig. 6.3c). Continuing this pattern would grow two crystals of the same material with different packing arrangements of the molecules that would both melt to form the same liquid, i.e., two different polymorphic forms. So what? What different properties might we expect from these different crystals? Firstly, Fig. 6.3b shows a denser crystal than Fig. 6.3c; there are more molecules (greater mass) in a unit volume. Secondly, it would take more energy to melt the crystal in Fig. 6.3b as each molecule has 12 nearest neighbors (6 in the plane, 3 above, 3 below) and so it would be necessary to break 12 optimal bonds to move a molecule out of the lattice. The molecules in the crystal in Fig. 6.3c only have six bonds (4 in the plane, 1 above, 1 below) holding

them into the lattice. We will return to the subject of polymorphism in Sect. 6.6.

Even something as simple as one size of spherical molecules can pack regularly into a number of crystalline forms (i.e., crystal polymorphs). Real crystals are often mixtures of different components with different sizes (e.g., salt crystals contain small sodium ions and large chloride ions, Fig. 6.4). The molecules are not spherical and specific intermolecular interactions may prefer one orientation over another. More complex cases may require a computer to calculate the optimal structure but conceptually the problem is the same—to find a regular structure that optimizes bonding. The regular structures of crystals can be described in terms of their geometry.

The first step in describing a crystal's structure is to divide it into asymmetric units. The asymmetric unit is a reflection of chemical composition, for example, the asymmetric unit of a salt crystal is one sodium ion and one chloride ion. However, just knowing the asymmetric unit does not describe overall arrangement of structure in the crystal. For this, we generate an imaginary lattice by connecting points marked on each asymmetric unit. The choice of where the point is
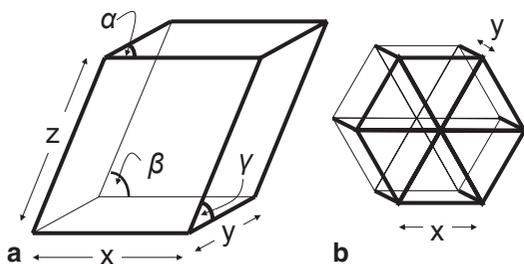
Fig. 6.5 a Generic crystal unit cell with labels for the *x, y,* and *z* repeat distances and the angles between the faces (*α, β,* and *γ*). **b** Hexagonal unit cell. Different levels of symmetry between angles and lengths define the different lattice systems. (see Table 6.1)

fixed is arbitrary, but is often taken as the center of an ion or molecule. The unit cell of the crystal is the smallest identical unit into which the lattice can be divided. A macroscopic crystal of any size can then be generated by stacking unit cells without rotating them in any way.

In a salt crystal, the lattice is usually drawn through the center of the sodium ions (Fig. 6.4). The unit cell is a cube with each face centered on a sodium ion (see highlighted region in Fig. 6.4). In 3-D, each unit cell consists of four sodium ions (one-eighth of each of the eight corner sodium ions plus one-half of each of the six sodium ions on the faces of the cube) and four chloride ions (one in the center of the cube plus one-quarter of each chloride on the 12 edges of the cube). There are 14 possible unit cell structures, the Bravais lattices, which group into seven crystal systems based on their symmetry (e.g., triclinic, orthorhombic, Fig. 6.5, Table 6.1).

The fixed and repeating interatomic distances of a unit cell are most readily measured by X-ray scattering. X-rays are high-energy, short-wavelength ($\sim 10^{-10}$ m) light rays. When a beam of X-rays hits a sample, some of the energy passes straight through, while some interacts with regions of high electron density (atoms) and scatters into different directions. An X-ray detector mounted just out of the beam path can be used to detect an angular dependence to the scattering. If the sample is amorphous (i.e., noncrystalline) then there is no particular pattern to the scattering, but the regular arrangement of atoms

in a crystal lattice gives a characteristic pattern. Figure 6.6 shows a regular crystal of model spherical "atoms." The regular spacing gives rise to multiple planes of atoms in the crystal, one example shown in Fig. 6.6d but many others could be drawn through the same material. A beam of X-rays incident on the crystal at an angle θ is partly scattered by the top plane and partly scattered by the second plane. The light scattered by the second plane travels a distance $2.d.\sin\theta$ further than the light scattered by the first. If that distance is an integer number (*n*) of wavelengths (λ) then the two scattered waves will positively interfere and a peak will be seen in the X-ray pattern, that is, the Bragg equation:

$$2d \cdot \sin \theta = n \cdot \lambda \qquad (6.1)$$

Measuring the positions of the peaks and knowing the wavelength of the X-ray source, it is possible to calculate the spacing between planes. Real crystals have multiple planes of atoms in their structures, so there are multiple peaks in the X-ray pattern. However, once they are measured, the 3-D configuration of all the atoms in the crystal can be calculated to good precision. In most food applications, X-ray crystallography is conducted on a mixture of small crystals rather than on a single large crystal (i.e., powder diffraction). The crystals are randomly orientated to the beam, and rather than the sharp points of a single crystal pattern, a powder pattern is a series of bands at defined angles to the beam. The position and strengths of the different peaks are used as an indication of the type and number of the crystals present. Peak width can also be used to calculate the size of the crystallites in the powder.

6.3　Nucleation

The fact that there is a thermodynamic driving force for crystallization does not mean crystals will form over the life of the product. A striking example is honey, which is available either as a pourable, clear liquid or a spreadable, turbid semisolid. Both forms of honey have the same chemical composition ($\sim 82\%$ sugar largely

Table 6.1 Crystal lattice systems with some food examples (adapted from Hartel 2001). The lengths x, y, and z are the lengths of the unit cell along the three axes while α, β, and γ are the angles between the X and Y, X and Z, and Z and Y planes respectively. (See Fig. 6.5 for detail)

	Unit cell geometry	Food examples
Cubic	$x=y=z;\ \alpha=\beta=\gamma=90°$	Sodium chloride
Hexagonal	$x=y\neq z;\ \alpha=\beta=90°,$ $\gamma=120°$	Ice triacylglyerol subcell[a] (α-polymorph) B-form of starch
Monoclinic	$x\neq y\neq z;\ \alpha=\beta=90°\neq\gamma$	Sucrose Lactose A-form of starch
Orthorhombic	$x\neq y\neq z;\ \alpha=\beta=\gamma=90°$	Triacylglycerol subcell[a] (β'-polymorph) Citric acid Fructose (D) Glucose
Trigonal	$x=y=z;\ \alpha=\beta=\gamma\neq90°$	
Tetragonal	$x=y\neq z;\ \alpha=\beta=\gamma=90°$	Urea
Triclinic	$x\neq y\neq z; \alpha\neq\beta\neq\gamma=90°$	Triacylglycerol subcell[a] (β-polymorph) Triacylglycerol full crystal (β'-polymorph)

[a] The concept of subcells is sometimes used to describe the packing of the fatty acid chains in triacylglycerols
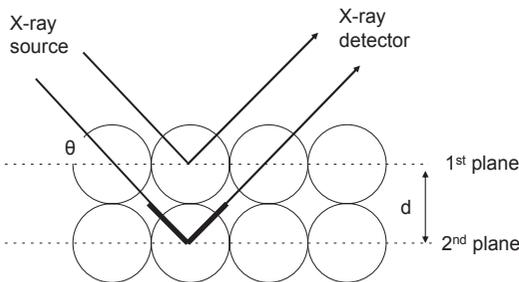
Fig. 6.6 X-ray scattering from two planes of atoms in a crystal. The two reflections will constructively interfere if the additional distance traveled by the reflection from the lower plane (shown as a *bold line*, $2.d.\sin\theta$) is an integer number of wavelengths. A crystal has other planes corresponding to other regular spacings (e.g., the distance between diagonal neighbors) that give rise to another peak at a different angle

fructose and glucose) and only differ because the sugars in the latter form have crystallized whereas those the former have not. The solid form can be converted into the liquid form by heating to dissolve the crystals, and then remaining stable as a liquid for a long period before grains of crystalline sugar begin to form over months or years. In fact, while honey is an extreme example, there is always a delay between the onset of supercooling/supersaturation and the appearance of crystals.

The time it takes for the first crystal to form (J^{-1}, s) is often expressed as its reciprocal, the nucleation rate (J, s^{-1}, i.e., the number of crystal nuclei formed per second). A sample is stable for a long time in a supercooled state if the rate of crystal nucleation is low. The relationship between the rate of nucleation and temperature given is by an Arrhenius-type relationship; the Fisher–Turnbull equation:

$$J = J_0 \exp\left(\frac{-\Delta G_n}{k_B T}\right) \qquad (6.2)$$

where ΔG_n is the free energy barrier for nucleus formation, J_0 is the frequency factor (a constant), T is absolute temperature, and k_B is the Boltzmann constant.

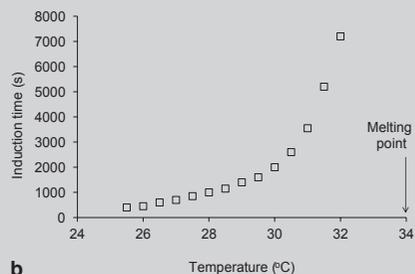**Example: Induction Time for Milk Fat Crystallization**

Milk fat can be separated into fractions for different applications by melting it and then cooling to a specific temperature so that some of the fat crystallizes. The crystalline fraction can be separated from the liquid oil by centrifugation. It has a higher melting point which is more suitable for bakery applications. Herrera et al. (1999) studied the processes of milk fat crystallization by rapidly cooling molten milk fat to below its melting point and measured turbidity as a function of time. Molten milk

fat is clear, but becomes turbid once crystals form. The induction time for crystallization was defined as the time before the turbidity increased (Fig. 6.7a). It should be noted that the measured induction time (τ) is a sum of the time it takes for the first crystals to form (J^{-1}) and then grow to a size that the analytical technique selected can detect their presence ($\tau_{instrument}$), i.e., $\tau = J^{-1} + \tau_{instrument}$. For example, a higher-magnification microscope would see crystals before a lower magnification microscope and τ would be different even if J^{-1} were the same. In general, studies of nucleation kinetics are conducted using techniques sensitive to the first appearance of tiny crystals so the measurement of τ is most directly related to J^{-1}. Turbidity is very sensitive to small crystals, so in this work, the $\tau_{instrument}$ term is neglected and the measured induction time is taken to be the nucleation time J^{-1}.

All samples were stored below their melting point ($T_m = 34\,°C$) and so were thermodynamically driven to crystallize, however the kinetics of the nucleation delayed the process; especially at low levels of supercooling (Fig. 6.7b). For example, while oil could be stored for 2 h at 2 °C below its melting point before crystals were detected, a sample stored 8 °C below its melting point was only stable for 7½ min. Knowing the nucleation time, the nucleation rate was calculated as its reciprocal. For example, if nucleation was observed after 400 s, then there was one nucleation event every 400 s and the number of nucleation events per second, i.e., the rate of nucleation, was $J = 0.0025$ s^{-1}. The temperature dependence of nucleation rate was used to investigate the energetics of nucleation. The logarithm of the nucleation rate constant was inversely proportional to reciprocal absolute temperature (Fig. 6.7c) and, according to Eq. 6.2, the slope of the line is $-\Delta G_n/k_B$ so the free energy for nucleation is 317 kJ mol^{-1}.
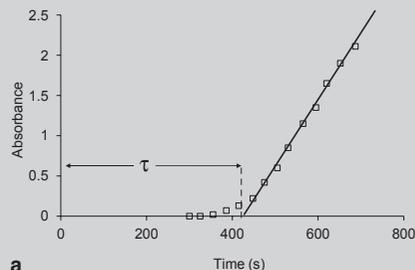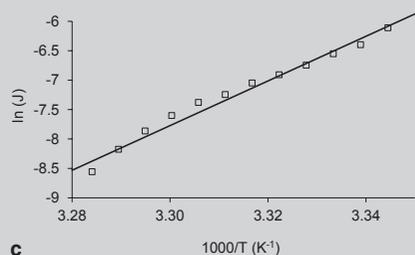
Fig. 6.7 a Turbidity of a supercooled milk fat sample showing the calculation of the induction time for crystallization (J^{-1}). **b** Effect of temperature on the induction time for crystallization of milk fat. **c** Fisher–Turnbull plot of induction time data. (Replotted from Herrera et al 1999)

So what is the molecular basis for the energy barrier slowing crystal nucleation? We saw in Sect. 5.7 that molecules in very small crystals have a higher chemical potential than similar crystals in the bulk due to surface curvature effects. The surface excess free energy creates a barrier to the formation of small crystals and acts as a high-energy intermediate to delay the formation of macroscopic crystals. As an illustra-

tion, imagine the process of crystallization as the formation of larger and larger crystals in a super-saturated or supercooled solution. When a crystal starts to form, the free energy of the system de-creases. The decrease in free energy depends on the number of molecules in the crystalline state and hence, the volume of the crystal (i.e., decrease in free energy is proportional to the cube of crys-tal size). On the other hand, there is a surface be-tween the crystal and the solution. Surfaces have an excess Gibbs free energy equal to the product of surface tension and surface area (Eq. 5.1). The increase in surface free energy depends on the number of molecules at the interface and hence, the surface area of the crystal (i.e., increase in free energy is proportional to the square of size). The net free energy change (ΔG) for forming a crystal of radius r is the sum of a positive term propor-tional to the surface area (r^2) and a negative term proportional to volume (r^3):

$$\Delta G = k_{surf} \cdot r^2 - k_{vol} \cdot r^3 \qquad (6.3)$$

where k_{surf} and k_{vol} are constant for the surface and volume terms. The net change in free energy for forming a crystal of given radius increases to maximum (ΔG_n) at a critical value of crystal size (r^*), then decreases (Fig. 6.8). Crystals smaller than r^* (i.e., crystal embryos) will reduce their free energy by getting smaller and melting, while crystals larger than r^* (i.e., crystal nuclei) will reduce their free energy by growing to a visible size. In a pure liquid, a crystal of size r^* is the unstable intermediate from kinetic theory and ΔG_n is an energy barrier that slows the rate of the crystallization.

For crystallization of a pure supercooled liq-uid, the volume term, k_{vol} is given by:

$$k_{vol} = \frac{\Delta H}{V_m} \cdot \frac{T_m - T}{T_m} \qquad (6.4)$$

where ΔH is the enthalpy of fusion of the crys-tals, V_m is the molar volume, T_m is the melting point of the crystals, and T is the temperature. In-creasing the supercooling ($T_m - T$) increases the importance of the volume term and both decrease r^* and the energy barrier for nucleus formation
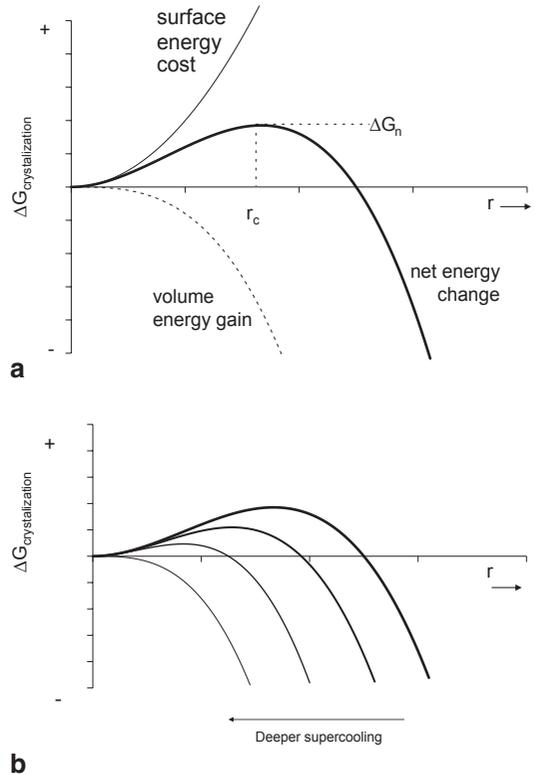
Fig. 6.8 The change in Gibbs free energy on forming a crystal of size, r, is the sum of a volume and surface con-tribution. **a** The combination of a negative volume term and positive surface term gives rise to a maximum en-ergy at $r=r^*$; the free energy for the formation of a stable nucleus. **b** Lower temperatures increase the magnitude of the volume term and reduce r^* and the energy required to form a stable nucleus

(Fig. 6.8b). The barrier to nucleation decreases at lower temperatures, so nucleation proceeds more quickly (see the milk fat data in Fig. 6.7).

How does this small crystal nucleus first form? It could simply be added, bypassing the need for a nucleation step in the liquid state. Molten choc-olate and sugar solutions are sometimes "seeded" with a small amount of finely ground solid choc-olate or fine sugar crystals respectively, to expe-dite nucleation. However, in their absence, the only way for a pure, supercooled liquid to crys-tallize is if the random movement of a few mole-cules spontaneously and very transiently arranges them into a configuration similar to a very small crystal. This process is illustrated schematically for triacylglycerol crystallization in Fig. 6.9. In

most cases, the region of order formed is smaller than $r*$, so ΔG decreases with decrease in r and the crystal embryo rapidly disperses without visible crystallization. Occasionally though, a larger number of molecules will spontaneously form a region of order bigger than $r*$. For these tiny crystals, ΔG decreases with increase in r, so the crystal nucleus will continue to grow and eventually a visible crystal will be seen. This process is called homogenous nucleation because the nucleus arises spontaneously from the homogeneous liquid. However, homogeneous nucleation almost never occurs in real foods because another mechanism—heterogeneous nucleation—occurs much more effectively.

Heterogeneous nucleation occurs at the surface of some solid impurity (usually the walls of a container or perhaps specks of suspended dust). A good example of heterogeneous nucleation is the manufacture of large rock candy crystals. A stick or piece of string is dipped into a supersaturated sugar solution and provides the heterogeneous nucleation sites that the large crystals grow from. The presence of the surface dramatically lowers ΔG_n compared to the homogeneous nucleation so that in practice, only the heterogeneous mechanism is seen. Different materials have different efficiencies as nucleation catalysts but the mechanism that allows a particular surface to be effective is not well understood. One widely reported theory is that a catalytic surface can make good contact with the crystal nucleus. In terms of surface energy, this means the crystal must wet the impurity surface effectively (see discussion of wetting angle, Chap. 5). For example, the crystal in Fig. 6.10a makes no contact with the surface while the crystals in Fig. 6.10b and c make increasingly good contact (characterized by a lower wetting angle, θ). The surface in Fig. 6.10a would not be catalytic to nucleation as crystals cannot develop on it while in Fig. 6.10b, it would be effective, and in Fig. 6.10c, it would be even more effective. Another theory for heterogeneous nucleation is the atomic spacing on the catalyst surface corresponds to the pattern of the crystal and this the "matching" facilitates nucleation. Measurements of nucleation kinetics can be used to compare the effectiveness of different catalysts.

a

b

c
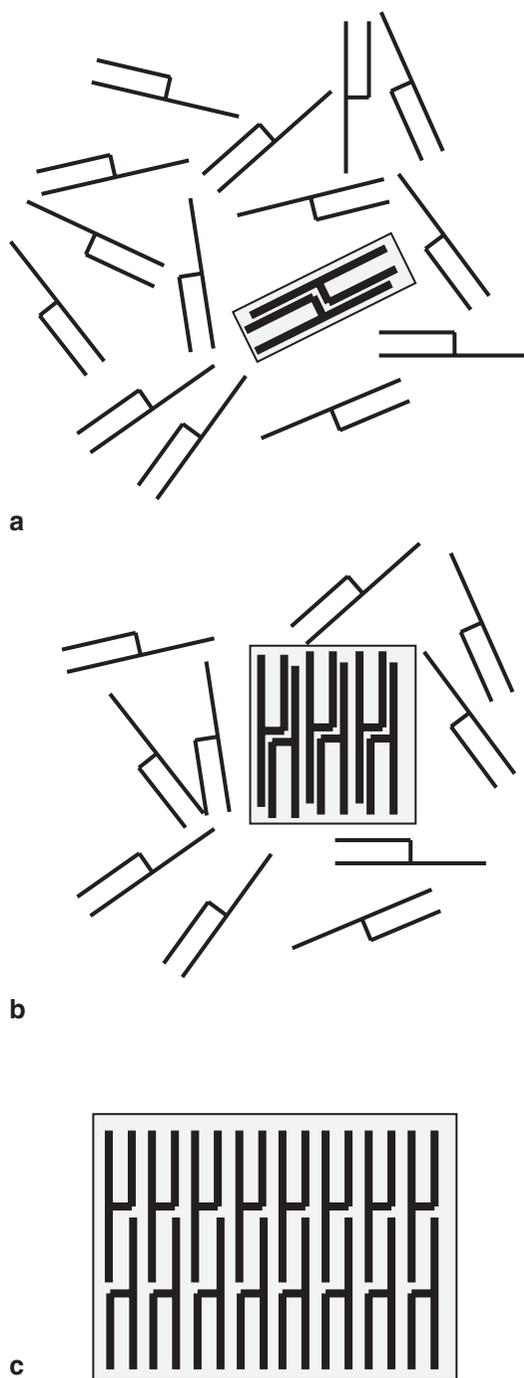
Fig. 6.9 The process of nucleation in triacylglyerol molecules. **a** Above the melting point, small, short-lived molecular assemblies can form but rapidly disperse, **b** as temperature decreases, the size of the crystal embryos and the average time before they dissociate, increases until **c** one exceeds critical radius and can grow to macroscopic size
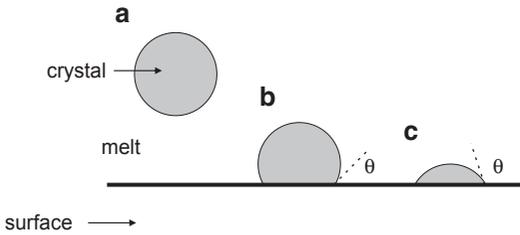
Fig. 6.10 Heterogeneous nucleation by a surface. **a** The crystal does not wet the surface and nucleation is homogeneous, while in **b** and **c**, the crystal wets the surface progressively better and the rate of heterogeneous nucleation increases

Example: Supercooling Coconut Oil

Coconut oil is semicrystalline at room temperature and has a soft butter-like texture but melts to a liquid on heating. Tangsuphoom and Coupland (2009) used differential scanning calorimetry to study coconut oil crystallization as a bulk fat and as fine emulsion droplets dispersed in water (Fig. 6.11). Peaks above (or below) the baseline are due to the release (or absorption) of heat due to the formation (or melting) of crystals. The melting

profiles of bulk and emulsified coconut fat were similar with the single endothermic peak ending at about 24 °C. This property makes coconut oil an attractive ingredient for cosmetics, it can be scooped out of a jar as a semisolid but melts when rubbed into the skin. When the liquid oil was cooled slowly, it started to crystallize (i.e., onset of the exothermic peak) a few degrees below the melting point because of the delay due to nucleation (ΔT_1). When it was cooled more quickly, it could reach a much lower temperature before crystallization was observed (ΔT_2), as there was less time for the oil to nucleate. (Also, note that fast cooling gave a double peak, suggesting the mechanism of crystallization was a two-step process.) When the coconut oil was emulsified, the onset of fat crystallization was much lower (ΔT_3). Presumably, a finite number of unknown impurities in the bulk oil acted as crystal nuclei and allowed the fat to crystallize heterogeneously (Fig. 6.12). When the same oil was emulsified into small droplets, the number of droplets vastly exceeded the number of catalysts. The small proportion of droplets

Fig. 6.11 Cooling and heating thermograms of coconut oil and coconut oil emulsion (volume fraction 10 %, diameter 0.5 μm stabilized with polyoxyethylene sorbitan monolaurate). Exothermic transitions are shown *downward* and are scaled to allow easy comparison of the position of peaks between systems. (Data replotted from Tangsuphoom and Coupland 2009 along with other measurements made in the author's laboratory)
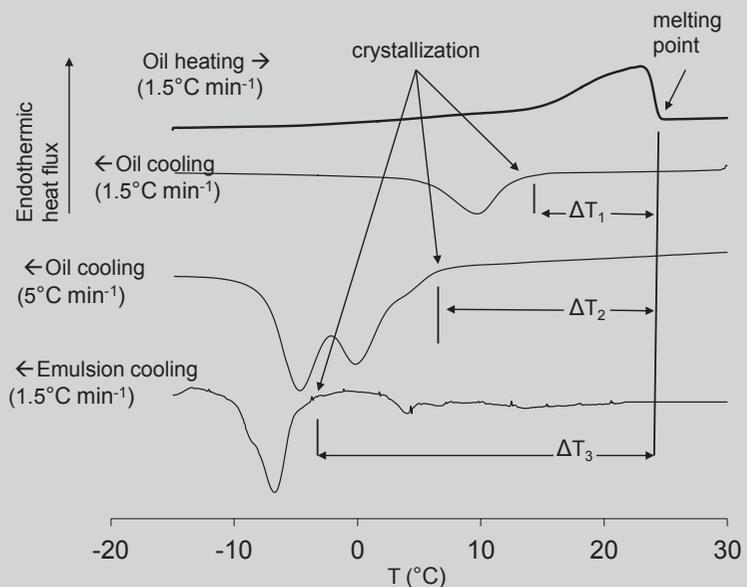
Fig. 6.12 A few nucleation catalysts are sufficient to catalyze the heterogeneous nucleation of bulk oil, but if that oil is distributed into a large number of small emulsion droplets, only those droplets containing a catalyst can nucleate heterogeneously and the remainder must nucleate via the less-efficient homogeneous mechanism at a lower temperature. The *lines* separating the individual crystals in the bulk solid fat are grain boundaries
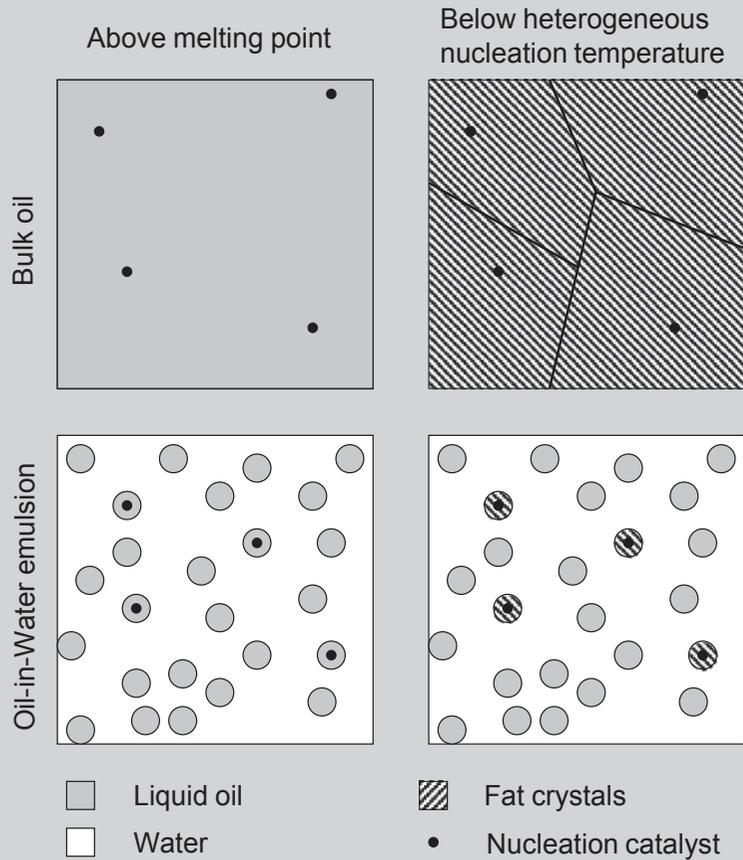
containing a nucleation catalyst nucleate heterogeneously, but the remaining droplets are effectively pure liquid oil and nucleate homogeneously. Indeed in many cases, the fraction nucleating heterogeneously will be negligible and all the observed crystallization is by homogeneous nucleation. Homogeneous nucleation is less efficient than heterogeneous, so allows a much deeper degree of supercooling.

6.4   Crystal Growth

The amount of material in the crystal nuclei is insignificant and the overwhelming majority of the mass crystallization occurs during its subsequent growth. The thermodynamic driving force for crystal growth is the free energy difference between the current conditions and the conditions at equilibrium. Thus, once a supersaturated solution nucleates, the nuclei will tend to grow until the concentration of solute in solution has fallen to the saturation point. As crystal growth progresses and the system gets closer to equilibrium, the thermodynamic driving force for growth and the rate of the processes decrease. The last stages of crystal growth may be so slow that many foods never reach their equilibrium state. The process of mass crystallization is frequently modeled using the Avrami equation:

$$X/X_0 = 1 - kt^n \tag{6.5}$$

where X is the crystal load at time t and X_0 is the equilibrium crystal content. The constant k is a measure of the rate and the n parameter is characteristic of the mode of growth.

Example: Use of the Avrami Equation to Model Milk Fat Crystallization

In the same paper discussed above, Herrera et al. (1999) used nuclear magnetic resonance (NMR) spectroscopy to measure the changing solid fat content (SFC) of milk fat and milk fat fractions as they crystallized isothermally at a range of temperatures. NMR spectroscopy depends on the property of certain atomic nuclei known as spin. For SFC measurements, the spin of hydrogen nuclei is usually used. Spin is a quantum property, so nuclei can only have two values, up or down. Under normal conditions, the nuclei are distributed between the two states, but in a strong magnetic field, all of the nuclei "line up" to the same value. In the SFC measurement, the nuclei are first aligned in a magnetic field and then a pulse of radio waves is used to disrupt their alignment. After the pulse ends, the nuclei once again align themselves in the magnetic field and in doing so, generate a measurable signal. The signal decays via a first order mechanism as the nuclei return to equilibrium, but the kinetics of the process is much faster for nuclei of atoms in the solid state than for those of atoms in the liquid nuclei. Thus the decay function for a semisolid fat is the combination of two overlaying exponential functions which can be deconvoluted to reveal the proportion of solid fat present. NMR has a limited sensitivity to low solid fat contents, so the induction times measured in this experiment are both longer and less representative of the nucleation process than the turbidity measurements described earlier. Herrera et al. observed an approximately sigmoidal increase in solid fat content with time (e.g., milk fat at 25 °C in Fig. 6.13).
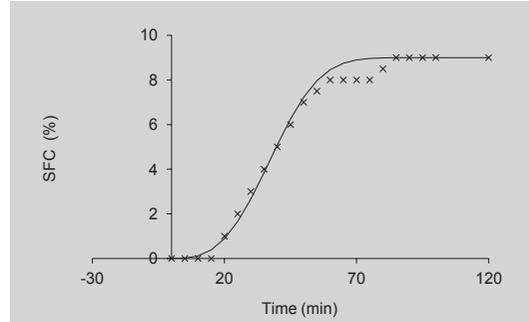
Fig. 6.13 Solid fat content of a milk fat sample during isothermal crystallization at 25 °C (Adapted from Herrera et al. 1999). *Line* shows the fit of the Avrami equation to the data

They modeled their data using Eq. 6.5 and found a value of $n=3$ best described their results. The rate constant of the crystallization process (k) increased at lower temperatures because the driving force for crystal growth was greater.

Crystal growth is a combination of mass and heat transfer processes. Before they can be incorporated, molecules in the liquid phase must diffuse to the crystal surface. As the crystallizing molecules diffuse to the growing face, the noncrystallizing molecules diffuse away from it. For example, when a dilute sugar solution freezes, water molecules are plentiful close to the ice surface and also diffuse quickly, so their movement to the crystal surface is not rate limiting. Instead, the growth of the ice crystals is limited by the diffusion of larger sugar molecules out of the way. If the noncrystallizing molecules do not diffuse away in time, then they may be trapped as a defect in the crystal. Mass transport can be increased up to a point by mixing, but there is always a stagnant layer close to the surface where diffusion is the only mode for mass transfer.

Once the crystallizing molecule reaches the surface, it is considered part of the adsorbed layer but must go through a series of further ordering steps before it is fully incorporated into the lattice. First, the molecule must align itself properly with the lattice structure of the face. This process

is particularly slow for large and asymmetric molecules such as triacylglycerols. Some molecules must also lose water of hydration (e.g., sucrose has up to six hydrating water molecules in solution, yet crystallizes anhydrously) or change their anomeric form (e.g., lactose molecules in solution shift between α- and β-ring forms via the straight chain form, yet the crystals are either purely α-lactose or β-lactose). The crystallizing molecules must also diffuse in two dimensions around the surface until they can find a suitable docking site where they can be incorporated into the lattice. If a crystal is growing quickly, molecules will tend to adsorb at the first potential binding site they reach and leave gaps and imperfections in the lattice resulting in a rough surface on the crystal. Finally, forming bonds molecule into the crystal leads to a release of heat and a rise in local temperature that must be dissipated either by conduction into the crystal itself or convection into the surrounding liquid. Water has a higher enthalpy (~ 330 J g^{-1}) of fusion than most sugars (~ 100 J g^{-1}) or fats (~ 200 J g^{-1}), so the rate of heat removal is often rate limiting in food freezing.

Any one of the steps in crystal growth (mass transport to or from the growing crystal surface, molecular rearrangement and alignment, surface binding, and heat dissipation) could limit the overall rate of the mass crystallization.

6.5   Crystal Size and Shape

The mass of crystalline material formed at equilibrium depends on the phase diagram (Chap. 4) but the number of crystals is equal to the number of nucleation events. The greater the number of nuclei, the smaller the average size considering the same mass of material must divide over more crystals (Fig. 6.14). Sometimes small crystals are desirable (e.g., for a smooth mouthfeel in a fondant filling) while in other cases, larger crystals may be better (e.g., to allow easy centrifugal separation of table sugar crystals from the remaining saturated solution during manufacturing). By similar reasoning, if nucleation occurs rapidly and then stops before substantial growth
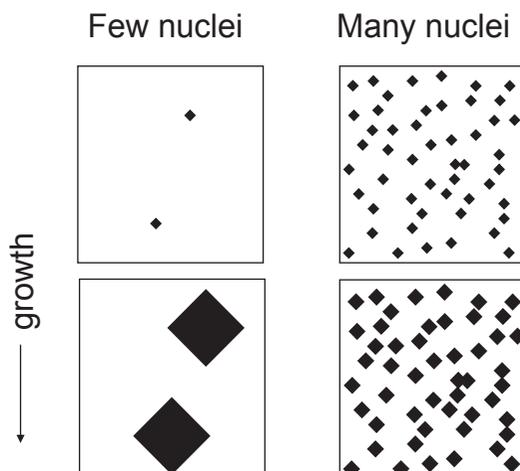
Fig. 6.14 Faster nucleation means the same mass of crystalline material is divided between more crystals and the average size is smaller

starts, then all the crystals will be similarly sized. Alternatively, if the rate of nucleation is comparable or even slower than the rate of growth then the nuclei that form first will have more time to grow and end up much larger than the nuclei that form later, and the crystals will be polydisperse (Fig. 6.15). Polydisperse crystals are harder to separate from a liquid by centrifugation or filtration.

The macroscopic shape of a crystal is determined by the packing of molecules in the unit cell but, is not necessarily the same shape as the unit cell. For example, lactose has a monoclinic unit cell but there is a large range of shapes observed (Fig. 6.1). The faces of crystals correspond to the planes of atoms in the lattice, but the relative size of each face depends on how quickly each one grows (i.e., a slow-growing face will tend to be smaller in the final crystal). For example, Fig. 6.16 shows a series of macroscopic crystals that could be formed from a cubic lattice. The angles between the faces are the same in all cases, and the different shapes result from different rates of growth of the shaded face. Defects in the crystal structure such as the presence of impurities can affect the shape of the crystals.

Even after the crystals have grown, there can be changes in shape to reduce the area of contact between phases through a process of crystal perfection. This can occur either by Ostwald ripening where molecules diffuse from small to larger
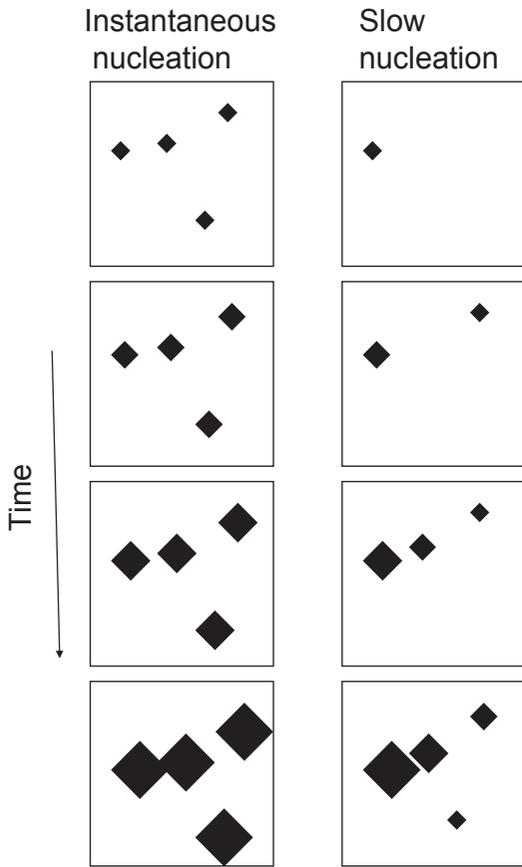
Fig. 6.15 Instantaneous nucleation leads to similarly sized crystals while slow nucleation leads to a wider distribution of crystal sizes
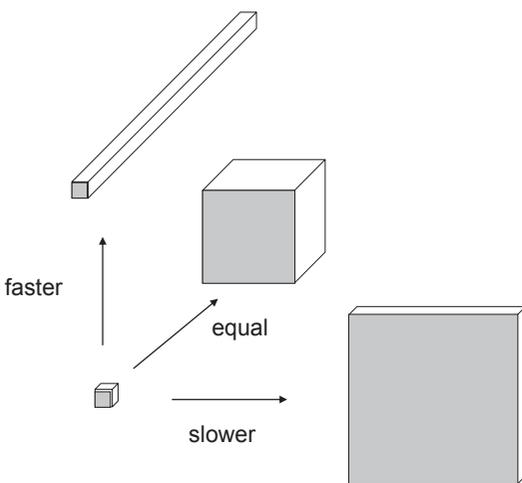

Fig. 6.16 Different shapes form due to the different rates of growth of different faces of a cubic crystal. If the *shaded face* grows faster or slower than the other faces, it will end up smaller or larger respectively in the final crystal
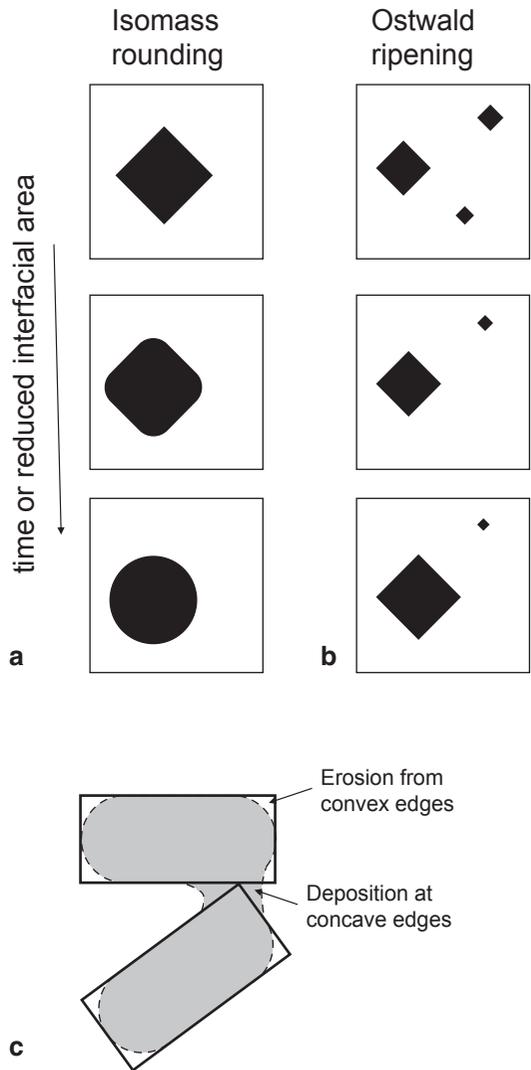

Fig. 6.17 Mechanisms of crystal perfection. **a** Isomass rounding, molecules diffuse from sharply curved to flatter faces. **b** Ostwald ripening from small to large crystals. Both mechanisms are driven by the thermodynamic pressure to reduce the area of contact between crystal and solution. **c** Crystal accretion due to isomass rounding, causing the erosion of material from convex regions and deposition at concave regions, and leading to two adjacent crystals (*thick lines*) bonding to one another over time (*shaded region, dashed lines*)

crystals (Fig. 6.17a) or by isomass rounding, where molecules diffuse from sharp edges and points to flatter surfaces (Fig. 6.17b). Ostwald ripening will lead to an increase in average crystal size and a reduction in the number of crystals. Isomass rounding will cause a change in crystal
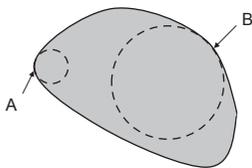
Fig. 6.18 In an irregularly shaped crystal, the effective radius at a given point is that of a sphere with similar surface curvature. Thus, point *A* has a lower effective radius than point *B,* and material will tend to erode from *A* faster than from *B*

shape. Isomass rounding can also cause adjacent crystals to bond to one another, as material will tend to be eroded from convex surfaces and be deposited at concave surfaces (i.e., accretion, Fig. 6.17c). These processes tend to increase the average size of crystals.

Both Ostwald ripening and isomass rounding are driven by the enhanced solubility of finely divided materials due to the effects of surface curvature (see Sect. 5.7). The solubility of a particle increases with decreasing radius. The effective radius at any point on the surface of a nonspherical particle is the radius of a sphere with similar curvature as that point on the surface, i.e., large at flat surfaces and small at sharp edges (Fig. 6.18). Consequently, the sharp edges tend to dissolve before the flat edges in response to small increases in temperature.

Example: Ice in Ice Cream

Hagiwara and Hartel (1996) used microscopy to measure the changing size distribution of ice crystals as a function of time and showed that average size was proportional to time raised to the power of one-third (Fig. 6.19). This type of coarsening kinetics is characteristic of Ostwald ripening. The rate of coarsening increased with increasing storage temperature as there was more liquid water to facilitate diffusion between the crystals in the frozen product and was usually slowed by the presence of a polysaccharide stabilizer. It should be noted that coarsening of ice crystals occurs via a variety of mechanisms and pure Ost-
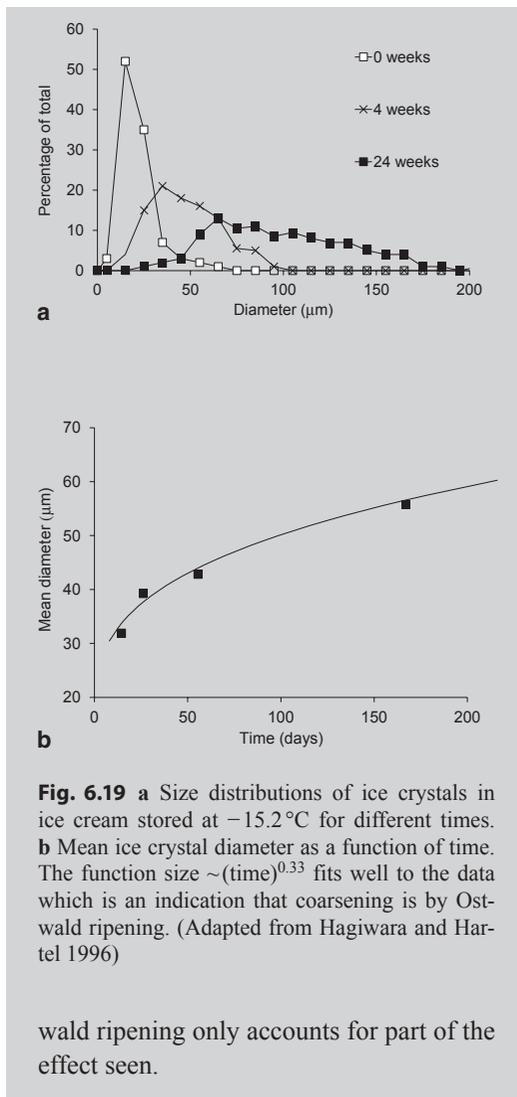
Fig. 6.19 a Size distributions of ice crystals in ice cream stored at $-15.2\,°C$ for different times. **b** Mean ice crystal diameter as a function of time. The function size $\sim(\text{time})^{0.33}$ fits well to the data which is an indication that coarsening is by Ostwald ripening. (Adapted from Hagiwara and Hartel 1996)

wald ripening only accounts for part of the effect seen.

6.6 Polymorphism

Even the simple spherical "molecules" discussed earlier could crystallize into different arrangements or polymorphic forms (Fig. 6.2). Indeed, many real substances are polymorphic (i.e., form multiple types of crystal that melt to the same liquid). The molecules in the different forms are packed differently with the more stable (i.e., higher-melting) forms being more ordered (i.e., lower entropy) with more optimized internal bonding (i.e., lower enthalpy). This is illustrated
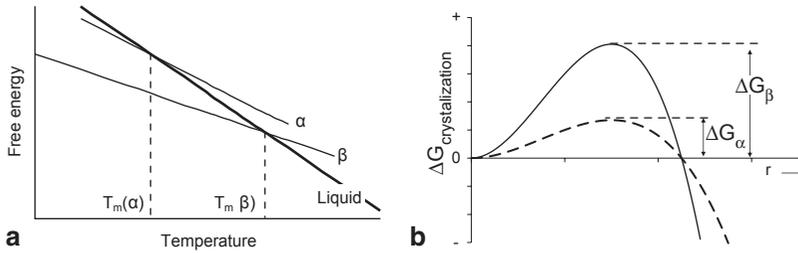
Fig. 6.20 **a** Free energy as a function of temperature for the liquid and two crystal polymorphs (α and β) of the same material. **b** Free energy as a function of crystal radius for the same crystals (Adapted from Rousset 2002)

in Fig. 6.20a, which shows the free energy of two different polymorphic crystals and their liquid melt as a function of temperature (compare with Fig. 4.4). The more stable β-polymorph has lower entropy (i.e., smaller slope) and lower enthalpy (i.e., lower y-axis intercept) than the less stable α-polymorph. Consequently, the β-polymorph free energy intersects with the liquid line at higher temperatures than the α-polymorph and has a higher melting point.

However, the free energy barrier to nucleation is greater for the more stable polymorphs (Fig. 6.20b, compare with Fig. 6.8). The k_{volume} term in Eq. 6.3 is a larger negative number because the "more perfect" crystal has a lower free energy, while the positive $k_{surface}$ term is also greater because of a greater interfacial tension between the crystal and the melt. Consequently, polymorphic materials usually crystallize initially into less-stable forms, although they may later undergo polymorphic transitions to more stable forms. Changes in crystal polymorphism are often quite slow and can be regarded as another form of crystal perfection. An important illustration of crystal polymorphism in foods is seen in chocolate tempering.

The fat in chocolate is cocoa butter—a mixture of triacylglycerols based largely around palmitic, stearic, and oleic acids. Cocoa butter can crystallize into six main polymorphic forms each with different molecular packing arrangements and different physical properties (Table 6.2). When liquid chocolate is poured into a mold and allowed to cool, the cocoa butter crystallizes, eventually reaching about 75 % SFC at room temperature. However, the type of crystals formed and the quality of the product depend on the cooling process. Cooling quickly from the molten state allows a mixture of the less stable polymorphic forms to develop. Therefore the chocolate is crumbly and tends to stick in the molds. Instead, a time–temperature process known as tempering is used to cause the cocoa butter to crystallize as the type *V* polymorph (Fig. 6.21). Correctly tempered chocolate breaks with a "snap" and detaches from the mold during cooling, so intact pieces can be packaged. From the models of polymorphism, we would expect the more stable polymorphs to be more dense (i.e., more efficient packing means more molecules per unit volume, Fig. 6.2). Dense Crystals mean the product contracts more on cooling and detaches from the mold. The product is also mechanically stronger, so it snaps rather than crumbles. Why then is it necessary to temper the chocolate to get the correct polymorphic form?

As liquid chocolate is cooled below the melting point of each polymorphic form, the melt becomes supercooled with respect to that crystal and thermodynamically crystallization can proceed. However, the nucleation time for more stable polymorphic forms is always longer than for less stable forms (Fig. 6.20b). For example, if hot cocoa butter is rapidly cooled to room temperature, it is supercooled with respect to the type III, IV, V, and VI polymorphs and crystals of any of these could form. Despite this, the less stable polymorphic forms tend to nucleate more quickly than the more stable, and in the mixture of crystals, the preponderance will be less-stable polymorphs. Less-stable polymorphic forms are less dense so the product does not contract on cooling and instead sticks to the mold and lacks the expected "snap."

Table 6.2 Properties of the main polymorphic forms of cocoa butter. (Wille and Lutton 1966)

Form	Melting Point (°C)
I	17.3
II	23.3
III	25.5
IV	27.3
V	33.8
VI	36.3

Density ↓   *Nucleation rate* ↑   *Enthalpy* ↓

In the tempering process, the liquid chocolate is cooled to 24 °C and held just long enough for a small fraction of the cocoa butter to crystallize. Only a small amount of crystallization is allowed, so the product remains liquid and the crystallization is not controlled, hence again a mixture of polymorphic forms will be present. The product is reheated to about 32 °C to melt all the polymorphic forms except Type V and then the tempered chocolate can be poured into a mold and allowed to cool. Because crystal nuclei preexist in this second cooling cycle, there will be no lag before additional nuclei form; mass crystallization will proceed by growth alone. The polymorphic form that grows is determined by the polymorphic form of the nucleus, so the fat crystals will be exclusively Type V. Type V crystals are denser than the less-stable polymorphic forms, so the product contracts sufficiently on cooling to detach from the mold and gives the expected "snap."

Freshly prepared chocolate is smooth and glossy but on storage can "bloom" and develop a matte, white surface, particularly if stored at high and fluctuating temperatures. This is due to the perfection of Type V crystals to the more stable Type VI. Type VI crystals are characteristically large and on the surface of the chocolate, scatter light differently causing the change in appearance. Type VI crystals do not form during chocolate manufacture as they are slow to nucleate, but they can begin to form over weeks or months of storage.

6.7 Crystallization in Viscous Solutions

Crystallization depends on the movement of molecules. Molecules in the liquid state must diffuse into a configuration to form a nucleus, other molecules must diffuse to (or away from) an existing crystal during growth, and molecules must diffuse between different environments in crystals during crystal perfection. If the viscosity of the liquid phase is high or the temperature is low, then the rate of diffusion is reduced and these processes will occur more slowly (Eq. 2.2). Viscosity is important as a measure of reduced molecular mobility but that relationship is not always robust. A highly concentrated sugar solution (e.g., a hard candy) becomes viscous, because in order for the liquid to flow, the sugar molecules have to move past one another; high viscosity is a good measure of their molecular mobility. A more dilute sugar solution with a small amount of polymer (e.g., a gummy candy) is also viscous, but in this case because of the effect of the polymer molecules. The sucrose molecules can freely diffuse through gaps in the polymer coils and their molecular mobility is not reduced. In general, the concept of reduced molecular mobility as an inhibitor of crystallization is most usefully applied in foods to supersaturated or supercooled carbohydrate and polymer solutions (i.e., dry and/or frozen products). The effects of high solution viscosity are generally not observed in fat crystallization.

The effect of supersaturated solution concentration on the diffusivity of a solute molecule is illustrated schematically in Fig. 6.21a. As
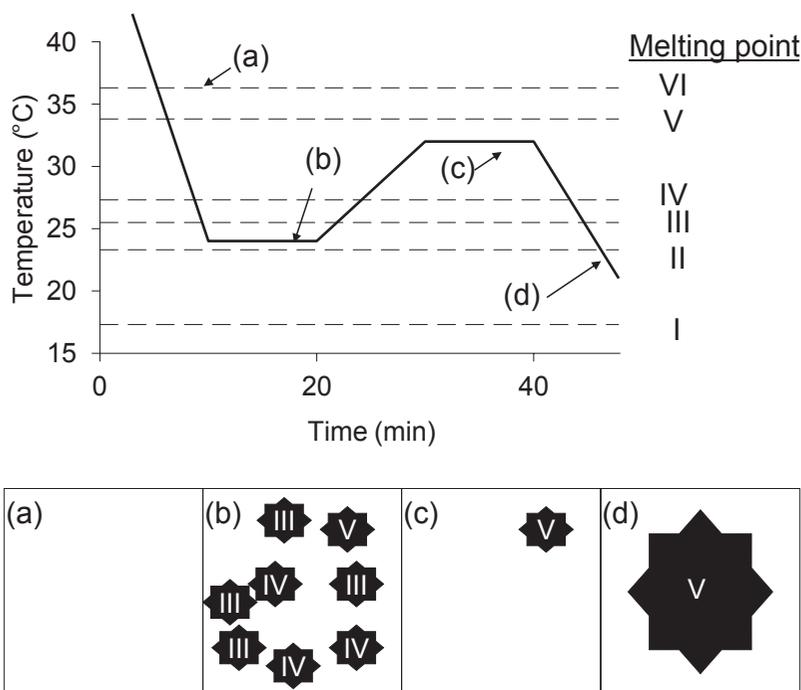
Fig. 6.21 Time temperature profile for chocolate tempering showing the melting points of the various polymorphic forms; alongside is a highly schematic illustration of the types of polymorphic forms present at different stages. **a** The cocoa butter liquid with no crystals present. **b** The liquid is cooled below the melting point of some of the polymorphic forms and a large number of nucleation events form small crystals primarily of the less stable forms. Type VI nucleates very slowly and is not seen. **c** The mixed crystals are reheated so all except Type V melt. **d** Mass crystallization proceeds by the growth of existing Type V nuclei and there is no further nucleation

concentration increases, the marked molecule's movement is increasingly blocked by other solute molecules, viscosity increases, and molecular mobility decreases. At a certain point, it becomes impossible for the marked molecule to move at all—the material becomes solid and molecular diffusion ceases. All of these are metastable as supersaturated solutions because crystal nucleation has not yet occurred. None of them, even the final solid, is crystalline. A crystal has regular spacing between atoms but these materials are all amorphous (i.e., without regular packing). When the molecules are in an amorphous state but unable to diffuse, they are in a glassy state. When the molecules are in an amorphous state but able to diffuse slowly, they are in a rubbery state. Glassy materials (e.g., hard candies, spray dried milk) are characteristically brittle, crunchy solids, whereas rubbery materials (e.g., a chewy caramel) are characteristically sticky and chewy. If a rubbery material is cooled, then the thermal energy of the molecules for diffusion is reduced and it enters the glassy state at a characteristic glass transition temperature (T_g). The glass transition temperature can be reduced by adding small amounts of water (or other miscible small molecule) as a plasticizer, a molecular-level lubricant, to increase molecular mobility.

Increasing solution concentration moves the system further away from the phase line and increases the thermodynamic driving force for crystallization. In theory, we would expect rates of crystal nucleation, growth, and perfection to increase. However, at a certain point, reduced molecular mobility starts to slow the rate of crystallization and eventually goes to zero at the glass transition point (Fig. 6.22b).

a   Concentration ——————▶
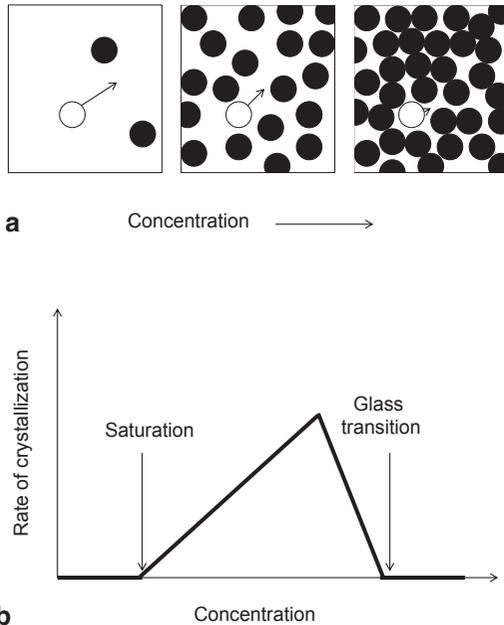
b   Concentration

Fig. 6.22 a Diagram showing the effect of increased solute concentration on the ability of a marked molecule to diffuse. At the highest concentration, the diffusing molecule cannot diffuse and the material is in a glassy state. **b** Effect of solution concentration on the rate of crystallization (adapted from Hartel 2001. Both illustrations are highly schematic)

Example: Cotton Candy

Cotton candy is made by melting sucrose ($T_m = 186\,°C$) and spinning the molten sugar in fine streams from small nozzles by a centrifugal force. The molten sugar cools extremely quickly and can be collected as a solid mat. Cotton candy must be eaten quickly, or packed in moisture-impermeable bags otherwise it will quickly collapse to powder. At a molecular scale, the crystals are melted to form liquid sucrose that is cooled so quickly that it reaches a glassy state before nucleation ($T_g = 60\,°C$). The cotton candy is stable in the glassy state so long as there is no moisture to act as a plasticizer and lower the glass transition temperature to room temperature. If the cotton candy enters the rubbery state, then
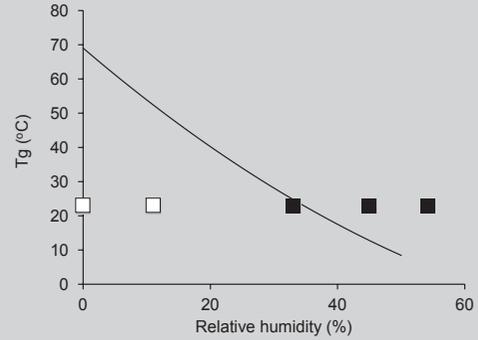
Fig. 6.23 T_g of sucrose as a function of relative humidity. *Points* show the state of cotton candy samples stored for 3 days at each relative humidity (*filled* = unstable, a mixture of crystals and a sticky rubbery residue, *open* = stable, glassy). (Adapted from Labuza and Labuza 2004)

it becomes sticky and the sugar can crystallize. Labuza and Labuza (2004) studied the effect of water activity on the storage stability of cotton candy by incubating samples at 23 °C in a series of controlled humidity chambers. They used literature data to show the glass transition temperature decreased with increased water activity (Fig. 6.23). Samples stored at water activities—so that $T_g < 23\,°C$—tended to become sticky and/or crystallize, while samples stored at lower water activities remained glassy and were stable.

6.8   Summary

Crystallization is a process of phase separation, driven by low temperatures and/or high concentrations of the crystallizing material. The process is conventionally divided into three phases: nucleation (the formation of new crystals from the liquid), growth (the mass deposition of material onto those crystal nuclei), and perfection (change in crystal form to further reduce free energy). The kinetics of these processes affects the microstructure of the crystals and the properties of the food. In most cases, when crystals are eaten as part of a food, they are very fine and dispersed

as part of a noncrystalline phase (e.g., a solid fat crystal network in margarine, fine sugar crystals in a fondant, and fine ice crystals in ice cream). The properties of solid-in-liquid dispersions (i.e., sols) will be discussed in Chap. 8 and their effects on the viscosity of liquids and the formation of solid gels in Sect. 7.7 and Chap. 9, respectively.

6.9 Bibliography

The processes of crystallization are covered in many specialized texts (e.g., Mullen 2001), but Hartel's (2001) "Crystallization in Foods" has the advantages of being explicitly focused on foods, comprehensive, and very readable. Chapters 14 and 15 of Walstra (2003) are also very useful.

The importance of amorphous states and the glass transition in foods is treaded here only very briefly. A good general overview is available in Chapter 4 of Hartel (2001), Chapter 16 of Walstra (2003) and in specialized reviews (e.g., Le Meste et al. 2002; Levine and Slade 1992).

Polymers

John N. Coupland and Rammile Ettelaie

7.1 Introduction

The simple picture of a molecule as a tiny particle that interacts with its neighbors while being buffeted by thermal motion has proved valuable. However when molecules become very large, they gain certain properties distinct from their smaller cousins that make them an interesting group to look at in their own right. "Very large" is intentionally a vague definition but if we take it to mean molecular weights in the tens of thousands and greater, then all of the molecules in this group are polymers. Polymers are formed from the combination of a series of smaller molecules (i.e., monomers) to form a chain. Thus, the fundamental requirement for a molecule to be able to polymerize is that it needs at least two reactive groups. If a molecule has one reactive group, then it can react with a second molecule to form a dimer but this blocks the reactive sites on both molecules and prevents further polymerization. If a molecule with two reactive groups forms a dimer then it blocks one reactive site but still has another available to continue the reaction and lengthen the chain. If a monomer has more than two reactive groups then it can form a branched chain.

J. N. Coupland (✉)
Department of Food Science, Pennylvania State University, University Park, PA, USA
e-mail: jnc3@psu.edu

R. Ettelaie
School of Food Science and Nutrition, University of Leeds, Leeds LS2 9JT, United Kingdom

The important polymers in foods are proteins and polysaccharides. Proteins are of primary importance to life because of their capacity to bind specifically to other materials and act as enzymes or structures. Proteins are consumed as part of whole foods, and used as ingredients to form gels (e.g., gelatin in desserts, soy proteins in tofu) and to stabilize dispersions (e.g., casein in homogenized milk, egg albumin in merengue). In some cases, the biological functionality of enzymes can be co-opted for use in food processing (e.g., corn syrup is manufactured from starch using amylase enzymes). Proteins contribute some energy as part of the diet as well as essential amino acids that we need but cannot synthesize (e.g., lysine and methionine).

Polysaccharides are made by plants, animals, and microorganisms typically either as part of their structures (e.g., pectin, cellulose, alginate) or to store energy (e.g., starch, glycogen). While most are consumed directly as part of whole foods, others are extracted, modified, and used by food technologists to increase the viscosity of liquids (e.g., xanthan gum in salad dressing), form gels (e.g., pectin in jams and jellies), form films (e.g., pullulan in breath freshening strips), or stabilize suspensions (e.g., carageenan in chocolate milk). Polysaccharides are important in the diet as sources of energy (e.g., starch) or components of dietary fiber (e.g., cellulose and most other nonstarch polysaccharides).

The properties of a polymer depend on its composition, but food scientists cannot control the composition of their polymers with the free-
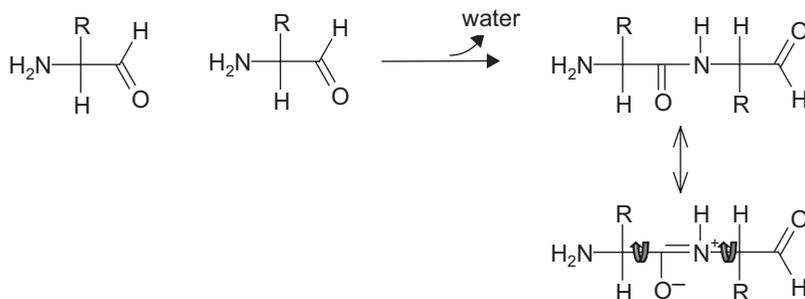
Fig. 7.1 Reaction of two amino acids to form a peptide bond. Note that the product, a dipeptide, still has an available amine group and carboxylic acid group, so it can continue to add more amino acids and eventually form a polymer. *Arrows* show examples of freely rotating bonds in the chain

dom that a synthetic polymer chemist enjoys, and must instead work with the variety provided by nature. In this chapter, we will examine how the structure of a polymer help define its shape, its interactions in solution, and importantly solution viscosity.

7.2 Polymer Chemistry

The properties of a polymer are determined by the sequence of monomers that make up its structure and how they interact with one another and with other components of the solution. Proteins are polymers of amino acids—multifunctional small molecules, each with an amine and a carboxylic acid functional group separated by a single carbon (the α-carbon). The amine group on one amino acid can react with the carboxylic acid group on another to form a dimer linked by a peptide bond (Fig. 7.1). The dipeptide formed still has a free amine and a free carboxylic acid, so it can react with more free amino acids to form larger polypeptides. Note that the repeating structure of the polypeptide is an amino acid residue rather than an amino acid, as it has lost the free amine and carboxylic acid in forming two peptide bonds. The synthesis of proteins *in vivo* is controlled by enzymes that use an RNA code to program the exact number and sequence of amino acids incorporated into each protein. A natural protein is therefore a heteropolymer (i.e., made from a range of different monomers) with

a defined primary structure (i.e., covalent structure, the sequence of amino acids in the chain).

The α-carbon is substituted with 1 of 21 naturally occurring side chains. The amino acid side chain is unchanged during peptide bond formation and contributes to the chemistry of the protein. Many side chains are nonpolar but others are polar. Some have weakly acidic or basic functional groups that gain or lose protons according to their pK and the pH. Proteins are therefore polyelectrolytes (i.e., many ionizable groups on the chain) and carry a net negative charge at high pH, no net charge when the pH is equal to the isoelectric point (pI), and a net positive charge at low pH (see Fig. 2.11).

Polysaccharides are formed by the polymerization of simple sugars. The small molecule sugars themselves are hydrocarbons with several alcohol groups and typically either an aldehyde or a ketone functional group (e.g., glucose and fructose, Fig. 7.2). Some sugars can also contain additional functional groups (e.g., negatively charged sulfate and carboxylic acid groups, positively charged amide groups). Alcohols can react reversibly with carbonyl groups to form a hemiacetal structure. When both the alcohol functional group and the carbonyl functional group are on the same simple sugar, the reaction will bring the molecule into a ring conformation where steric constraints mean that five- (i.e., furanose) and six-membered (i.e., pyranose) form are dominant. Forming the hemiacetal makes the first carbon (C_1) optically active so that there are two isomers of the final ring structure with distinct shapes (i.e., the α and β anomers, Fig. 7.2b).

Fig. 7.2 a Open chain and ring configurations of glucose and fructose. The numbers shown are commonly used to identify specific carbon atoms in the molecule.
b The two different forms of the glucose pyranose ring have very different shapes, in particular note the axial vs. equatorial orientation of the alcohol substituent in first carbon (C_1)
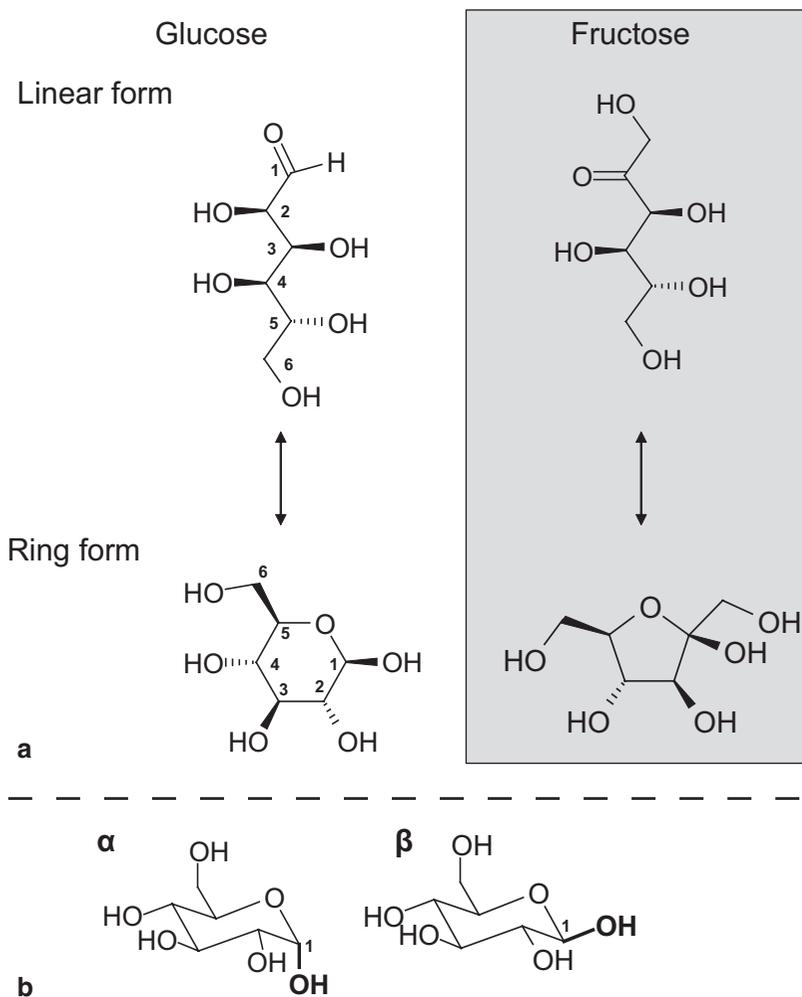

The hemiacetal group in a cyclic sugar can react irreversibly with the alcohol of a second simple sugar, locking the ring closed and bonding the two molecules together as a dimer (Fig. 7.3). The disaccharide has a single reactive hemiacetal so that it can continue to add more monomers and grow to form a polymer. Additionally, each sugar residue has several reactive alcohol groups that can bond several hemiacetals on different sugars and act so as a branch point on the chains.

Polysaccharides tend to be homopoylmers (e.g., cellulose, a linear sequence of glucose residues), with either semi-regular repeating structures (e.g., xanthan gum, another glucose homopolymer but alternate residues have tri-saccharide branches attached) or block copoly-mers—sequences of one type interspersed with sequences of another (e.g., alginate gum contains long sequences of either mannuronic acid, gulu-ronic acid, or alternating mixtures of the two). Unlike proteins with their precisely controlled sequence of amino acids, polysaccharides almost always exist as a distribution of different molecu-lar weights and chemical structures.

The chemical bonds in Fig. 7.1, 7.2, and 7.3 in-dicate the connections between atoms but do not represent the shapes of the molecules. There is a very small energy barrier preventing the rotation of a single bond, yet doing so changes the shape of a molecule completely (Fig. 2.3b). However, not all of the bonds in a protein or polysaccharide chain are so free to rotate. The ring structures of the sugar residues in a polysaccharide cannot be rotated without breaking bonds, leaving only the

Fig. 7.3 Polymerization of glucose molecules to form a dimer as a starting point for polymerization. Whatever its size, the polymer still has one reactive hemiacetal and alcohol groups so it can continue to add more glucose molecules and eventually a polymer. *Arrows* show rotatable bonds
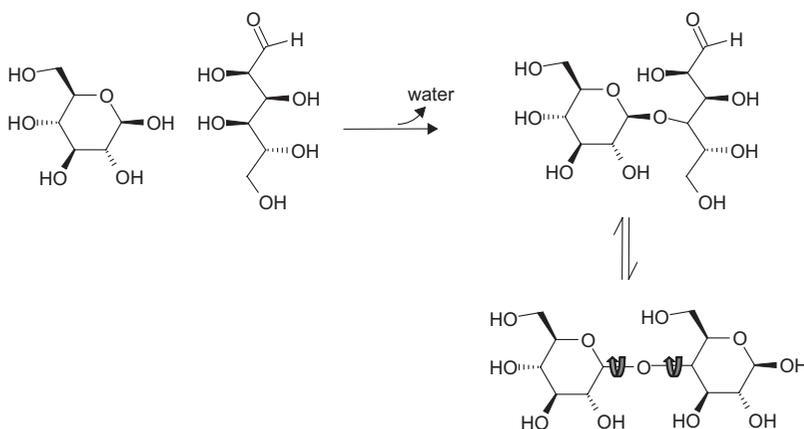
two bonds adjacent to the oxygen in the glycosidic link as rotation points (Fig. 7.3). The peptide bond in a protein chain does not contain rings, but the double bond electrons with the oxygen (the carboxylic acid residue) are partly delocalized into the chain providing the carbon–nitrogen bond with some degree of double-bond character (Fig. 7.1). Double bonds are much more resistant to rotation so, in practice, the only bonds free to rotate on a peptide chain are those adjacent to the α-carbon. Although some of the rotation angles of these bonds will be blocked as they would require two parts of the chain to overlap, the physical picture of a polysaccharide or protein as a series of rigid rods connected by freely rotating joints is more useful than the rigid forms implied by the chemical structures. In the next section, we will explore what flexibility means for the shape of a polymer molecule.

7.3 The Shapes of Polymer Molecules

We can model a polymer as a series of rigid rods (the nonrotatable parts of the structure) connected by freely moving joints (the rotatable bonds). If the bonds are truly freely rotating,[1] then any

Fig. 7.4 Example random coils of a 10-unit model polymer. The *straight-line* structure shown to the left is statistically improbable and in practice is not seen. Any one of the coils is equally as improbable as any other

bond angle is as good as any other and we can draw a shape for the polymer by setting a random value for each. Some examples of short polymers built in this manner are shown in Fig. 7.4. All of them have the same primary structure but all have different shapes and none show any regular pattern (i.e., no straight rods, no helices). It is possible a given polymer could look like on other polymer or have a regular shape but because each bond angle is randomly assigned, it would be staggeringly unlikely (see the discussion around Fig. 1.2d where this argument was used to develop the ideal of entropy). Each individual polymer is not locked into the initial shape generated, but will flex and twist as the individual bonds rotate.

---

[1] In many cases, steric hindrance and other factors mean the chain is much less flexible than suggested here. In these cases, the chain can be described as equivalent to another, ideally flexible, chain made up of a number of Kuhn segments where the length of each segment, the Kuhn length, is greater than one monomer.

Over time a polymer will sample the very large set of possible conformations available. The "shape" of the model polymer molecule is therefore some sort of statistical average of the set of possible conformations allowed.

There is a close analogy between the random coil model for a polymer and the problem of calculating the diffusion of a gas (see discussion in Sect. 2.1). A gas molecule moves in a straight line until it collides with another molecule and bounces off in different direction. This process is known as a random walk because the direction of each step is unaffected by the direction of the last. There is no particular direction to a random walk but on average, the distance from the starting point increases with the square root of the number of steps taken. Our model polymer is also a random walk with the rods representing each step and the joints the random change in direction. By analogy, we can say that the average magnitude of the end-to-end distance of the polymer (r_{rms}) is proportional to the square root of number of segments (n):

$$r_{rms} \sim n^v \qquad (7.1)$$

where $v = 1/2$. This relationship is shown in Fig. 7.5; the average size of the polymer increases ever more slowly with increasing number of segments, and large flexible molecules will tend to spontaneously fold up into a small region of space.

One problem in drawing such a precise analogy between polymer conformation and gas diffusion is that while the gas molecule could cross its own path as often as it liked, the polymer chain cannot pass through any region of space already filled with polymer. Rather than following a truly random walk, the polymer follows a self-avoiding random walk. In practice, this forces the chain to take a slightly more expanded conformation, as more of the steps towards the more densely packed center will be blocked than steps away. The radius of a self-avoiding chain increases with segment number raised to a power greater than the $1/2$ in Eq. 7.1. This effect is illustrated with an exponent of $v = 3/5$ in Fig. 7.5.
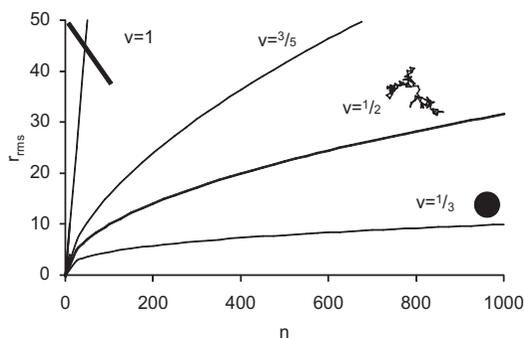
Fig. 7.5 Average end-to-end length of a random coil polymer (r_{rms}) as a function of number of segments (n) and different values of v. *Diagrams* illustrate representative shapes of some of the molecules

The second major problem in comparing polymer conformation to gas diffusion is we neglect the role of chemical interactions between the polymer and solvent. We used the exchange parameter, χ, to describe the effects of molecular interactions on small molecule solubility (Eq. 4.10). The exchange parameter is the enthalpy cost to move one molecule of solute out of its own phase by breaking solute–solute bonds, make a hole in the solvent by breaking solvent–solvent bonds, and inserting the solute into the hole and forming solute–solvent bonds. A positive χ means the energy cost of breaking bonds outweighs the energy released from bonds formed, and would oppose mixing and force the monomers into a pure phase (c.f., oil, and water). In the polymer case, χ is applied to a single segment of the polymer chain equivalent in volume to one solvent molecule. A high χ favors segment–segment and solvent–solvent contacts, so it will tend to force the polymer to fold in on itself to form a smaller volume than predicted by the entropy rule. Solvents for polymers are often described in terms of their quality with a good solvent having $\chi = 0$ and poor solvents have $\chi > 0$.

To form a truly random coil, the polymer–solvent interactions have to be the same as the polymer–polymer interactions ($\chi = 0$, for example, polystyrene in a styrene solvent); then, there would be no enthalpic modification of the basic self-avoiding shape defined by the chain entropy (i.e., $v < 3/5$ in Eq. 7.1 and Fig. 7.5). If the sol-

vent was slightly poorer quality, so that bonding slightly opposes chain expansion enough to balance the expansion due to self-avoidance, then the real polymer would behave as an ideal random walk ($n=0.5$ in Eq. 7.1). This is known as a θ solvent $\chi=0.5$. Similarly, increased temperature favors entropic effects over enthalpic ones (Eq. 1.10, $G=H-TS$) so polymers at higher temperatures will tend to have more expanded shapes (i.e., greater v in Eq. 7.1). The temperature where the expansive effect of temperature overcomes bonding and self-avoidance effects, and causes a real polymer to behave as an ideal chain ($n=0.5$) is called the θ temperature. If a polymer was mixed with a very poor solvent (e.g., polystyrene in water), then we would expect it to coil more tightly, than a random coil, to reduce the polymer–solvent interactions and maximize polymer–polymer interactions (i.e., $v\sim\frac{1}{3}$). If a polymer was mixed with a very good solvent (e.g., a highly charged polymer, a polyelectrolyte, in water), then the coil would stretch out to maximize the polymer–solvent interactions and minimize polymer–polymer interactions (i.e., $v>\frac{1}{2}$). Most food polymers, especially proteins, are somewhat hydrophobic so water is a poor solvent and $v<\frac{1}{2}$.

This is a statistical model of polymer shape without accounting for any specific interactions or details of structure. There is no fixed arrangement of the coil but rather a set of conformations defined by the combination of bond angles and enthalpic interactions that minimize free energy. Any change to that arrangement will increase the free energy and as a consequence:

- Polymer coils in solution will tend to repel one another, as their overlap would increase the local chain density and create an osmotic pressure gradient with the remaining solution (Fig. 7.6a)
- Polymer coils will resist deformation elastically, as stretching them would reduce the randomness of the bond angles and change the balance of polymer–polymer and polymer–solvent interactions (Fig. 7.6b)

The picture of a random coil as the preferred structure of a polymer in solution is a helpful step forward from the rigid bonds implied by the
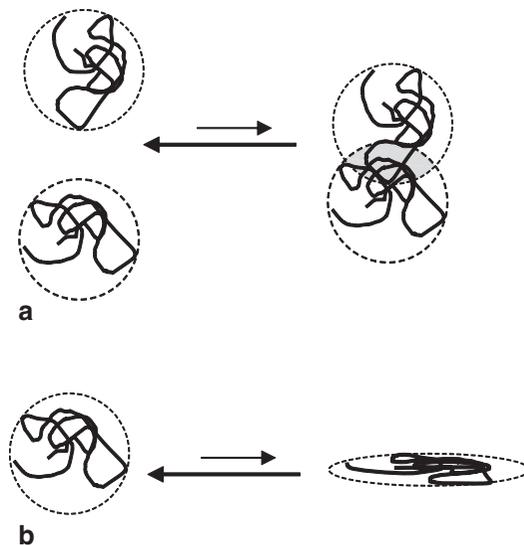
Fig. 7.6 Polymer coils will **a** repel one another and **b** resist deformation

covalent structure diagrams. However, real biopolymers have well-defined structures that allow them to perform their biological function and in the next section we must modify our picture to understand biological functionality.

7.4 The Shapes of Protein Molecules

Figure 7.7 shows some representations of the structure of lysozyme, a small (129 amino acid residues) protein found in egg white. To generate the images, the protein was first carefully purified and crystallized from the solution, then precise X-ray scattering experiments were used to calculate the positions of all the atoms in the molecule. These coordinates were uploaded to the Protein Data Bank (www.pdb.org) from where they, along with tens of thousands of other protein structures, can be freely downloaded, manipulated, and displayed. The lysozyme backbone is about 50 nm long but packs into a dense near-spherical globule a few nanometers in diameter (Fig. 7.7a); in some ways typical of a random coil polymer in a poor solvent. However, there are some important differences suggesting the coil is not genuinely random. Firstly, most of the hydrophilic amino acids are at the surface (Fig. 7.7a)
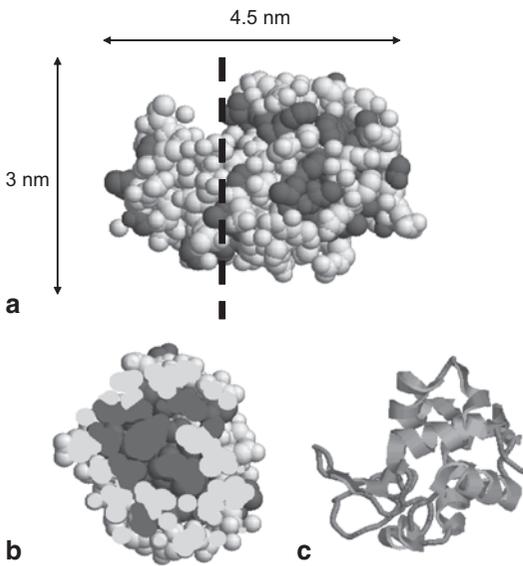
Fig. 7.7 Structure of lysozyme, a small globular protein found in egg white. **a** Space-filling structure of the protein surface with the hydrophobic amino acids shown with a darker shade than the polar amino acids. **b** The *dashed line* shows the position of cross section. **c** A cartoon representation of the polypeptide backbone showing helices and sheets. All are representations of an X-ray determination of the molecular structure developed by Naresh et al. (2007)

but the core of the molecule contains the hydrophobic residues and excludes water molecules (Fig. 7.7b). Secondly, the peptide backbone contains regular coils and helices that would not be seen in a random structure (Fig. 7.7c). Finally, the very fact that lysozyme could be crystallized from solution in the first place suggests that all the molecules must have the same shape. Indeed, at the time of writing, the Protein Data Bank contained over 130 structures of hen lysozyme, isolated and characterized by workers from around the world, and with some minor differences, all the molecules are the same. A protein is a coil but not a random one.

On reflection, this is not surprising. The hen depends on the protein to perform a very specific function—in this case to act as an antimicrobial agent and to protect the developing chick embryo from infection. Random coils could not be expected to function so specifically. The primary structure (the sequence of amino acids) is programmed by the genetic code of the organism,

but the folding and twisting of the chain to form the characteristic shape necessary for function (the tertiary structure) is a spontaneous process. Out of all the billions of potential configurations available to a flexible polymer, the freshly synthesized polypeptide reliably and rapidly finds the unique native configuration. We can get a sense of how this happens by thinking about a polypeptide chain in terms of the generic random coil model for polymer shape introduced above.

In general, a polymer will tend to minimize its free energy by taking a random coil configuration modified by the interactions of its segment with the solvent. Many of the functional R-groups on amino acid residues are hydrophobic so they have a large χ in the aqueous cell cytoplasm and would tend to fold into a dense ball to minimize contact with water (i.e., $v = {}^1/_3$ in Eq. 7.1).[2] Polar amino acids have a much lower χ in water and favor a more expanded random coil configuration, allowing more water-chain contacts and maximizing the chain entropy. The native configuration of the molecule balances the "needs" of its different residues by folding itself into form a dense, dry core filled with hydrophobic amino acids while allowing the more hydrophilic amino acid residues to remain at the surface. This simple argument of the conflict between chain entropy favoring a loosely packed random coil and the hydrophobic effect favoring a densely packed sphere provides many of the gross features of protein folding, but other noncovalent bonds (e.g., hydrogen bonding, charge–charge interactions) are important in supporting and refining the exact configurations of the chain. Their role is clear in the small set of repeating structures seen over part of the chain in a range of different proteins (secondary structures). For example the helices in lysozyme (Fig. 7.7c) are α-helices supported by hydrogen bonds between peptide –NH groups and peptide C=O groups four residues

---

[2] Note that χ is large because of the hydrophobic effect "repelling" water and nonpolar molecules. As we saw in Chap. 2, the hydrophobic effect is largely due to entropy changes resulting from ordering of water molecules. We have talked about χ as a purely enthalpy term but it is quite straightforward to treat it as having an entropic component as well.

down the chain. A small set of covalent bonds can also play a role supporting the configuration of a protein. In particular, the thiol groups of two cysteine residues on the chain can be oxidized to form the cysteine dimer, a disulfide bond holding two parts of the same chain in close proximity.

This gross description of lysozyme's shape—a rough sphere with a hydrophilic surface and a hydrophobic core—is characteristic of many proteins, especially enzymes. These globular proteins have reasonable solubility in water because of their hydrophilic surface. Under many conditions, intact globular proteins behave as nanoscale particles suspended in water, and in some cases it is more helpful to understand their behavior as that of a dispersion (see Chap. 8) rather than as expanded polymers using the approaches discussed here. Other proteins however, are clearly nonspherical—many structural proteins have an extended fibrous shape. In order to maintain a nonspherical shape with a low chain entropy, fibrous proteins often exist as bundles of polypeptides with a strong repeating secondary structure reinforced with covalent links. For example collagen, found in muscle and connective tissue and used as the raw material in gelatin manufacture, has a characteristic triple helical structure[3] formed from three polypeptide chains reinforced with intermolecular covalent bonds. The controlled association of several polypeptide chains under physiological conditions is known as quaternary structure. Having noted some of the diversity of structure we will continue to use globular proteins such as lysozyme as the example in further discussion (unless otherwise stated).

In summary, the native state of a protein—the biologically active form taken under physiological conditions—is due to the specific sequence of amino acid residues (i.e., primary structure), folding both locally (i.e., secondary structure) and overall (i.e., tertiary structure), and possibly forming multiunit aggregates (i.e., quaternary

structure). The number of configurations of the polypeptide is drastically reduced by forming a defined structure (i.e., the entropy of the chain is low). However the hydrophobic effect, supported by other covalent and noncovalent interactions, is sufficient to compensate and make the native state stable, but only barely. When the conditions change, the magnitude of the entropic and enthalpic factors contributing to stability change. If the change is enough, then the protein will denature and seek a new configuration appropriate for the new conditions. For example:

- Proteins can be denatured by changing the solvent. Under physiological conditions, the main stabilizing factor for the folded chain is to avoid contact between the hydrophobic amino acids and water, but in a less polar solvent the χ for the nonpolar residues is reduced and the protein will unfold. In a completely nonpolar solvent, the protein may eventually seek to fold with its polar residues buried in the core, away from contact with the solvent.

- Proteins can be denatured by a surface (e.g., whipping egg white to form a merengue). The folded protein has a configuration optimized for an aqueous solution with most of the hydrophobic residues buried in the core. When the protein adsorbs at a surface (see Fig. 5.7), it can partly unfold so that some of the hydrophobic residues are away from water in the air phase.

- Proteins can be denatured by the changes in pH (e.g., the "cooking" of fish in lemon juice to make ceviche). A protein is a polyelectrolyte with a charge dependent on the state of ionization of the various side chains. If the pH is very high or very low, the charges will be predominantly negative or positive. Like charges on the chain repel one another and tend to "stretch" the protein into a denatured configuration. In other cases, the denaturation of a protein may depend on an aggregation process and highly species (i.e., pH far from pI) will be less prone to aggregation.

- Proteins heated above a critical temperature will denature (e.g., cooking an egg white, "Kinetics of milk protein denaturation" in Chap. 3). Consider the free energy difference between

---

[3] The structural requirements for polymer helix formation will be discussed below in the context of polysaccharides. In anticipation of this, it is worth noting that collagen has an unusually simple and repetitive primary structure.

the native state and a hypothetical denatured state, the free energy change for denaturation: $\Delta G_d = \Delta H_d - T\Delta S_d$. The major factor favoring destabilization is the increase in the number of possible configurations of the polypeptide (i.e., chain entropy) upon denaturation. Increases in temperature increases the importance of the difference in chain entropy to the free energy of denaturation and therefore, destabilization is favored. However, the major factor opposing destabilization, the hydrophobic effect, is also largely entropic so its importance increases with temperature. For small changes in temperature, these effects cancel one another out and the protein remains stable. However, beyond a critical point, changes in the structure of water mean the increase in the stabilizing effect of hydrophobicity can no longer keep pace with the chain entropy. ΔG_d becomes negative and the protein is denatured.

While there is one native configuration, there are many possible denatured configurations. Denatured proteins are typically somewhat more expanded, with more hydrophobic and other reactive residues exposed at the surface. If the denaturing factors were removed, then the protein would be expected to return to its native configuration (i.e., regeneration). In dilute solution, simple proteins will readily regenerate but in foods, denatured proteins will aggregate first or otherwise never be able to recover. For example during cooking, egg white turns into an opaque gel but does not return to a clear viscous liquid on cooling.

Example: Inactivation of an Antimicrobial Protein

Lysozyme can be used as an antimicrobial additive in foods and drinks, but its functionality is lost if it is denatured. Makki and Durance (1996) were interested how thermal processing of a fluid food in combination with other solution conditions (e.g., different pH values, salt, and sugar concentrations) could lead to the denaturation of lysozyme. After various thermal treatments, they determined the amount of residual active lysozyme by adding an aliquot to a suspension of *Micrococcus*
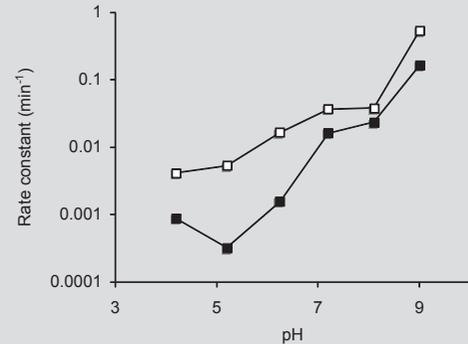
Fig. 7.8 Rate constant for the thermal inactivation of lysozyme as a function of pH at 75 °C (filled points) and 90 °C (open points). (Adapted from Makki and Durrance (1996))

lysodiekiticus and measuring the loss in turbidity as the protein lysed the cells. The protein activity decreased logarithmically with thermal treatment time, suggesting the inactivation of lysozyme followed apparently first order kinetics.

The rate of inactivation increased with temperature in an Arhennius model but also decreased as pH was decreased (Fig. 7.8). The pI of lysozyme is high (=10.7), so the greater the net positive charge on the protein, the more stable it was to heat. Makki and Durrance distinguished the process of protein denaturation, a reversible unfolding of polypeptide structure, from protein aggregation, an irreversible step leading, that is:

Protein ↔ Denatured protein → Aggregated protein.

Differential scanning calorimeter (DSC) analysis showed an endothermic peak corresponding to lysozyme denaturation at 71 °C (pH=7) but presumably much of this denatured protein regenerated as the samples were cooled before they were used in the bacteriological assay. The assay used by Makki and Durrance was a measure of the activity of the lysozyme and did not account for any regeneration. Presumably the rate-limiting step for inactivation was the aggregation of the denatured proteins, which was slower at low pH when they were highly charged.

7.5 The Shape of Polysaccharide Molecules

Sugar residues are typically more hydrophilic than amino acid residues, so a polysaccharide in aqueous solution more closely approximates an open coil configuration than does a protein in aqueous solution. If the solvent quality decreases (e.g., reducing water activity, adding salts to reduce the range of repulsions between charged residues, lowering the temperature), then the coil will begin to collapse, but the nature of coil collapse is quite different for polysaccharides than for proteins. A protein depends on the different χ values for each type of amino acid residue in the programmed sequence to allow it to collapse into a defined native structure under physiological conditions. Polysaccharides have simpler and more repetitive primary structures which will tend to maximize the polymer–polymer interactions in response to a high χ by forming repeating helical structures.

It should come as no surprise to see that polysaccharides, with their simple repetitive structures, tend to form helices. Helices, are always built from a single type of repeating unit. As an illustration, Fig. 7.9 shows an object built from the same structural unit stacked again and again. If the building blocks and the connection between units are fixed, then a helix inevitably emerges. Any more complex shape would need multiple types of building blocks or links similarly the simple repeating structure of a polysaccharide makes the helix the dominant shape seen. Each new residue lengthens the chain and its preferred orientation to the previous residue forces the growing chain to corkscrew. We will explore some of the consequences of helix formation by looking at two chemically similar, but functionally very different, polymers of glucose—amylose and cellulose (Fig. 7.10).

Both amylose and cellulose are linear chains of glucose residues linked by glycosidic bonds between the carbons adjacent to the oxygen (#1) and the carbons (#4) on the opposite side of the molecules (i.e., 1–4 links). The difference between amylose and cellulose arises from the fact

Fig. 7.9 A *spiral staircase* is a helix built by stacking the same piece over and over again with the same rotation between units. Polymer helices are characteristic of simple, repeating structures

that the rings are different anomeric forms (α and β respectively, Fig. 7.2b). In solution, the monomeric sugar cycle through both configurations rapidly, but once the acetal forms, the conformation is fixed and can have a persistent effect on the properties of the polymer formed. In amylose, the substituent on carbon # 1 is equatorial to the ring while the substituent on carbon # 4 is axial. The arrangement of bonds α 1–4 link twists the polysaccharide to form a tightly coiled helix supported by intramolecular hydrogen bonds along the axis. In cellulose, both substituents are equatorial to the ring so the β 1–4 link is flat relative to the planes of the rings. In cellulose, the polysaccharide forms a much flatter structure.

There is an analogy between the shape of a polysaccharide and the shape of an unsaturated small molecule; the pyranose ring, like the un-

Fig. 7.10 Structures of dimers of glucose as occurring in amylose and cellulose alongside structures of different cis and trans bond orientations in unsaturated hydrocarbons. The *double* bonds are analogous to the nonrotating ring structures and the cis/trans configurations of the *double* bond correspond to the axial–equatorial/equatorial–equatorial orientation of the glycosidic bond
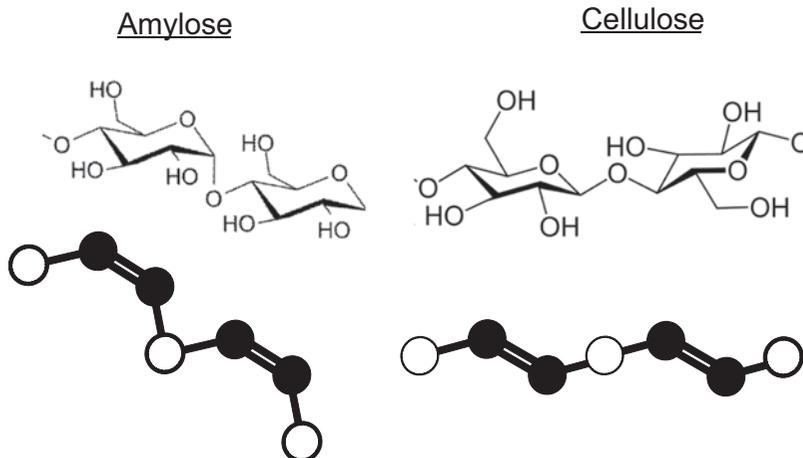
saturated bond is inflexible and cannot rotate. Just as the substituents on a double bond can be *cis* or *trans* (Fig. 2.3), the acetal bonds can be *cis* or *trans* to the nonrotating ring. The α-link in amylose is *cis* to the rings while the β-link in cellulose is *trans*. A *trans* bond puts a "kink" into the chain while the *cis* bond allows the chain to lie flat.

Polysaccharide helices can pack together to form larger structures to further reduce the number of polymer-solvent interactions. The "*cis*" type helices often interpenetrate with others to form double (e.g., amylose) or triple (e.g., xanthan gum) helices while the flatter "*trans*" type helices pack together in fibrous structures supported by intermolecular hydrogen bonds. Supramolecular associations of helices are crystalline in nature and share many features with the small molecule crystals described in Chap. 5, but there are some important differences.[4] Notably, polymers tend to be only semi-crystalline, with the repeating helical structure extending over part of their lengths. The

size of the polymer crystals can be limited by irregularities in the primary structure (e.g., sugar residues that cannot fit into the repeating pattern, branch points, chain ends) and also by kinetic constraints. Frequently, polymer solutions become highly viscous as they crystallize and the molecules do not have the mobility to align properly in the time available (see the discussion of the glass transition in Sect. 6.7, high molecular weight polymers have low glass transition temperatures so easily become immobilized before they can fully crystallize). The partly crystalline, partly random coil configuration of polymers in solution provides the structural basis for amylose gelation while the dense fibers of cellulose are important in providing the mechanical strength of plant cell walls.

7.6   Polymer Solutions

Our discussion regarding shapes of the different food polymers was based on forming some sort of coil, to maximize the entropy of the rotatable bonds, modified by the intermolecular interactions between polymer segments and the solvent. We can develop a more quantitative model of polymer–solvent interactions in a solution using a lattice model (Fig. 7.11). We used this approach in Chap. 4 for small molecule solutions, and calculated the entropy and enthalpy of mixing by randomly placing black "solute molecules" onto

---

[4] Be careful here—polymer crystallization was used to calculate the protein structure in Fig. 7.7. In that case, a crystal was made by packing many polymer molecules together with each molecule serving as a unit of the crystal lattice. Protein crystals can be grown in the laboratory but it takes a lot of patience and special conditions. In polysaccharide crystallization, the crystal structure occurs over part of the chain length and can occur spontaneously in food and in nature.

Fig. 7.11 Lattice model for the mixing of a small molecule solvent (*open circles*) with polymers (*filled points*). There is one unmixed configuration (*left*) and very many mixed configurations of which this is an example (*right*). However the number of mixed configurations is much less than the corresponding mixture of small molecules. (c.f. Fig. 4.14)

a lattice and filling in the gaps with white "solvent molecules." In a polymer solution, the solute molecules are much larger than the solvent molecules so we represent the polymer as a chain of monomers (strictly a unit of the chain of equal volume as the solvent molecule) adjacent to one another on the lattice. Placing the first monomer on the lattice is the same as placing a small molecule on the lattice; if there are N sites then there are N possibilities. However, for the second small molecule on the lattice, there are $(N-1)$ options but because the polymer must be in a continuous chain for the second monomer in the chain, there are only five options (i.e., coordination number -1). The number of possible mixed configurations determines the entropy of mixing so the entropy of mixing for a polymer into solvent is much lower than for a corresponding small molecule solution. For large polymers, mixing entropy is negligible and the free energy of mixing is determined solely by enthalpic interactions.

The mathematics for calculating the free energy of mixing a solvent and a polymer can be approached in a similar way to the small molecule case (i.e., the Hildebrand model, Sect. 4.6 and clearly described by Dill and Bromberg (2003)) but here it suffices to simply state the major result the Flory–Huggins equation (the corresponding equation for the small molecule case is shown below for comparison):

$$\text{Polymers}: \frac{\Delta F_{mix}}{RT} = \chi_{BW}\phi_B\phi_W + \frac{\phi_B}{M_B}\ln\phi_B + \frac{\phi_w}{1}\ln\phi_w \tag{7.2}$$

Small molecules:

$$\frac{\Delta F_{mix}}{RT} = \chi_{BW}x_Bx_W + x_B\ln x_B + x_w\ln x_w \tag{7.3}$$

In these equations, the subscript W refers to the white "solvent molecules" each filling a single lattice space and the subscript B refer to the solute molecules either a small molecule or a polymer of M monomers where each monomer is the size of a single solvent molecule and occupies one space. The fractions of each component are expressed in terms of volume fraction (ϕ) for polymers and mole fraction (x) for small molecules. The other parameters are the same as used in Chap. 4: RT is the thermal energy, and χ is the exchange parameter (expressed in terms of the interactions between monomer units of the polymer and solvent). The first term in the equations describes the enthalpic interactions, the second term is the contribution of the polymer to the entropy of mixing, and the third is the contribution of the solvent to the entropy of mixing. The contribution of the polymer to the entropy of mixing decreases as polymer molecular weight increases, and becomes negligible for high polymers.

Figure 7.12 shows the calculated free energy of a short polymer as a function of solution composition. While the small molecule phase diagrams in Chap. 4 (e.g., Fig. 4.16) were symmetrical about the midpoint, the size difference between the molecules shows that the polymer phase diagrams are highly asymmetric. Rela-

Fig. 7.12 Calculated polymer + solvent phase diagram ($M = 6$, $\chi = 1.5$). The compositions sharing a common tangent (i.e., same chemical potential) are the phase boundaries. The *shaded areas* are one-phase regions of the phase diagram (*inset* shows a zoom close to the pure solvent axis)
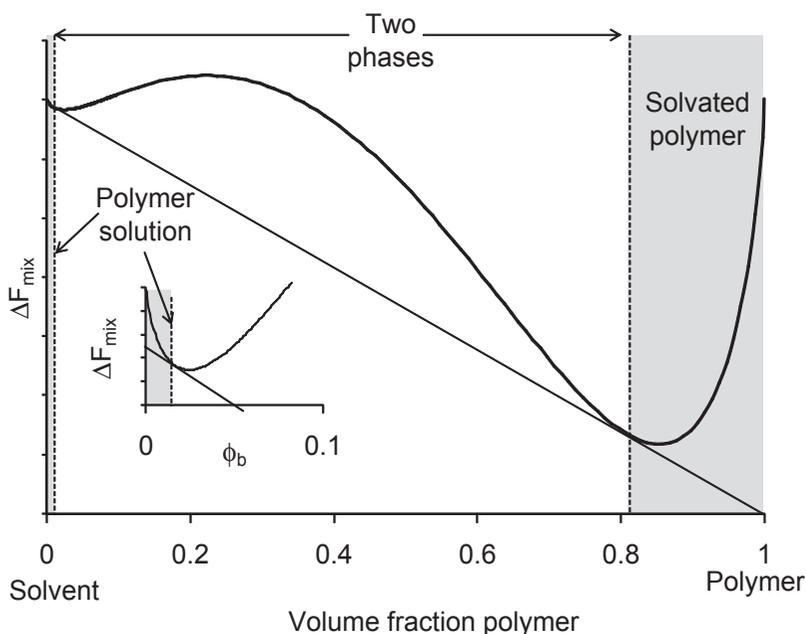
tively, small amounts of polymer will dissolve in solvent, while a dry polymer can absorb significant amounts of solvent. One way of thinking about this asymmetry is that moving a polymer molecule from a pure polymer phase to a solvent phase is a relatively small entropy gain. On the other hand, moving a solvent molecule from a pure solvent phase into a polymer phase causes a large increase in entropy. In binary mixture of polymers (i.e., replacing the third term in Eq. 7.2 with a term in molecular weight similar to the second term), only the enthalpy plays a role. As a consequence, polymer blends almost always phase separate.

It is possible to extend this approach to mixtures of two polymers and a solvent. Here, there needs to be an additional two exchange parameter terms for the enthalpy of Polymer 2–solvent interactions and Polymer 1–Polymer 2 interactions. In practice, however, most polymer–polymer–solvent ternary blends phase separate into two phases with different compositions (see Fig. 4.8 as an example). There are two main types of polymer phase separation (Fig. 7.13):

- Associative, if there are strong Polymer 1–Polymer 2 interactions leading to the forma-

tion of complexes which can either precipitate or form a stable suspension of fine particles. This is commonly seen in mixtures of an anionic polysaccharide with a protein below its isoelectric point.

- Segregative, if the Polymer 1–Polymer 2 contacts are not preferred and the mixture separates into regions enriched in one type of polymer and depleted in the other. One way to understand this process is to imagine slowly concentrating a mixed polymer solution. In the dilute case each coil has enough solvent to take on its optimum configuration without overlapping with any other. As concentration increases, the coils start to overlap. While any overlap is unfavorable, each polymer would "prefer" to be surrounded by more segments of similar polymer rather than dissimilar polymer so they tend to concentrate together. The phase-separated regions commonly form some sort of opaque water-in-water emulsion with one polymer phase dispersed in the other, but these can often be separated by centrifugation if they do not gel (Fig. 7.14). Segregative phase separation can sometimes be coupled to
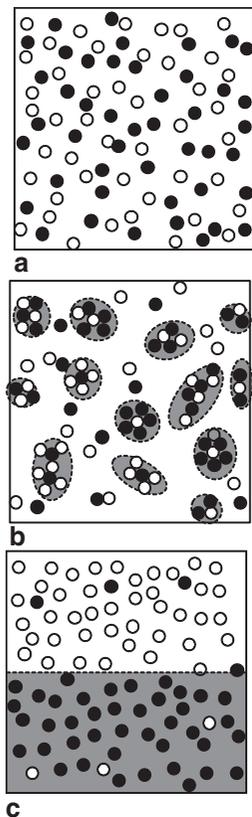
Fig. 7.13 Blends of two types of polymer (represented by *black and white circles*) in solvent. **a** Fully miscible, **b** associative phase separation, and **c** segregative phase separation. Differently shaded regions show separate phases

Fig. 7.14 Segregative phase separation of two polymers in solvent. The *different shadings* represent different compositions. **a** Fully phase separated, **b** Polymer 1-rich phase dispersed in Polymer 2-rich phase, and **c** Polymer 2-rich phase dispersed in Polymer 1-rich phase

polymer gelation of one or both of the polymer phases (Chap. 9) to form a textured solid.

The texture and stability of phase-separated polymer mixtures are often understood in terms of dispersions using the approaches outlined in the following chapter. Associative phase separation leads to the formation of fine particles that may sediment, gel, or aggregate depending on their interactions with one another and gravity. Segregative phase separation gives rise to water-in-water emulsions that behave like oil in water emulsions, but with much lower interfacial tensions (in the order of a couple of mNm^{-1}). In this chapter, we will confine ourselves to the properties of polymer solutions, particularly their effect on viscosity.

Example: Structured Gels

B-type gelatin is derived from animal collagen by vigorous alkali extraction. It is a highly disordered protein that is soluble at high temperatures but tends to progressively form intermolecular triple helices that lead to its gelation at lower temperatures. Pectin is a polysaccharide extracted from plants and can be used to form a gel in the presence of calcium ions. The major monomer of pectin is galacturonic acid (pK=3.5) and, as the pI of B-type gelatin is about 4.9, at pH 5.5 both polymers are negatively charged and there is no strong
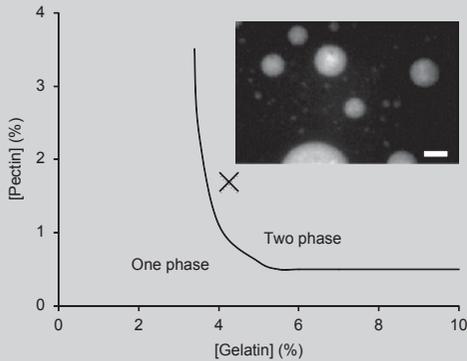
Fig. 7.15 Pectin–gelatin phase diagram. *Inset image* shows phases of labeled pectin in a continuous (unlabeled) gelatin gel formed from the composition marked on the phase diagram (scale bar = 10 μm). (Adapted from Nordmark and Ziegler (2000))

electrostatic attraction that might lead to associative phase separation. Instead, blends of gelatin and pectin solution tend to phase separate segregatively above a critical concentration. Nordmark and Ziegler (2000) carefully purified gelatin and pectin, then mixed their solutions and separated the two phases by centrifugation. They then measured the pectin and gelatin content of each of the separated phases and used the values to create a phase diagram (Fig. 7.15). If the mixtures were stored hot, the liquid phases would eventually separate under gravity but if they are cooled, the gelatin will solidify, trapping "blobs" of pectin-rich phase inside the gel. Depending on the rate of solidification compared to the rate of phase separation, different microstructures can be "locked in" to the gel. Nordmark and Ziegler used covalently attached fluorescent labels to image the intact gel microscopically (Fig. 7.15 inset). The polymer blend phase separated to form large droplets of pectin-rich phase inside a continuous gelatin-rich phase—a water-in-water emulsion. If the mixture were given longer to phase separate in the liquid state before cooling to form the gel, the particles would have time to coalesce and grow

larger. Determination of the polymer composition of the different phases from the brightness of the images was in fair agreement with the phase diagram determined by chemical analysis of the separated phases.

7.7 Defining Viscosity

Even a small amount of dissolved polymer dramatically increases the viscosity of a solution. For example, 0.1 % xanthan gum can be added to low-fat salad dressings to make them thick enough to stay on a lettuce leaf whereas even 60–70 % of dissolved sucrose would not be so viscous. To understand why, we need to first define what we mean by viscosity.

Flow occurs in a variety of ways as liquid ingredients are combined, processed, and eaten (e.g., stirring, pumping down a pipe, swallowing), but in all cases a force is applied and, as a result, the liquid moves. For our purposes, it is better to keep the flow as simple and controlled as possible, and we will do this by examining the movement of a liquid in a concentric cylinder viscometer (Fig. 7.16). The sample is poured into the outer cylinder and a second solid cylinder is lowered down the axis of the first, trapping fluid in the gap between them. Next, a force is applied to the inner cylinder causing it to rotate; the greater the force applied, the faster is the rate of rotation. We are applying a force and measuring the response of the fluid.

We can better picture the movement of the liquid inside the viscometer by mentally "unrolling" the cylinders along their axes. Now, rather than one tube rotating inside another, we have one plate moving parallel to a second stationary plate with the sample in the gap between them. The liquid closest to the moving plate will move at the speed of that plate while the liquid closest to the stationary plate will be stationary. This condition is known as the no slip boundary condition and holds true for many liquids, though not all. We can divide the intervening fluid into infinitesimally thin layers moving at intermediate speeds
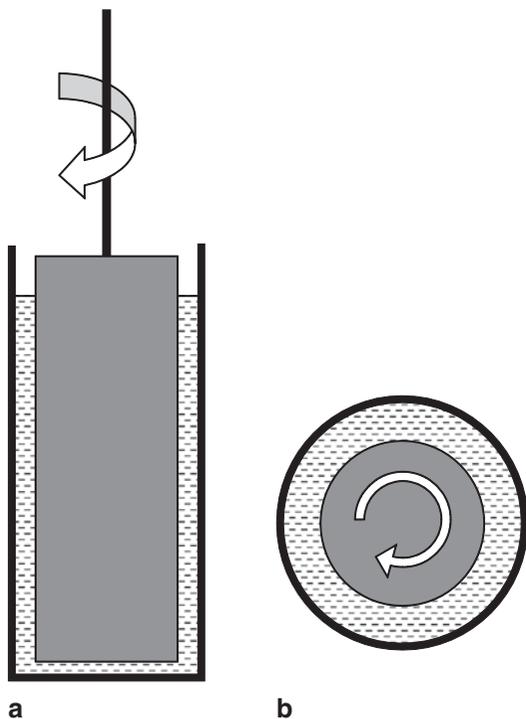
Fig. 7.16 Schematic illustration of the design of a concentric cylinder viscometer showing **a** a cross-sectional and **b** top-down view. The sample is trapped in the narrow gap between two concentric cylinders. A force is applied to the inner cylinder causing it to rotate at a speed that depends on the geometry of the rheometer and the viscosity of the fluid under investigation

(Fig. 7.17). The resistance to flow is represented by the molecular friction between the layers of fluid moving against one another. (A partial analogy would be to press a stack of cards between your hands then try to slide one palm over the other. The cards slip past one another and the resistance felt depends on the friction between the cards; waxed playing cards slide more easily than rougher index cards). Frictional forces increase with the speed of movement, so the liquid will flow at the speed where the frictional losses match the force applied to cause the flow. Applying a greater force will cause the inner cylinder to spin more rapidly.

We want to use the rheometer to measure the properties of the fluid, so we need to free our measurement from the influences of the instrument design. First to rotate cylinders that are twice as long will require twice the force, as
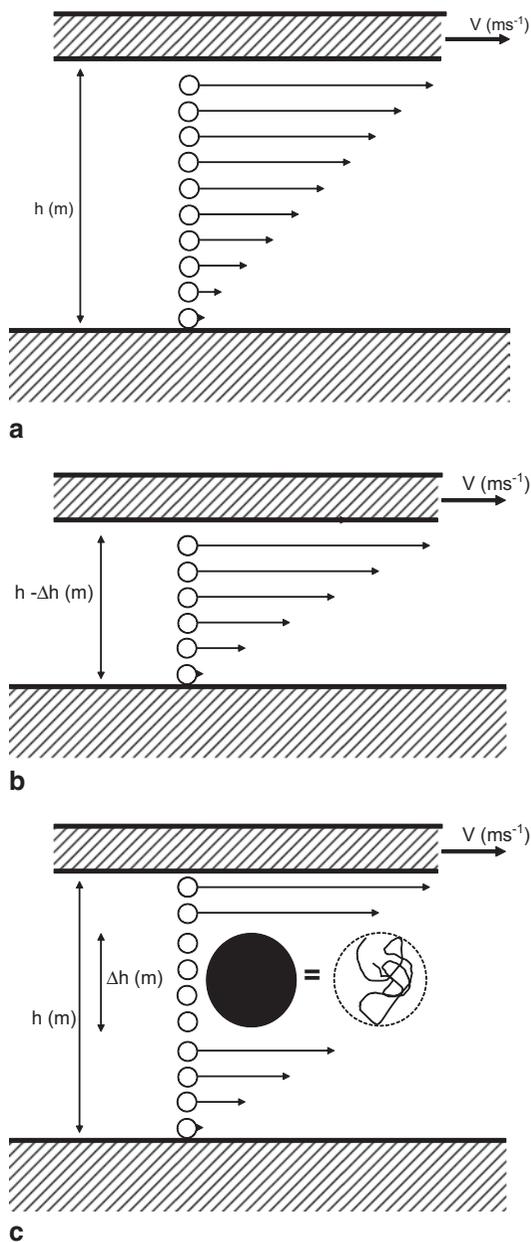
Fig. 7.17 Schematic diagram showing the motion of elements of a fluid trapped between a moving and a stationary plate. **a** In a simple fluid, an applied force F (Nm^{-2}) generates a velocity gradient of V/h (s^{-1}). **b** If the gap is narrower, a greater force is needed to generate the same rate of speed in the moving plate as the velocity gradient is greater. **c** The presence of a particle (not to scale) in the fluid stream stops some fluid elements and effectively lowers the gap size in a similar manner. The apparent viscosity of the fluid is greater because more force is needed to generate apparently the same velocity gradient across the sample. The "particle" in **c** can be a model of a polymer coil as shown

there is twice the drag from the liquid. Similarly, larger diameter cylinders will have greater drag again due to their larger surface area. The same liquid measured in a larger rheometer will therefore need more force to get the same rotation rate. We can correct this effect by normalizing applied force to the surface area of the cylinders; so the applied force F is expressed in units of force per unit area (Nm^{-2} or Pa). Second, if the gap between the cylinders were narrower, there will be fewer layers of fluid present. To get the same rate of rotation of the inner cylinder, they will have to flow past one another faster, generating greater frictional drag and requiring more applied force (Fig. 7.17b). This agrees with everyday experience—it takes a much bigger pump to get the same amount of fluid to flow down a narrow pipe than a wide one. We can account for this by reporting the velocity gradient achieved in the fluid rather than the rate of rotation of the inner cylinder. If the same rotational force (per unit area) is applied to the same fluid in two viscometers with different sized gaps, then there would be a different rate of rotation in each case, but the velocity gradient (i.e., rotation rate divided by the gap size) will be the same.

Having normalized the forces and velocities in this way, we are left with the relationship between applied force per unit area (F, Pa) and velocity gradient (dV/dh, s^{-1}) in the fluid. Fluids that require more force to get the same velocity gradient are more viscous, so we can define viscosity (η, Pa.s) as the proportionality constant (Fig. 7.18):

$$\eta \frac{dV}{dh} = F \tag{7.4}$$

Viscosity is a measure of the capacity of a fluid to resist flow. In some cases (i.e., Newtonian liquids) viscosity does not change as a result of flow but in other cases it may increase or decrease (see below). By defining our terms carefully, we have been able to move from using the word viscosity in a general way to describe our everyday experience to a formal definition obtained from a fundamental measurement. We have entered the field of rheology—the study of the response of solids and liquids to applied forces.
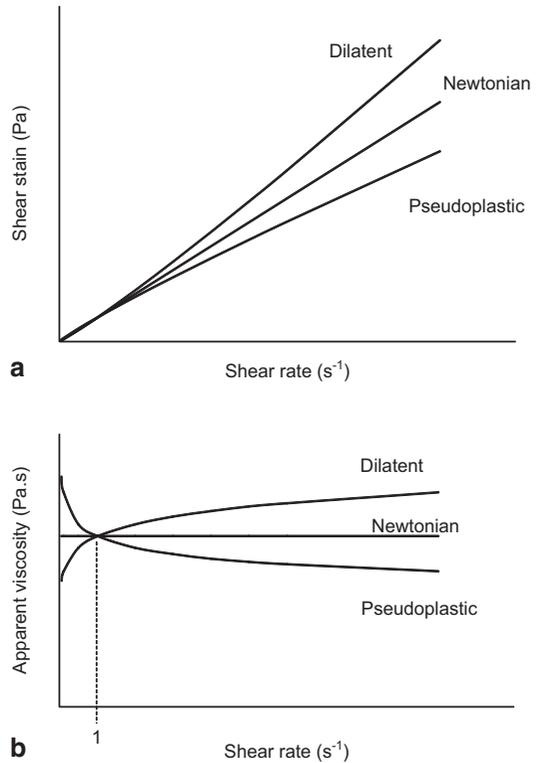
Fig. 7.18 Relationship between **a** stress and rate of strain and **b** viscosity and rate of strain for Newtonian and non-Newtonian fluids

To take a (slightly) more formal look at this problem in terms of rheology, we need to imagine a small volume of the fluid within the viscometer (Fig. 7.19). A force is applied along the plane of one of the surfaces to generate a small deformation. The force applied per unit area is the stress (Pa or Nm^{-2}). This type of sliding motion is known as shear deformation, but it is easy to imagine other types of motion (e.g., stretching, compression) induced by forces acting in other directions. The extent of deformation is described as tan θ or dx/h, and this is defined as being the strain (a dimensionless quantity). The rate of change of deformation with time is the rate of strain (or velocity gradient in terms of our earlier discussion). In a Newtonian liquid, shear stress (τ) is proportional to the rate of strain ($\dot{\gamma}$ = d tanθ)/dt) and the proportionality constant is viscosity:

$$\eta \dot{\gamma} = \tau \tag{7.5}$$
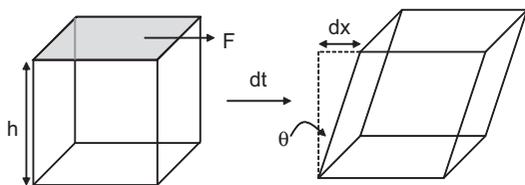
Fig. 7.19 Simple shear applied to a volume element of fluid. An applied force per unit area of face of the fluid (F) leads to a deformation of $dx/h = tan\theta$ in time dt

Where Eq. 7.4 is a definition of viscosity in terms of the operation of the viscometer Eq. 7.5, is the general case and can be applied to any type of shear flow.

7.8　Viscosity of Dilute Polymer Solutions

So how can we take our improved definition and physical model of viscosity and use them to understand the effects of added polymer? The common approach is to neglect the complexities of polymer shape and to simply treat the various random coils and globules as hard spheres suspended in solvent. Clearly a simplification, but one with some justification; globules are on average spherical and to some extent resist deformation (Fig. 7.6). We will return to the limitations of this simplification later, and first examine the useful results that follow.

The viscosity of a suspension of rigid spheres increases with volume fraction as:

$$\eta = \eta_s \left(1 + 2.5\phi_v + x_1\phi^2 + x_2\phi^3 + \cdots\right) \quad (7.6)$$

where η is the viscosity, η_s is the viscosity of the solvent, ϕ_v is the volume fraction of the particles, and x is a series of constants for the expansion that depend on the type of flow. The linear term in volume fraction represents the properties of the isolated spheres themselves. A rigid particle suspended in flowing liquid blocks streamlines of flow (Fig. 7.17c). If we were to maintain the same rate of rotation in the inner cylinder of our viscometer, the remaining layers not blocked by the particles must flow past one another more quickly with a consequently greater frictional

force. In effect, the polymer reduces the "effective" gap size. A greater force is needed to achieve the same rate of rotation so the measured viscosity is higher.

As concentration increases, the higher order terms in Eq. 7.6 become significant and viscosity is higher than predicted by the properties of the isolated molecules. The higher order terms describe hydrodynamic interactions between the spheres and account for the additional energy dissipated by their moving past one another. However, even with higher order terms, Eq. 7.6 is only reliable at quite dilute systems where the polymer coils do not overlap with one another and the liquid flow around one coil is not strongly affected by the presence of neighboring chains. The properties of more concentrated solutions are described in the next section, but first we will see how Eq. 7.6 can be used to investigate the properties of the polymers themselves in the dilute limit.

The key parameter here is the constant 2.5—the universal number that relates the measured viscosity of a dilute polymer solution to the properties of the polymer coil. We can isolate it by first rearranging Eq. 7.6 to define a new term:

$$(\eta - \eta_s) / (\eta_s \phi_v) = 2.5\phi_v + x_1\phi^2 + \cdots \quad (7.7)$$

The limiting value of $(\eta - \eta_s) / (\eta_s \phi_v)$ as ϕ_v tends to zero is the intrinsic viscosity $[\eta]$; equal to 2.5 if the polymer is a sphere. Intrinsic viscosity can be measured from a plot of measurements of $(\eta - \eta_s) / (\eta_s \phi_v)$ as a function of concentration extrapolated back to zero concentration (Fig. 7.20). Intrinsic viscosity is useful to us as we can take measurements of viscosity at real concentrations, which depend on a multitude of complex interactions and, by extrapolating back toward zero concentration, one can say something about the individual particle.

The hard-sphere model allows us to deal with highly idealized objects in suspension but how can we translate this approach into properties of a polymer? Intuitively, we can expect a relatively small mass of polymer to have a large effect on viscosity as the extended nature of the coil means a lot of solvent is entrained within the "sphere"
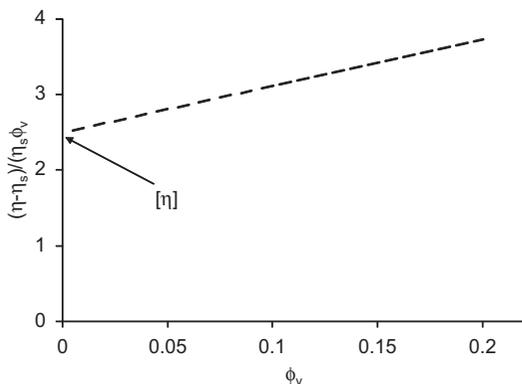
Fig. 7.20 A plot of $(\eta - \eta_s)/(\eta_s \phi_v)$ as a function of concentration yields the intrinsic viscosity by extrapolating to zero concentration

(i.e., most of the volume of the sphere drawn around the coil in Fig. 7.17c is entrained solvent). A small molecule in solution does not entrain solvent in the same way so a glucose solution is much less viscous than an amylose solution at the same mass concentration. However, to practically use the equations we have developed so far, we need to translate the volume fraction occupied by chains into terms in the mass concentration and molecular weight of the polymers.

First, the volume fraction of a polymer solution (ϕ) can be expressed as $n.V_{sphere}/V$, where n is the number of polymer molecules, V_{sphere} is the volume occupied by each polymer molecule, and V is the overall volume of the system. Substituting into Eq. 7.6 and excluding higher-order terms in ϕ (as concentration is low) gives:

$$\frac{\eta - \eta_s}{\eta_s} = 2.5\phi = 2.5\left(\frac{n}{V}\right)V_{sphere} \qquad (7.8)$$

The number of polymer molecules per unit volume of solution (i.e., n/V) is given by cN_{av}/M_w where c is the mass concentration, N_{av} is Avogadro's number, and M_w is the molecular weight of the polymer, therefore:

$$\frac{\eta - \eta_s}{\eta_s} = 2.5c\frac{N_{av}}{M_w}V_{sphere} \qquad (7.9)$$

Dividing through by c, the mass concentration of the polymer, allows us to calculate an intrinsic viscosity of the polymer coil:

$$\lim_{c \to 0}\frac{\eta - \eta_s}{\eta_s c} = [\eta] = 2.5\frac{N_{av}}{M_w}V_{sphere} \qquad (7.10)$$

Finally we need to modify the expression to replace the V_{sphere} with a more useful parameter related to the properties of the polymer.

V_{sphere} is hard to calculate, but it is proportional to the cube of the root mean squared end-to-end distance, that is, $V_{sphere} = k.r_{rms}^3$ where k is a proportionality constant, which tends to vary from one polymer to another (i.e., is not dependent on the size or M_w of the chain, but rather on the chemical identity of monomers comprising the chain). We already know that the radius of a polymer coil is proportional to its molecular weight raised to a power (i.e., $r_{rms} \sim n^v$, Eq. 7.1) so V_{sphere} is proportional to molecular weight raised to the power 3v. Substituting this term into Eq. 7.10 and combining all the nonmolecular weight parameters as a constant (K) gives the Mark–Houwink equation:

$$[\eta] = KM_W^{3v-1} \qquad (7.11)$$

For an ideal random coil (i.e., in a θ solvent) $v = 0.5$, but other values are possible depending on the precise shape that the polymer adopts in a given solvent ($v = 3/5$ for good solvent and $v = 1/3$ for poor solvent, Sect. 7.3). The relationship between viscosity and molecular weight is sometimes used to measure the molecular weight of an unknown polymer from a calibration curve of the viscosity of chemically similar standards of known molecular weight.

7.9   More Concentrated Polymer Solutions

Our physical picture of a viscous polymer solution shown in Fig. 7.17c is valid for dilute solutions when the polymer coils do not overlap. Polymers coils in solution tend to repel one another (Fig. 7.6), but beyond a critical concentration, c^*, they are forced to overlap and the increased interactions between the entangling polymers further contributes to the viscosity. The actual value of c^* generally decreases with the

molecular weight of the polymer coils and with solvent quality. For most food applications of polysaccharides $c*$ is exceeded.

Example: Characterizing Guar Gum

Funami and co-workers (2005) were interested in the effects of different guar gum preparations on the gelatinization behavior of corn starch. To characterize their guar samples, they calculated the intrinsic viscosity from a plot of viscosity vs. concentration. To measure solution viscosity, they used a capillary viscometer; the liquid was allowed to flow through a narrow capillary under gravity and the time it takes was proportional to viscosity. The specific viscosity per unit concentration increased with polymer concentration, suggesting polymer–polymer interactions were important in these samples (Fig. 7.21a). How-
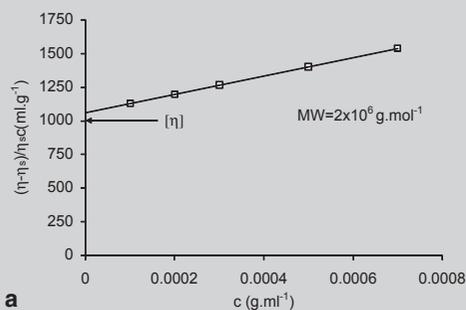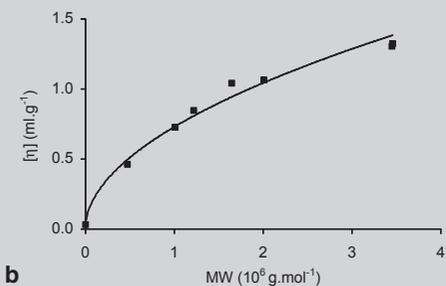
Fig. 7.21a Specific viscosity per unit concentration as a function of guar gum concentration. **b** Intrinsic viscosity as a function of molecular weight for a series of guar gum samples. *Line* shows fit of the Mark–Houwink equation. (Adapted from Funami et al. (2005))

ever, by extrapolating a straight line to zero concentration they were able to calculate the intrinsic viscosity of each guar sample. Intrinsic viscosity increased with molecular weight (as measured by a combination of chromatography and light scattering) and the data fit with the Mark–Houwink equation (Fig. 7.21b) with $\alpha = 0.51$, close to the expected value for an ideal random coil polymer.

The viscosity of a dilute polymer solution typically increases with concentration as $\eta \sim c^{\alpha}$ and experimentally the exponent lies in the range 1.3–1.5. However, beyond the critical concentration $c*$ the variation of the viscosity with the concentration of polymer becomes much more pronounced, with an exponent of 3.3–3.5. Such rapid variation of the viscosity above $c*$, allows the food scientist to tailor the viscosity of a product by several orders of magnitude using only a modest amount of a polysaccharide gum.

However, the behavior is more complex than simply an increase in viscosity. As polymer concentration increases, non-Newtonian behavior is increasingly important. In a Newtonian liquid, viscosity is independent of shear rate (Fig. 7.18; Eq. 7.5). The most common type of non-Newtonian behavior is a decrease in viscosity with increased shear rate (i.e., shear thinning or pseudoplasticity) but the opposite behavior is also possible (i.e., shear thickening or dilatency) (Fig. 7.18). In some cases, the viscosity also becomes dependent on the time for which the flow

Example: Viscosity of Xanthan Gum

Dhami and co-workers (1995) measured the viscosity of xanthan gum solutions using a concentric cylinder viscometer (Fig. 7.22). Viscosity was proportional to concentration up to a critical concentration $c* = 0.5$ g/l but beyond that point increased more rapidly with concentration.

Fig. 7.22 Viscosity of a xanthan gum solution as a function of concentration showing the limit of the dilute regime (*c**) (adapted from Dhami et al 1995). Insert shows a cartoon representation of a polymer solution above or below *c**. In dilute polymer solutions, the coils are isolated but above *c** there is no space to available to accommodate nonoverlapping coils
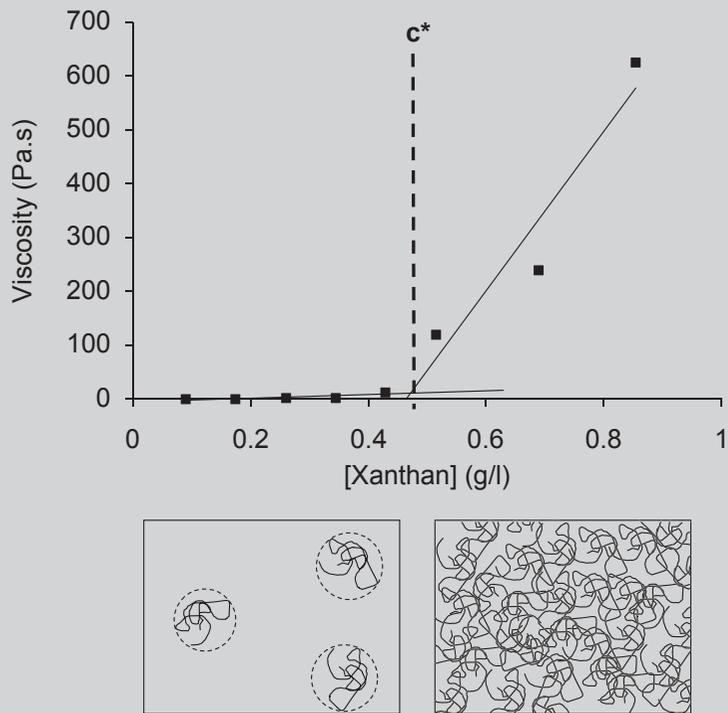
is applied. i.e., rheopectic and thixotropic fluids, where viscosity increases or decreases with time, respectively.

Newtonian flow could be explained in terms of hard spherical particles obstructing the streamlines in the flowing solvent, and non-Newtonian flow can be understood in terms of the particles being modified in some way by the flowing liquid. For example, Fig. 7.23 shows spherical particles randomly dispersed in a stationary fluid. If, in response to flow, they either deformed to present a smaller aspect to the flowing liquid or formed trains to block fewer streamlines then the viscosity would decrease. The magnitude of the change in structure is usually dependent on the magnitude of the applied shear force so the flow is shear-thinning. Dilatency is less common but is sometimes seen if the polymers increasingly interact with one another as a result of shear. It takes time for a fluid microstructure to respond to the application or removal of an applied force which explains the dependency of viscosity on time and flow history.
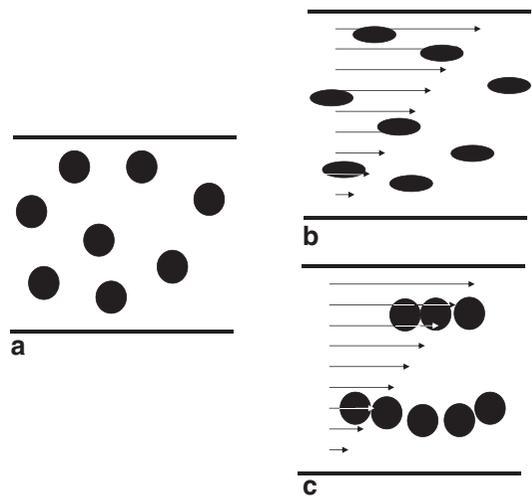
Fig. 7.23 a A random suspension of spherical particles in a stationary fluid can be shear thinning if the particles either, **b** deform to present a smaller aspect to the shear gradient, or **c** arrange form streamlines

Whether non-Newtonian behavior is actually observed in an experiment depends on the

Fig. 7.24 A suspension of rods progressively align with increasing shear rate leading to a decrease in viscosity from one Newtonian plateau to another
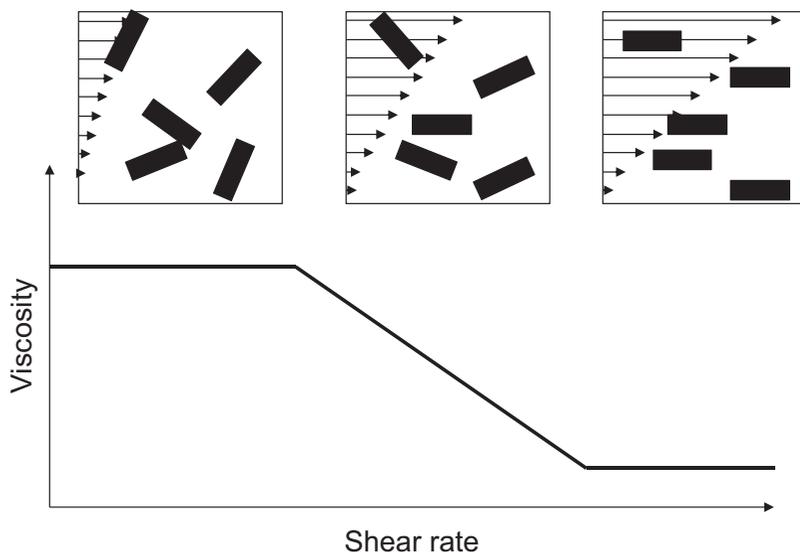
type of microstructure in the fluid, and the times and shear rates applied. In general, larger structures (e.g., bigger polymers) will respond more strongly to lower shear gradients than small polymers do. In some cases, there may be a limit to the degree a sample can respond to applied shear gradients; beyond this limit, the liquid will again become Newtonian. As an illustration, consider a suspension of rods (Fig. 7.24). At rest, they are aligned randomly to the flow and the viscosity is high. Low shear rates impose too small a shear gradient on the rods to cause a change in their average alignment and the fluid is Newtonian. As shear rate increases they progressively align and the liquid is shear-thinning. Beyond a certain point the rods are fully aligned, presenting their smallest possible aspect to the flow so further increases in shear rate cannot reduce the viscosity.

Non-Newtonian behavior is often modeled by rewriting Eq. 7.5 as a power-law function, the Ostwald–de Waele equation:

$$\tau = K\dot{\gamma}^n \qquad (7.12)$$

Where K is the consistency coefficient and n is the flow behavior index. The consistency coefficient is the viscosity of the sample at a shear rate of $1\ s^{-1}$ and the flow behavior index is a measure of how viscosity changes with shear rate. If $n < 1$ the liquid is shear thinning, if $n > 1$ the liquid is

shear thickening, and if $n = 1$ Eq. 7.12 reduces to Eq. 7.5 and the liquid is Newtonian. Non-Newtonian fluids are often described in terms of their apparent viscosity (η_{app}), the ratio between stress and rate of strain at any point in the flow. Apparent viscosity is often a power-law function of shear rate.

7.10 Summary

We started this chapter by thinking about how the simple model of molecules as particles that attracted and repelled one another could be adapted to deal with polymers. The model that emerged was of a random-walk coil modified by the interactions of the monomers with one another and with water. This simple physical model had to be adapted to explain the precise structure of native proteins and the coils of polysaccharides, but provided a reasonable physical picture of the shapes that may be present. Importantly, the more expanded coils, notably polysaccharides, take up

Example: Ketchup Rheology
Some manufacturers add polysaccharides as thickeners to improve the texture of tomato ketchup. Koocheki and co-workers
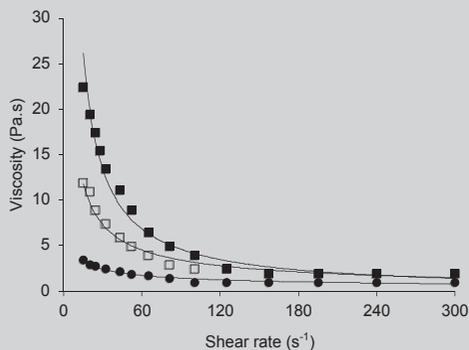
Fig. 7.25 Viscosity as a function of shear rate for ketchup samples with added xanthan gum with a power-law function alongside the data. (From Koocheki et al. (2009))

Table 7.1 Power-law parameters for ketchup with added xanthan gum at 25 °C. (From Koocheki et al. (2009))

[Xanthan] (wt%)	K ($Pa.s^n$)	n
0	19.34	0.228
0.5	23.82	0.218
0.75	24.84	0.211
1	26.73	0.204

(2009) used a concentric cylinder viscometer to measure the viscosity of ketchup with added xanthan gum as a function of viscosity (Fig. 7.25). In all cases, the viscosity of xanthan gum decreases with increasing shear rate and fits well with the power-law equation. The intrinsic structure of the ketchup (larger K) increases with added xanthan but this structure changes more rapidly and causes a bigger decrease in viscosity with increasing shear (lower n) (Table 7.1). Practically this means that adding xanthan to ketchup would make it very difficult to start it flowing out of the bottle but once it was moving the effects would not be seen. Viscosity is decreasing most sharply with shear rate at low shear rates but the viscometer used here did not provide useful data below 3 s^{-1}. It is likely that either the viscometer was rotating too slowly to be measured, or not rotating at all. A liquid that flows only after a critical stress, the yield stress, is exceeded is a plastic. At stresses lower than the yield stress the plastic will stretch and deform like a solid but not flow like a liquid. A ketchup with a higher yield stress would "stand up" better on the plate as the xanthan help support the liquid against the force of gravity. The properties of solid-like materials is the subject of Chap. 9.

much more of the solution than their own molecular volume, and are much more viscous than a corresponding small molecule solution. The effects of polymers on food texture become more complex at higher concentrations with as the polymers conformation responds to the flow. In Chap. 9, we will continue this topic and see how polymers can give rise to solid structures, gels.

Polymers tend to be less soluble than their monomers as there is little entropy gain on adding a polymer to solvent. Phase-separated polymer solutions can be used to generate microstructure in foods (e.g., dispersions of particles or dispersions of one liquid phase inside another). The properties of dispersions in general are the topic of the next chapter.

7.11 Bibliography

There are many excellent books on the physics and chemistry of polymers, and the introductory ones provide more depth to the material covered here without becoming too mathematical. I found "The Colloidal Domain" (Chap. 7, Evans and Wennerström 1994) particularly good. Grosberg and Khoklov (2010) take a more playful approach to the major concepts and their use of "toy" simulations in the CD accompanying the book provides a strongly intuitive explanation of how polymer molecules behave. "Painter and Coleman on Polymers" (Painter and Coleman 2004) is another highly interactive look at polymer science and also recommended. Polymer phase diagrams are described in Chap. 31 of "Molecular Driving Forces" (Dill et al. 2003) using an extension of the regular solution model used in Chap. 4. Another approach to polymer phase diagrams is provided by Picuell et al.

(1995) while Morris (2009) describes the properties of gels formed from mixtures of polymers.

"An Introduction to Food Colloids" (Dickinson 1992; Chap. 3) offers a very readable discussion of the basics of rheology. A more complete discussion is "Rheology" by Macosko (1994) and Chaps. 2, 5, and 11 are most relevant to the material described here.

Fennema's Food Chemistry provides good detail on the chemistry of proteins (Chap. 5; Damodaran 2008) and polysaccharides (Chap. 3; BeMiller and Huber 2008). Walstra (2003) discusses polymers generally (mainly polysaccharides) and proteins in Chaps. 6 and 7 of "Food Physical Chemistry." Rees' (1977) "Polysaccharide Shapes" takes an unusual approach to explain how the bonds of primary structure set the tendency to form helices and deserves to be more widely read.

Dispersions

8.1 Introduction

A central theme to this book has been the formation of multiple phases within foods. We have seen how mismatched intermolecular interactions can lead to phase separation (Chaps. 4 and 7), how the properties of phase separated systems depend on the interface (Chap. 5), and even considered crystallization as an example of phase separation (Chap. 6). In this chapter the focus is on the properties of dispersions.

Foods that are fine dispersions of one or more phases in another phase share some common properties regardless of what the phases are and can be helpfully considered as a group. Importantly, we are interested in small particles with at least one dimension larger than the molecular scale but not so large that the particle constitutes a distinct macroscopic object. Thus, while an ice cube in a drink falls outside our discussion, we can expect dispersions of fine ice crystals in a sorbet or bubbles in soda or oil droplets in mayonnaise to have some properties in common. Philosophically we are still trying to understand the properties of a whole system using knowledge of the interactions of its component parts, but to do so we are moving from the study of the molecular scale ($\sim 10^{-10}$–10^{-9} m) to the study of mesoscopic or colloidal scale assemblies of molecules ($\sim 10^{-8}$–10^{-5} m).

Dispersions have different names depending on the nature of the phases present. For example, a liquid-in-liquid dispersion is an *emulsion*, a solid-in-liquid dispersion is a *sol*, and a gas-in-liquid (or solid) is a *foam* (or solid foam). Some examples of food emulsions, sols, and foams are given below

and to give some insights into the types of common properties, we will examine in this chapter:

- Milk is an emulsion of dairy fat in an aqueous continuous phase. In fresh milk, the oil phase is about 3.5 % of the mass and the droplets have a diameter of about 3.5 μm. A cream layer will slowly form on the surface of fresh milk unless it is homogenized to reduce the milk fat globule size.
- Ouzo (and many related aniseed-flavored alcoholic drinks from the northern Mediterranean) are clear but become cloudy when mixed with cold water. The aniseed oil (anisole) is soluble in strong ethanolic solution but precipitates out as fine oil droplets when the alcohol is diluted with water.
- Peanut butter is made by milling peanuts. In the final product, fine peanut particles are dispersed in liquid peanut oil as a sol. Sometimes lecithin is added to make the product smoother and less prone to oiling off.
- Meringue is a foam made by whipping air into a sweetened egg white solution. The bubbles are relatively large (~mm) and their net volume is many times that of the egg. Uncooked meringues are unstable and will break down over a few hours unless they are stabilized by cooking at a high temperature for a short time to denature and partially gel the proteins or by cooking at a low temperature for a long time to dry and form a crisp solid. In both cases, the liquid foam is converted to a solid foam.

Although these products are very different from one another, they share some common properties with each other and with other dispersions:

- Dispersions are often turbid even if their ingredients are not.
- Dispersions are more viscous than the continuous phase and can even behave as solids.
- Dispersions have at least three ingredients (two phases and an emulsifier).
- Dispersions tend to separate into two bulk phases over time; sometimes this is significant in a food (e.g., a beer foam collapses quickly) and sometimes not (e.g., the milk emulsion spoils by bacterial action before the phase separates).
- Dispersions are manufactured, most often by breaking up one phase in another with a homogenizer or mixer but sometimes by controlling a phase separation.

In this chapter, we will examine these common features of food dispersions, but we will start by developing a language to describe the structure of dispersions.

8.2   Characteristics of Dispersions

A dispersion can be characterized by a relatively small set of measurable parameters describing what is dispersed in what as well as the concentration and the size/shape of the particles.

Type of Dispersion In a dispersion, discrete particles of one phase are dispersed in a second continuous phase. For example, in an oil-in-water emulsion the droplets of oil are dispersed in an aqueous phase. All of the water is effectively interconnected but each droplet of oil is discrete. If the phases were *inverted* so that water droplets were dispersed in oil, it would be a water-in-oil emulsion. A common error is to assume the phase present in the largest amount is the continuous phase but this is not the case (e.g., mayonnaise is a dispersion of 80 % oil droplets in a water continuous phase). It is usually possible to determine which phase is continuous because a dispersion can only be readily diluted by the continuous phase (e.g., peanut butter is readily diluted in oil but not water because it is a dispersion of peanut particles in peanut oil while cream can be diluted in water but not oil because it is a dispersion of oil droplets in water). Alternatively, a water-soluble

dye will rapidly color a water-continuous dispersion (such as milk), but not an oil-continuous one (such as peanut butter).

Dispersion   Concentration Having   established the type of dispersion, the next important parameter is the relative concentration of phases present. Concentration is readily expressed as a mass fraction, that is, mass of the particles relative to the total mass of the system, (ϕ_m) and can be determined from proximate analysis, density measurements, or simply from knowing the formulation. In many cases, droplet concentration is better expressed as a volume fraction, that is the volume of the total of the dispersion taken up by the dispersed phase. Volume fraction (ϕ_v) can be calculated by multiplying the volume fraction by the ratio of densities of the dispersed and continuous phases, for example, a salad dressing made from 200 g of olive oil ($\rho \sim 900$ g cm^{-3}) and 800 g of vinegar ($\rho \sim 960$ g cm^{-3}) would have a mass fraction of 20 % ($=200/[800+200]$) and a volume fraction of 21.3 % ($=20\% \times 960/900$).

In a foam, the mass of the gaseous dispersed phase is effectively zero, so volume fraction is the only useful measurement. Foam concentration is often expressed as a fractional overrun:

$$\text{Overrun} = \frac{\rho_{\text{liquid}} - \rho_{\text{foam}}}{\rho_{\text{foam}}} = \frac{\phi_v}{1 - \phi_v} \qquad (8.1)$$

where ρ is the density of the foam (ρ_{foam}) or unwhipped liquid phase (ρ_{liquid}), which can readily be measured as the mass of foam or un-whipped starting liquid needed to fill a standard volumetric container. For example, when ice cream mix is frozen, an approximately equal volume of air is incorporated into the mix so $\rho_{\text{foam}} \sim \frac{1}{2} \rho_{\text{liquid}}$ and the overrun is 100 % and the volume fraction is 50 %. A low-density foam such as the head on beer might have a dispersed phase volume fraction of 90 % and an overrun of 900 %.

There is no minimum limit to the dispersed phase fraction in a dispersion (e.g., particles of cloud in apple juice or the flavor emulsions used in beverages can have $\phi_v < 0.01$ %). However, the geometric constraint of packing rigid particles into a confined space sets a maximum volume
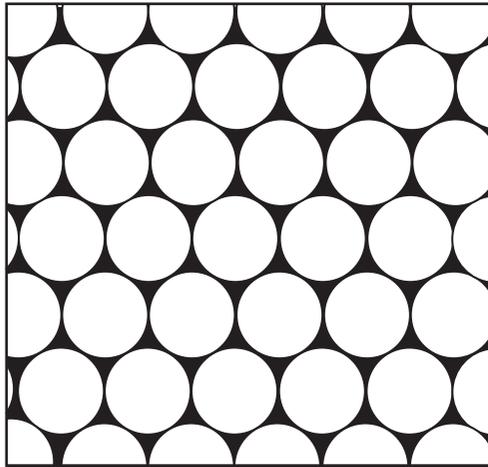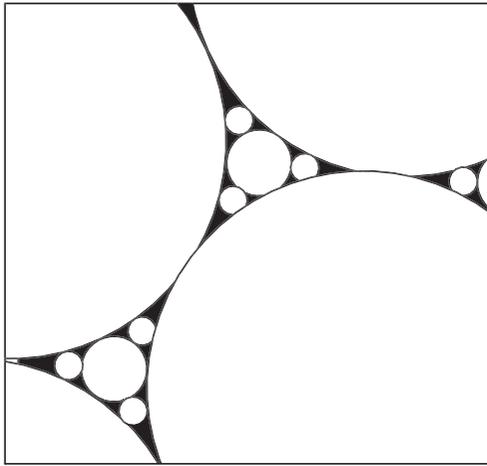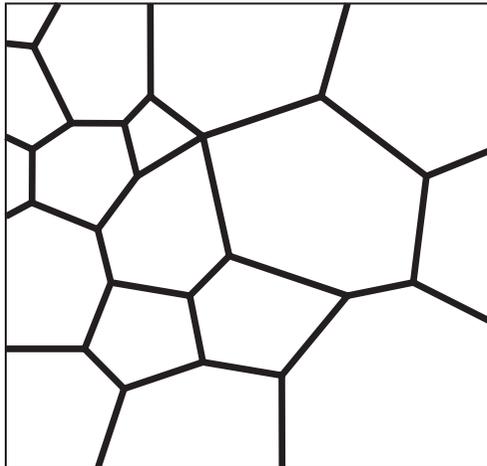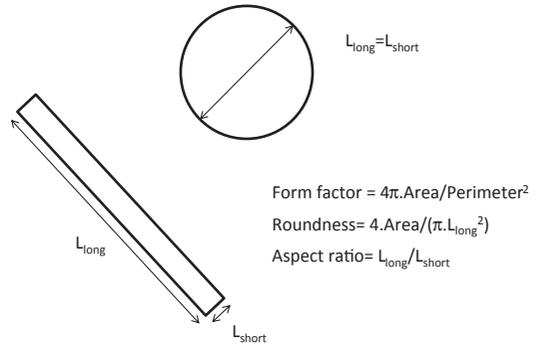
Fig. 8.2 The size of sphere can be described with a characteristic diameter while various characteristic parameters can be used to describe the size and shape of an irregular object (see Russ 2007 for more detail)

fraction, ϕ_{max}. A good illustration of this phenomenon is that it is possible to add water to a bucket that has previously been filled with marbles. The geometry of the marbles is such that there are always gaps between them that leave space for about a quarter of a bucket of water (Fig. 8.1a, the maximum dispersed phase volume fraction $\phi_{max} \sim 74\,\%$). Note that this value of ϕ_{max} is for spherical particles that have effectively "crystallized" into a regular packing pattern. Real dispersions would be unlikely to be so well ordered and a random packing of the same particles would have a lower maximum volume fraction. On the other hand, if there is a distribution of particle sizes (Fig. 8.1b) or if the particles are able to deform (Fig. 8.1c), then much larger dispersed phase volume fractions are possible. Very high dispersed phase volume fraction dispersions such as beer foam and mayonnaise show both of these features. Nonspherical, rigid particles tend to have lower values of ϕ_{max}.

Particle Shape and Size Fine fluid particles are spherical because the Laplace pressure tends to equalize surface curvature (see Sect. 5.7) while larger fluid droplets can more easily deform and solid particles may have irregular shapes. While spheres can be characterized with a unique diameter, irregular shapes cannot, and various parameters can be measured (usually by microscopy) and used as characteristic sizes (Fig. 8.2).

Whichever dimension is chosen and however it is measured, real food dispersions always contain a range of differently sized particles (i.e., they are

Fig. 8.1 **a** Maximum volume fraction for uniform spheres, **b** different-sized spheres, and **c** deformable particles

	(a) Monodisperse	(b) Polydisperse	(c) Skewed	(d) Bimodal
Diameter	Number of particles			
1	0	0	0	4
2	0	2	0	0
3	7	3	4	0
4	0	2	2	2
5	0	0	1	1
Median	3.0	3.0	3.0	1.0
Mode	3.0	3.0	3.0	1.0
d_{10}	3.0	3.0	3.6	2.4
d_{20}	3.0	3.1	3.6	3.0
d_{32}	3.0	3.4	3.9	4.2
d_{43}	3.0	3.5	4.0	4.4

Table 8.1 Size distribution data for model dispersions

polydisperse rather than monodisperse). Size distributions can be represented as histograms (i.e., the proportion or number of the particles in a given range of sizes), or if the number of bins is large, a scatter plot where size is shown as a continuous variable. Size distributions are sometimes shown as cumulative distributions, that is, the proportion of particles smaller than a given size.

It is often convenient to report a single average diameter rather than showing the overall distribution, but there are many ways to calculate an average. Table 8.1 and Fig. 8.3 show four very simple particle size distributions. The median of the distribution is the size that half the particles are bigger than and half smaller. The mode of the distribution is the maximum of the distribution, the size with the largest number of particles. Alternatively, the average could be taken by simply averaging the diameters of all the particles, for example, for the polydisperse distribution: $[(2 \times 2) + (3 \times 3) + (4 \times 2)]/7 = 3$. This is the length average mean diameter (d_{10}):

$$d_{10} = \frac{\sum_i d_i n_i}{\sum_i n_i} \qquad (8.2)$$

where n_i is the number of particles of diameter d_i. (The subscripted numbers mean there is a first power of length in the numerator and a zeroth power in the denominator.) If surface area were more important than length, then the area average mean diameter (d_{20}) could be calculated by taking the average of the surface areas of all of the particles then taking a square root to calculate a diameter:

$$d_{20} = \sqrt[2]{\frac{\sum_i d_i^2 n_i}{\sum_i n_i}} \qquad (8.3)$$

If all the particles were identical with diameter d_{20}, then the total surface area would be the same as the original distribution. A volume average can also be calculated by the same approach. Two widely used distributions for dispersions are the Sauter or surface–volume mean diameter (d_{32}) and the volume fraction-length diameter (d_{43}):

$$d_{32} = \frac{\sum_i n_i d_i^3}{\sum_i n_i d_i^2} = \frac{6\phi_v}{A_s} \qquad (8.4)$$

$$d_{43} = \frac{\sum_i n_i d_i^4}{\sum_i n_i d_i^3} \qquad (8.5)$$

The Sauter mean diameter is particularly useful as it is readily related to the average interfacial area per unit volume (A_s) and volume fraction (ϕ_v) of the dispersion.

Which average diameter is most appropriate will vary on a case-by-case basis according to which gives the most useful prediction of other properties of the dispersion. It should be stressed that going from a full distribution to a single average always discards some information. In many
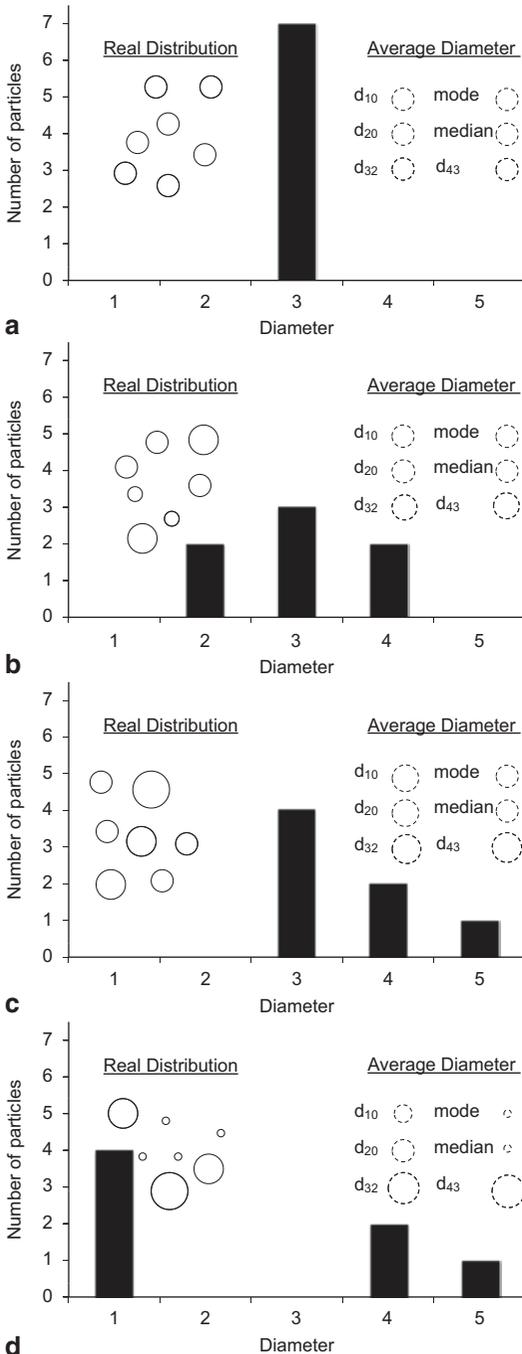
Fig. 8.3 a Particle size distributions for monodisperse, **b** polydisperse, **c** skewed, and **d** bimodal distributions of particles. Actual particles and representations of various average diameters are shown *inset*. Data from Table 8.1

cases, it is also helpful to include some measure of the polydispersity of the distribution, such as, the width of the peak at half maximum height or the difference between the 25th and 75th percentile of the distribution.

Particle size can be measured by a variety of methods. The most direct is to visualize the structure using some form of microscopy. However, for microscopy to be reliable, the images generated must be representative of the structure present. A major challenge with imaging methods is that the technique selected is capable of seeing the smallest particles present but if the magnification is too great, then only very few particles will be seen in the image and may not provide a representative sample of the overall distribution. The maximum possible resolution of a microscope (i.e., the smallest objects that can be resolved) is in the order of the wavelength of the radiation used, so for optical microscopy (i.e., visible light, $\lambda \sim 500$ nm) many colloidal particles cannot readily be measured. The high-energy electron beam used in electron microscopy can visualize much smaller objects but sample preparation is often more disruptive and may create artifacts.

Although less direct, light scattering methods are usually preferred over microscopy for the characterization of fine dispersions. When a beam of light passes through a dilute suspension of particles, some light is scattered in different directions. Light scattering explains why many concentrated colloidal suspensions appear turbid and white (e.g., smoke, foams, milk). In static light scattering experiments, a laser beam is passed through a very dilute suspension of particles and the intensity of the scattered light is measured as a function of angle relative to the incident beam. The scattering pattern can be theoretically predicted in terms of the wavelength of the light and the size and optical properties of the particles, so a size distribution can be calculated to give the best fit between theory and experiment (i.e., the inverse scattering problem). In an alternative group of methods, dynamic light scattering, the intensity of a scattered light at a single angle is measured as a function of time. The intensity of the signal changes as the positions of the particles move relative to one anoth-

er and the rate of change is related to how fast the particles are moving, that is, their diffusion coefficient. Diffusion coefficient can be related to the size of the particles if the viscosity of the continuous phase is known (Eq. 2.2). In general, dynamic light scattering is preferred over static light scattering for finer, more mobile particles. An advantage of scattering methods is that a very large number of particles are measured simultaneously so the sampling issues common in imaging methods are not significant. However, in most cases the dispersion must be extensively diluted before measurement (e.g., a typical sample for dynamic light scattering looks something like a drop of milk in a glass of water), which can disrupt inherent structure. All scattering methods suffer from the fact that the particles are not directly seen and sizing is based only on a theoretical model of their interaction with light.

Example: Characterizing Lactose Powder

Yucel and Coupland (2010) were interested in developing an ultrasonic sensor to determine the properties of suspension of lactose crystals in water and they needed to first characterize their crystals. Butanol was selected to disperse the particles for a static light scattering experiment because it is clear and colorless and, more importantly, the sugar will not dissolve in it. The results from light scattering are shown (Fig. 8.4) along with an optical micrograph of the same sample. The light scattering method depended on a theoretical prediction of scattering originally developed for dilute, spherical particles. While the butanol may have served to disperse and dilute the particles effectively to meet the first condition, the crystals were not spherical so the scattering theory could not be expected to work perfectly. Despite this, the average particle size calculated from light scattering was in reasonable agreement with the results from microscopy.
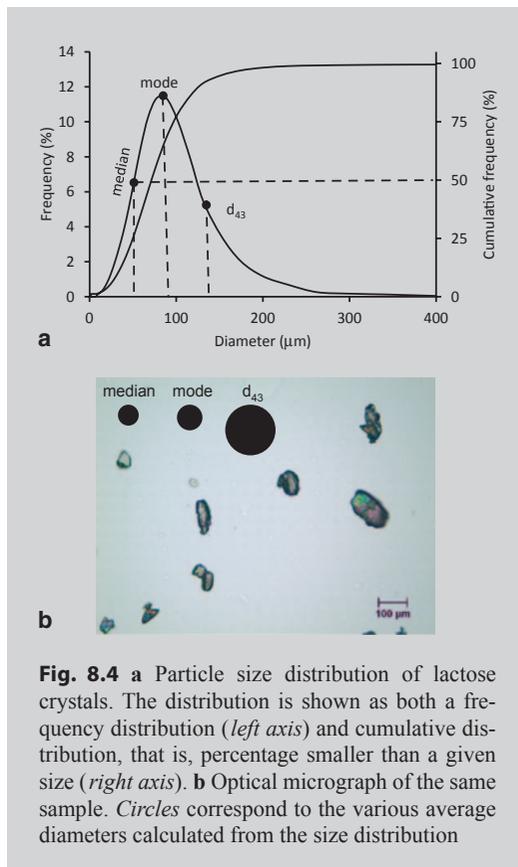
Fig. 8.4 a Particle size distribution of lactose crystals. The distribution is shown as both a frequency distribution (*left axis*) and cumulative distribution, that is, percentage smaller than a given size (*right axis*). **b** Optical micrograph of the same sample. *Circles* correspond to the various average diameters calculated from the size distribution

Particle Aggregates In many cases, the particles in a dispersion are present not as individual objects but as clusters of particles or flocs. Each individual particle retains its own identity but the forces holding them together (see Sect. 8.3) are greater than any shear forces that might break them up.

Flocs can be characterized in terms of an effective size (see dotted line in Fig. 8.5) and/or an effective density (i.e., average density of the material inside the dotted line, a value between that of the dispersed and continuous phase). Note that the effective volume fraction of a flocculated dispersion (i.e., the sum of the effective volumes of the flocs) is always greater than the volume of the nonflocculated particles as each floc contains a volume of the continuous phase. Flocs typically have an open branching structure sometimes characterized with a fractal dimension, that is,

Isolated particles
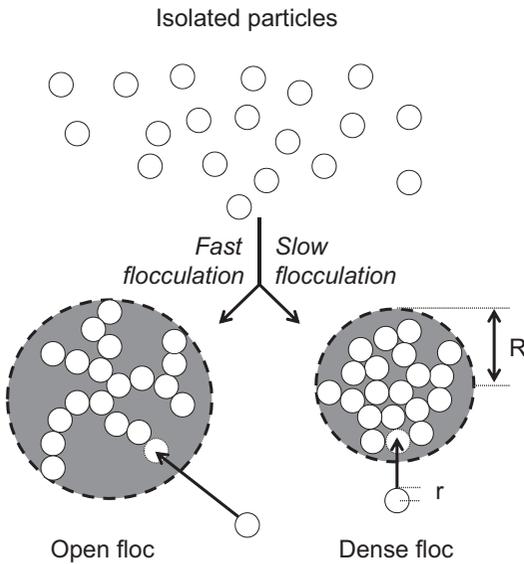
Fig. 8.5 Particles can combine to form open and loose flocs. The loose floc has a larger effective radius and more voids containing continuous phase. A particle adding to the open floc will adhere at wherever it first collides with the growing surface because of the strong attractive forces. A particle adding to the dense floc will only stick if it collides with the denser center of the floc where it makes more contacts because each individual contact is weaker

$$R = rn^{1/D} \qquad (8.6)$$

where R is the radius of a floc of n particles with radius r each. The parameter D is the fractal dimension where $D=3$ if the individual particles merge to form a single spherical particle, and $D=1$ if they form a string. In a fractal floc $3>D>1$ with smaller fractal dimensions corresponding to more open structures. In practice, floc structures can be hard to measure as they are readily damaged by many sample preparation methods. Common techniques include light scattering, imaging, and sedimentation/creaming measurements.

Flocs form from the aggregation of the primary particles following collision. If the particles are very "sticky" (i.e., strong attractions), then the floc will grow rapidly with new particles attaching to the outside of the existing aggregate at the point when they first come into contact. On the other hand, if there are only weak attractions, then many of the collisions will be ineffective

and only the collisions with multiple connection points will stick to the existing aggregate. Consequently strong attractions between particles leads to fast-growing open aggregates with low fractal dimensions while weak attractions lead to slow-growing dense aggregates with large fractal dimensions. Flocs can readily modify their structure after formation.

8.3 Interparticle Forces

To understand the behavior of colloidal dispersions, we need to understand how the particles interact with one another. We faced a similar problem in Chap. 2 where we calculated an interaction potential between two molecules and used that to understand the behavior of solutions (Chap. 4). Some example interparticle potentials are shown in Fig. 8.6; compare these with the intermolecular potential in Fig. 2.8, but remember that zero separation here is the two surfaces coming into contact. Remember that the potential is the energy surface upon which the particles move as they interact. Particles will tend to move "downhill" on this surface and accumulate at minima:

a. Repulsive at all separations. The particles will repel one another with a force that increases at smaller separations.

b. Attractive at all separations. The particles will begin to attract one another when they come within range and will eventually come into contact with one another. Once in contact there is a large energy cost in separating the particles. (Note that separation is the distance between the surfaces of the particles, whereas in Chap. 2 separation was the center-to-center distance between molecules.)

c. Intermediate energy maximum. Particles repel one another at long separations but attract strongly at short separations providing an energy barrier slowing or preventing the particles coming into contact. Once the particles are in contact, the reverse pathway is energetically highly unfavorable and they will tend to remain in contact. This model is very similar to the kinetic theory used in Fig. 3.1.
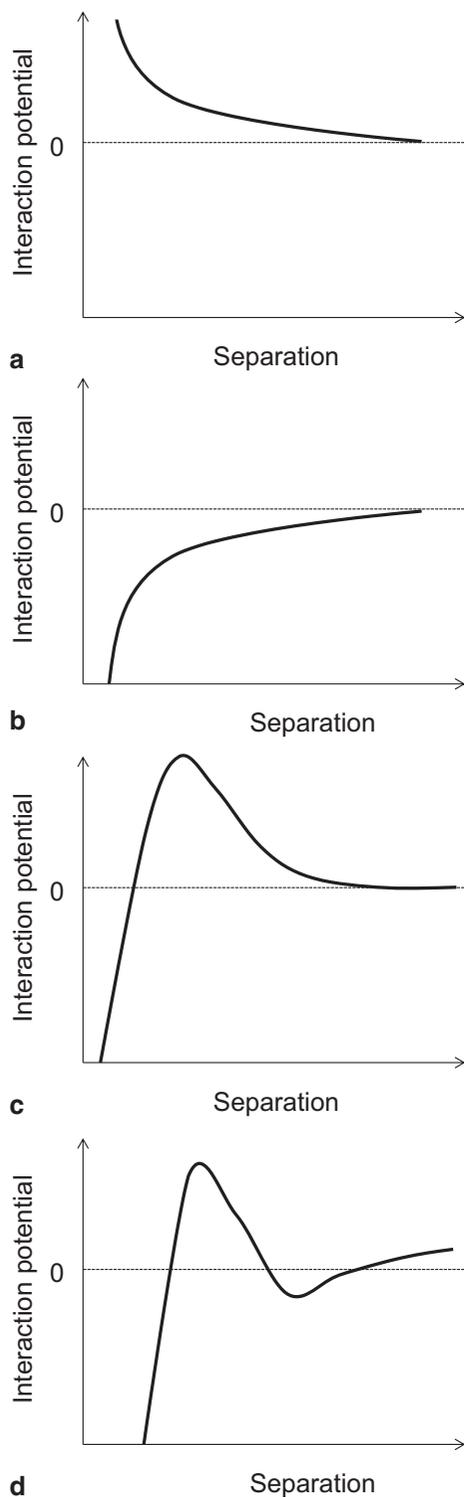
d. Intermediate energy minimum and maximum. The behavior of this type of dispersion can be understood in terms of the distribution of particles between three different energy minima (i.e., infinite separation, at the minimum at small separations, and in contact). If the energy difference between infinite separation and the small separation minimum is large, particles will tend to stick together to form loosely associated flocs. If the energy difference between the small separation minimum and the maximum is large, the particles will tend to only slowly move to the true minimum (i.e., particles in contact). Again once the particles are in contact, the reverse pathway is energetically highly unfavorable and they will tend to remain in contact.

It is possible to directly measure colloidal interaction potentials from the forces acting between two particles (see example below). However, it is instructive to be able to predict the interactions in terms of the properties of the system. In principle it would be possible to calculate the behavior of a colloidal dispersion simply from the sum of all of the intermolecular forces acting. But in practice, it is easier to explain the behavior of the colloidal system in terms of the interactions of the particles. This approach is particularly useful as the expressions are often in terms of parameters that can be experimentally altered to achieve predictable changes in the stability of the dispersion. Our goals here are similar to the molecular case: we will consider the various types of interaction in turn, then sum them to produce an interparticle pair potential describing the energy change required to bring two particles to a given separation. Although several of the interactions have similar names to their molecular counterparts, their nature can be quite different. In particular, while the fundamental molecular interactions all depended on some derivation of Coulomb's law and are entirely energetic in nature, many colloidal interactions have significant entropic components. We will examine a few important examples of intermolecular interactions, the texts in the bibliography offer many more.

Fig. 8.6 Model interparticle potentials. **a** Repulsive at all separations. **b** Attractive at all separations. **c** Intermediate energy maximum. **d** Intermediate energy minimum and maximum
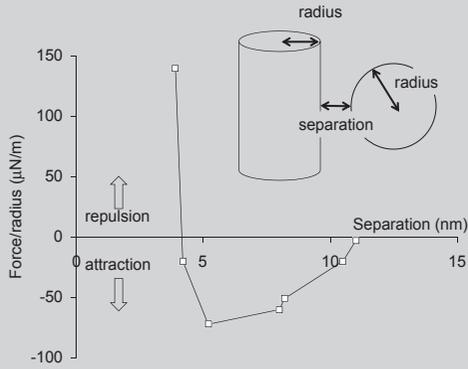
Fig. 8.7 Forces between two crossed mica cylinders in 0.1 M KNO$_3$ as a function of separation. Adapted from Israelachvili and Adams (1978)

Example: Measuring Colloidal Interaction Potentials

A colloidal force balance works by very precisely measuring the forces between two crossed mica cylinders as a function of separation. Jacob Israelachvili was a pioneer of this technique and in one of his early papers he measured the forces between charged surfaces in ionic solution (Israelachvili and Adams 1978). Sample results in 0.1 M KNO$_3$ solution are shown in Fig. 8.7. Beyond about 10 nm, there is no interaction between the surfaces but moving closer there is an increasingly strong attraction reaching a maximum at about 5 nm. Moving the surfaces yet closer together leads to a weakening attraction and eventually a repulsion as the cylinder come into contact. If the surfaces were the surfaces of charged particles, they would be expected to aggregate at a characteristic separation of 5 nm.

Van der Waals Interactions Van der Waals forces between molecules are weak attractions between transient and induced dipoles in most types of molecules (Sect. 2.8). Molecular scale Van der Waals forces have a characteristically short range (proportional to the inverse sixth power of separation, Eq. 2.9). The sum of the

molecular Van der Waals interactions give rise to an attraction between colloidal particles, albeit one with a longer range. At close separations the Van der Waals pair potential between similar spherical particles is given by:

$$\mathrm{u}_{VDW}(s) = \frac{-Ar}{12s} \qquad (8.7)$$

where A is the Hamaker function, s the interparticle separation, and r the particle radius.

The Hamaker function is divided into a zero frequency component (corresponding to the orientation and induced contributions to the molecular Van der Waals interactions) and a frequency-dependent component (corresponding to the transient contributions to the molecular Van der Waals interactions). The frequency-dependent part reflects the time it takes for the changing electrical field from a one transient dipole in one particle to be felt by the second particle and is a function of the optical properties of the dispersed and continuous phases (i.e., refractive index and frequency of the major UV absorbance maximum). The zero-frequency part depends on the fixed dielectric properties of the particles. McClements (2005) calculates that for oil/water food emulsions 42 % of the Van der Waals interactions are from the zero frequency component and 58 % from the frequency-dependent component. In simple treatments, it may suffice to take the Hamaker function as a constant. For example, Walstra (2002) gives 1.2 kT, 0.7 kT, and 0.2 kT for liquid oil, crystalline fat acid, and biological cells in water at room temperature and 0.5 kT for triglyceride crystals in liquid oil. Using these figures, the force Van der Waals attraction is significant to well over 10 nm from the particle surface (Fig. 8.8).

In fact, the effective range may be somewhat less, as Eq. 8.7 is likely to overestimate the magnitude of the Van der Waals attraction. Firstly, ions accumulate near the surface of many particles (see below) and electrostatically screen the zero-frequency part of the interaction. Secondly, at long separations the finite time for the dielectric field from one droplet to reach the second
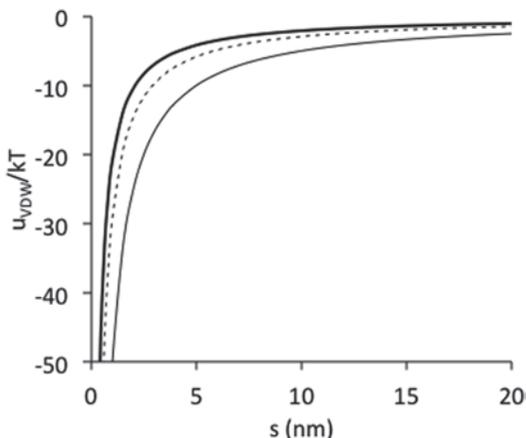
Fig. 8.8 Van der Waals interaction between fine ($r = 500$ nm) spherical dispersions of liquid oil (*thick line*) and solid fat (*dashed line*) in water, and solid fat in oil (*fine line*). Values of the Hamaker constant from Walstra (2003)

droplet will reduce the magnitude of the frequency-dependent part of the interaction in a process known as retardation. In addition, because molecular Van der Waals interactions are short range, molecules at the droplet surface (e.g., proteins, surfactants) have a disproportionate effect on the magnitude of colloidal Van der Waals forces.

In summary, colloidal Van der Waals forces provide an attraction between particles extending over quite long ranges. In the absence of a repulsive interaction, colloidal particles will tend to quickly aggregate.

Electrostatic Interactions Similarly charged ions repel one another by the Coulombic force (Eq. 2.5). Similarly charged particles in an aqueous solution also repel one another, but the force between the particles is largely an indirect effect of the charged surface on the surrounding ions in solution.

Most surfaces are charged due to the adsorption of charged emulsifiers or the binding of ions. The surface charge density (i.e., the number of charges per unit surface area) is proportional to the surface potential (ψ_0, i.e., the energy needed to create that charge density). Ions with charges opposite to that of the surface (i.e., counter ions) will cluster around the interface, while ions with

the same charge as the surface (i.e., co-ions) will be pushed away (Fig. 8.9). Thus, close to the surface there will be an imbalance of charge with a nonzero potential. Moving further away from the influence of the charged surface, the imbalance of charges will tend to subside and the potential will decrease. Eventually the randomizing effects of thermal motion will dominate the ordering effect of the surface charge and, sufficiently far from the surface, there is no imbalance of charge and the potential is zero. Thus a charged particle in aqueous solution is surrounded by a region of solution where the ionic balance is affected by the presence of the particle. It is convenient to divide this region into an inner and an outer part together known as the electrical double layer.

The properties of the inner part, the Stern layer, are dominated by ions closely bound to the surface. Especially important here are charged surfactants and proteins attached to the surface by hydrophobic interactions. The types and amount of adsorbed charges depend on the interactions of the various molecules with the solvent, with each other, and particularly with the surface. In general, the more charged groups adsorbed per unit area, the greater the surface potential. Many charged groups are readily protonated and deprotonated by changes in pH (Eq. 2.13), which can affect the magnitude and sign of the surface charge.

The potential at the edge of the Stern layer (i.e., the Stern potential, ψ_d) can differ from the surface charge in magnitude and even in sign. For example, consider a negatively charged surface adsorbing positive ions (Fig. 8.9a). The Stern potential has a lower magnitude than the surface potential. On the other hand, if an anionic polymer adsorbs to the positive surface, the Stern layer might be considerably thicker and the Stern potential negative (Fig. 8.9b). The properties of the Stern layer are hard to predict for a given mixture of ingredients but is often taken to be similar to the readily measured ζ-potential.

When a charged particle is placed in an electrical field, it moves toward the oppositely charged electrode. For larger particles the speed of movement can be measured using a microscope but
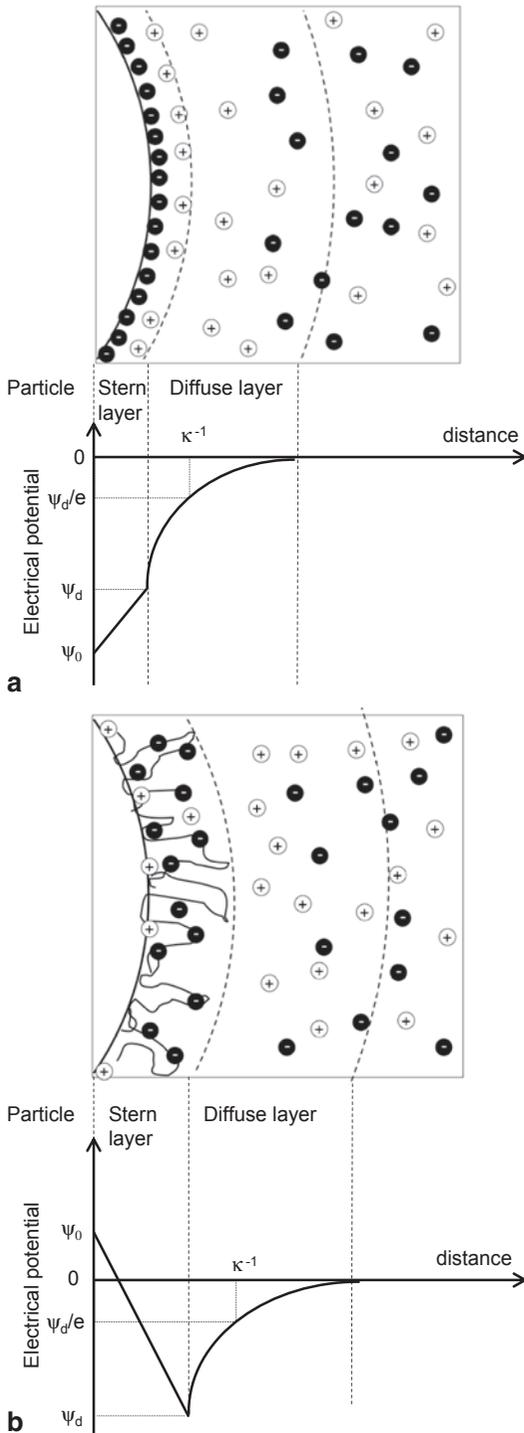
generally dynamic light scattering gives a better average velocity for smaller particles. However, the moving object measured is the particle itself along with tightly bound emulsifier, some water, and some strongly associated ions. The speed measured is neither the surface potential nor the Stern potential but rather the potential at the boundary between the parts of the double layer that move with the particle and the parts which do not, that is, the ζ-potential measured at the plane of slip. The plane of slip is hard to define, but the idea that it encapsulates the tightly bound material means that it is often used as a more readily measurable proxy of the Stern potential.

The region of the solution affected by the charged surface but outside the Stern layer is the diffuse layer. Here the distribution of ions can be calculated from the Poisson–Boltzmann distribution which, for low surface charges, approximates to an exponential decay:

$$\psi = \psi_d e^{-\kappa s} \tag{8.8}$$

where ψ is the potential, a distance s away from the surface of the Stern layer, and κ is the reciprocal Debye screening length. The Debye length is a measure of the effective range of influence of the charged surface (i.e., the distance over which the potential falls to $1/e$ of its maximum value). In room temperature water, κ approximates to $0.3/\sqrt{I}$ nm (where I is the ionic strength, Eq. 2.5). Increasing ionic strength reduces the region where the properties of the solution are affected by the properties of the charged particle. For examples in a molar sodium chloride solution, the Debye length is 0.3 nm and in a 10 mM solution, 3 nm (Fig. 8.10).

The electrostatic repulsion between two similarly charged particles occurs when the diffuse

Fig. 8.9 Distribution of ions around a charged surface (i.e., the electrical double layer) determines the potential in the solution adjacent to the surface. **a** Monovalent cations form a thin Stern layer around a strongly cationic

surface and the magnitude of the Stern potential is slightly less than the surface potential. **b** An anionic polymer adsorbs hydrophobically to a weakly cationic surface and forms a thick Stern layer. The magnitude and sign of the Stern potential is different from the surface potential. Beyond the Stern layer the potential decreases approximately exponentially in the diffuse layer
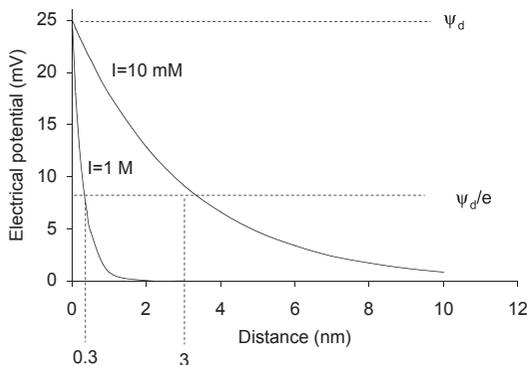
Fig. 8.10 Effect of ionic strength on the electrical potential adjacent to a charged surface ($\psi_0 = 25$ mV) in 1 M and 10 mM NaCl. The Debye length, i.e., the distance over which the potential falls to $1/e$ of its initial value, is much reduced at higher ionic strength
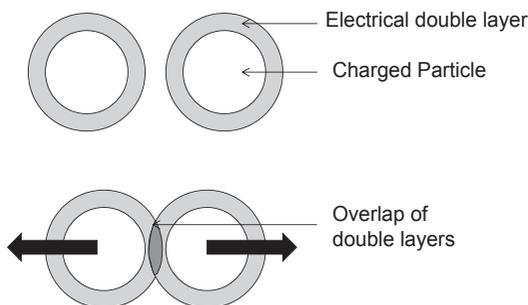
Fig. 8.11 When the electrical double layers of two charged particles overlap, there is an electrostatic repulsive force between the particles largely due to the excess concentration of ions in the overlap region

layers begin to overlap (Fig. 8.11). Surprisingly the enthalpic interaction between the various charged species present is slightly attractive but relatively unimportant. Electrostatic repulsion between charged colloidal particles is largely an entropic due to the excess concentration of ions in the overlap region. Assuming ions at the surface do not adsorb or desorb due to the approach of the second particle, the repulsive force between two similar charged spheres of radius r separated by a distance s is given by:

$$u_{electrostatic}(s) = 2\pi\varepsilon_0\varepsilon_r\psi_0^2 r.\ln\left(1 + e^{-\kappa s}\right) \quad (8.9)$$

where ε_0 and ε_r are the dielectric constant of a vacuum and the relative dielectric constant of the solution. This equation is reasonably reliable for small separations ($s < r$) and low to moderate
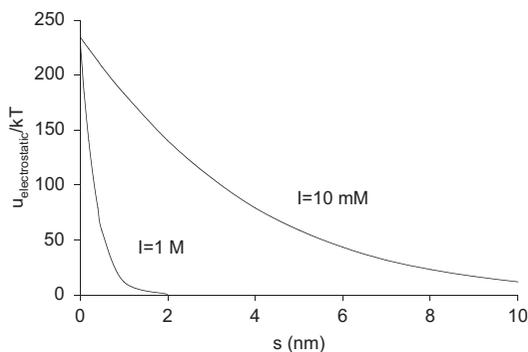
Fig. 8.12 Effect of ionic strength on the electrostatic pair potential between two charged surfaces ($\psi_0 = 25$ mV) in 1 M and 10 mM NaCl.

surface charges ($\sim < 30$ mV). The range of the electrostatic repulsion, approximately equal to the Debye length, is quite long especially at low ionic strength (Figs. 8.10 and 8.12). The magnitude of the force increases with the surface potential and the size of the particles.

Example: Solubility of Soy Protein Isolate

Soy protein is isolated from soybeans for use in other foods. During commercial extraction, the protein is denatured and is often highly aggregated. Thus soy preparations are often better described as sols of micron-scale protein particles suspended rather than dissolved in water. The suspension is stable at neutral pH but tends to precipitate in acids so while soymilk is quite shelf-stable, fruit beverages fortified with soy often form a precipitate. Malhotra and Coupland (2004) measured the stability of a soy protein isolate suspension, as a function of pH, and at the same time measured the ζ-potential of the protein particles. Stability was measured as the proportion of the protein remaining suspended after a gentle centrifugation (Fig. 8.13).

The ζ-potential was positive at low pH, but decreased with increasing pH as the hydrogen ion concentration in the solution decreased and the protein progressively released bound protons. At neutral and very low pH, the protein is highly charged (posi-
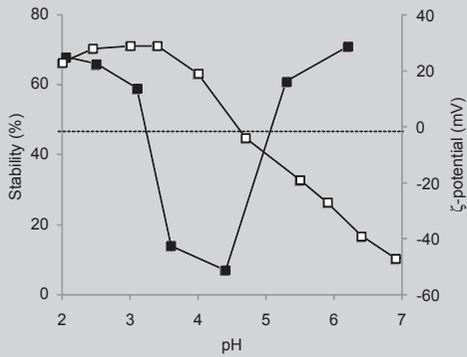
Fig. 8.13 Stability (*filled squares*) and ζ-potential (*open squares*) of soy protein isolate as a function of pH. Adapted from Malhotra and Coupland (2004)

tive and negatively, respectively) and there are strong repulsive intermolecular forces. At pH~4.6 the electrophoretic mobility is zero, so there is no repulsion between the protein particles. The uncharged particles aggregate and precipitate during the centrifugation. Note the electrophoretic mobility of the protein changes approximately linearly with pH rather than the characteristic sigmoid form seen for single functional groups (Fig. 2.17). A protein is a polyelectrolyte with many different weak acid and weak base functional groups whose dissociation curves overlap giving the diffuse shape seen.

In summary, electrostatic interactions provide long-ranged repulsion between similarly charged particles in water. The strength of the interaction is largely determined by the material adsorbed at the surface and is strongly affected by ionic strength and pH.

Electrostatic and Van der Waals interactions are often summed to give the DLVO interaction potential between two particles. While this simple interaction is rarely sufficient to describe the behavior of real food dispersions, it provides a simple model to begin to see how interactions can be related to stability. Figure 8.14 shows
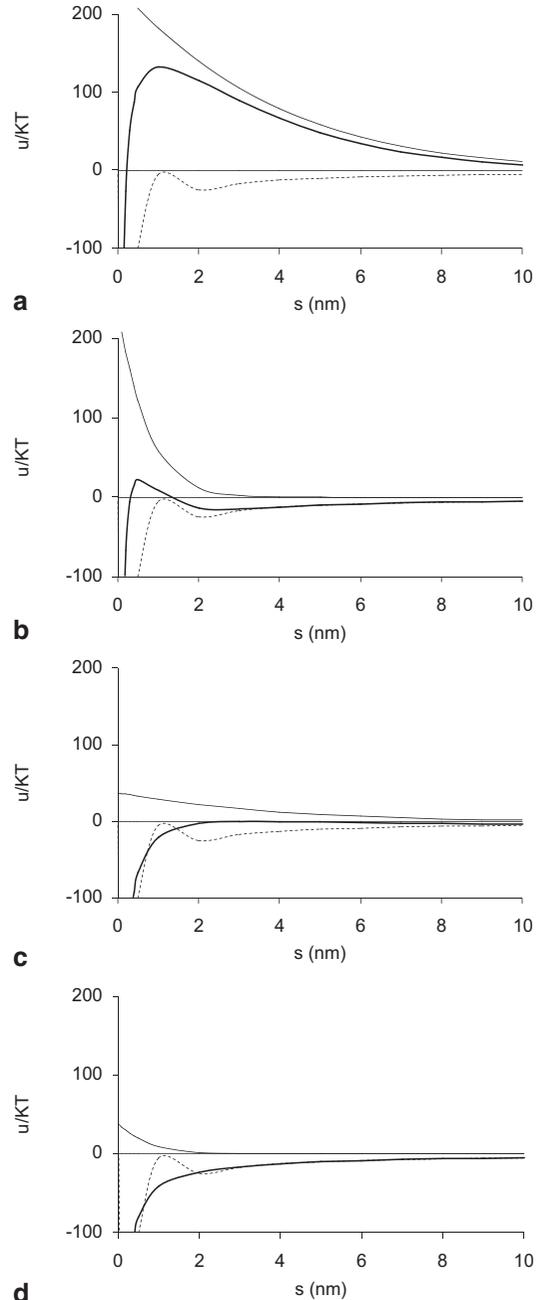
Fig. 8.14 DLVO potential between two identical oil droplets in water ($r=500$ nm, $A=1.2$ kT). **a** High surface potential ($\psi_0=25$ mV, low ionic strength ($I=10$ mM). **b** High surface potential ($\psi_0=25$ mV, moderate ionic strength ($I=250$ mM). **c** Low surface potential ($\psi_0=10$ mV, low ionic strength ($I=10$ mM). **d** Low surface potential ($\psi_0=10$ mV, moderate ionic strength ($I=25$ mM)
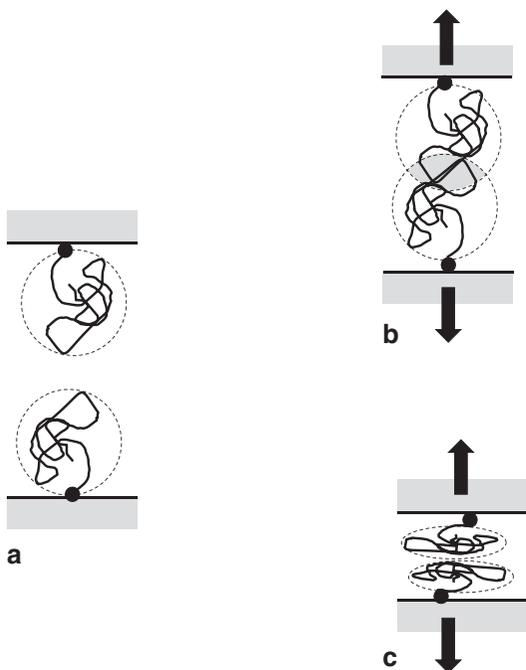
Fig. 8.15 Surfaces with tethered random coil polymers repel one another once the coils come into contact (**a**) as overlapping the coils generates an osmotic pressure gradient (**b**) and/or deforming the coils reduces the chain entropy and creates an osmotic pressure gradient (**c**)

the DLVO potential between two droplets of oil ($r=500$ nm, $A=1.2$ kT) in water as a function of separation for four cases generated by changing the surface potential (e.g., by changing the pH of a protein-stabilized emulsion or adding a charged surfactant) or changing the Debye length by changing the ionic strength. Remember that colloidal potential functions can be read in a similar way to the intermolecular potential functions in Chap. 2— a surface over which particles move accumulating at the minima and only passing the maxima when they have sufficient energy to clear the barrier.

1. High surface potential, low ionic strength. The electrostatic repulsion is strong with a long range. It exceeds the Van der Waals interaction at all but the shortest separations and the net interaction is repulsive and particles will tend to repel one another.
2. High surface potential, high ionic strength. The electrostatic repulsion is strong but with a short range. There is a weak energy mini-

mum at about 2 nm and a weak maximum at about 0.5 nm. Depending on the exact shape of the potential, some particles may tend to accumulate in the energy minimum but lack the energy to reach the true minimum. A potential like this has the potential to allow the droplets to flocculate (i.e., to remain at some short separation from one another without touching).

3. Low surface potential, low ionic strength. The electrostatic repulsion is weak but with a long range and serves to cancel out the Van der Waals attraction at longer ranges, but at shorter ranges the attractive force dominates. The effective range over which particles attract one another is decreased.
4. Low surface potential, high ionic strength. The electrostatic repulsion is weak and has a short range and does not significantly modify the Van der Waals attraction.

Steric Interactions Polymers in solution tend to resist overlapping with one another because of the osmotic pressure gradient between the overlapped region and the rest of the solution (Fig. 7.6a). Similarly, polymers in solution resist deformation from their original shape because this reduces the entropy of the chain and also increases the local polymer chain concentration (Fig. 7.6b). Now, if we imagine similar polymers tethered to the surface of a colloidal particle (Fig. 8.15). As a second similar surface approaches, there would be no interaction until the coils came into contact at which point the polymer coils must either interpenetrate or compress. Both interactions would be unfavorable and lead to repulsion between the droplets.

It is hard to formulate an interaction potential for the interaction of surfaces covered with real polymers, but in general it will be zero until the polymers come into contact and then strongly repulsive. A polymer would be expected to give strong steric stabilization if:

1. It adsorbs strongly at the surface so that it cannot simply desorb when close to a second droplet. This usually means having sufficient hydrophobic groups that can attach the polymer firmly to the particle surface.

a

b

Fig. 8.17 Diagram illustrating the mechanism of depletion interactions. *Fine particles* are excluded from a region near the surface of *large particles,* generating a region with a lower particle concentration and lower osmotic pressure than the bulk fluid. The osmotic pressure gradient leads to an attractive depletion force between the large particles

Fig. 8.16 Combined potential (*bold line*) resulting from a steric repulsion (*thin line*) plus a Van der Waals attraction interaction (*dashed line*) between two coated with **a** a thin (1 nm) coating of polymer and **b** a thick (5 nm) coating of polymer. In **a,** the energy minima is close to the particle surface so the particles will tend to stick to one another at this separation and flocculate. In **b,** the thick steric layer is of similar range to the Van der Waals attractive force and they effectively cancel one another out so the net interaction is approximately zero and the particles will not aggregate

2. It absorbs in large enough quantities to fully cover the surface.
3. It forms a thick coating to provide repulsion some distance away from the particle surface before the Van der Waals attractions become dominant (see illustrative example in Fig. 8.16). The typical thickness of a globular protein layer at an interface is in the order of a few nanometers although caseinate and adsorbed polysaccharides can extend tens of nanometers.

This model assumes no specific interactions between the polymers. For example a positively charged polymer on one particle and a negatively charged polymer on the second or a disulfide bond forming between two cysteine residues on different droplets, can lead to strong bonds between droplets.

In summary, steric interactions provide a strong repulsion between particles but only at quite short separations when the surface layers begin to overlap.

Depletion Interaction Small nonadsorbing particles can cause larger particles to aggregate via a depletion mechanism. Imagine a mixture of large and small noninteracting spherical particles (Fig. 8.17). The centers of the small particles cannot be any closer than their radius from the surface of the large particles, so the concentration of small particles in this space is reduced compared liquid to the concentration of small particles in the bulk away from the large particles. This concentration difference creates an osmotic pressure gradient with a higher concentration of water close to the surfaces and a lower concentration further away. The only way to resolve this is if the two large particles approach one another to reduce the volume of solution depleted in small particles. The osmotic pressure gradient is responsible for an attractive depletion force between the large particles. The force is entropic in nature, meaning it results from increasing the volume the small particles have available to occupy.

Depletion forces are attractive and have a range approximately equal to the size of the small particles. The magnitude of the force increases with the concentration of the small particles as well the size of the particles. Any nonadsorbing particle can cause depletion effects so long as it can be dispersed in sufficient concentrations and does not adsorb at the surface (e.g., fine emulsion droplets, nonadsorbing proteins or polysaccharide coils, surfactant micelles).

Example: Creaming in Flavor Emulsions

Flavors are often added to drinks in the form of very dilute oil-in-water emulsions. These emulsions are often stabilized with food starches modified with hydrophobic functional groups to make them surface active or by gum Arabic, a naturally surface-active polysaccharide. Chanamai and McClements (2001) hypothesized that excess polysaccharide in the aqueous phase of the emulsion might promote droplet flocculation by a depletion mechanism. They prepared a series of emulsions and diluted them in different concentrations of the two polysaccharides. Flocculated emulsions often cream faster so they measured the creaming rate using an optical scanning instrument. The device measures the reflected light intensity (alternatively the transmitted light intensity) as a function of height in a column of emulsion as it creams (Fig. 8.18). The amount of light reflected from the droplet-free serum layer at the bottom of the tube is much less than the amount of light reflected from the droplet-rich cream layer at the top of the tube (and vice versa for transmittance) so the boundary can readily be measured. Measuring the position of the boundary as a function of time gives a creaming rate. The initial creaming rate is plotted as a function of polysaccharide concentration in Fig. 8.19.

The rate of creaming in the absence of added polysaccharide was low and greater for the larger droplets. The emulsion surfaces
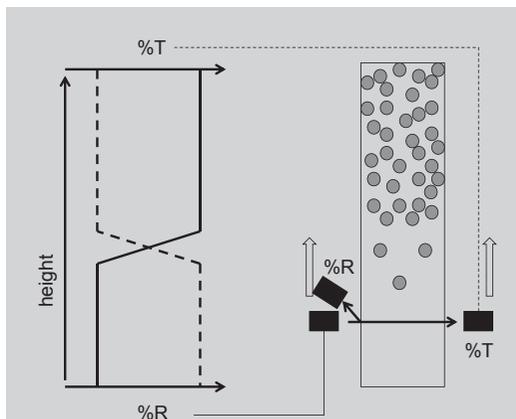
Fig. 8.18 Optical scanning instrument used for creaming measurements. The device measures reflected and/or transmitted light as a function of height in a tube of emulsion
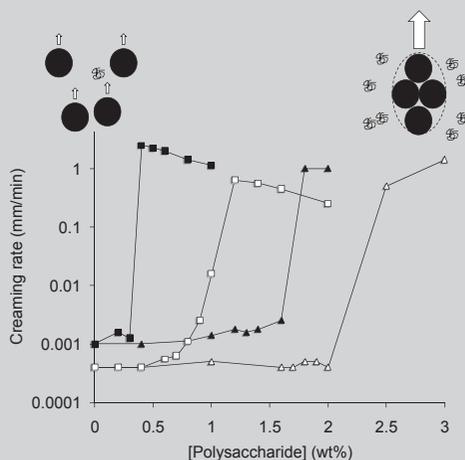
Fig. 8.19 Creaming rate of 5% oil in water emulsion (*open points* $d=1.04$ μm, *filled points* $d=0.34$ μm) as a function of aqueous modified starch (*black triangle, white triangle*) or gum Arabic concentration (*black square, white square*). *Inset* shows isolated droplets at low polymer concentration creaming more slowly than flocculated droplets at high polymer concentration. Adapted from Chanamai and McClements (2001)

were stabilized with a nonionic surfactant so any effect of the polysaccharide must be as a nonadsorbed polymer. Small amounts of polysaccharide caused no change in the creaming rate but beyond a critical point the creaming rate increased abruptly. Presum-

ably at this value, the depletion attraction was greater than the stabilizing effects of the surfactant and the droplets flocculated and creamed more rapidly. The critical concentration of polymer was less for the larger droplets because depletion attraction is greater for larger droplet. The critical concentration of gum Arabic was less than for modified starch because the gum Arabic has a higher molecular volume. Slight decreases in creaming rate at high polymer concentrations are probably due to an increase in solution viscosity.

Fig. 8.20 *Spherical* and *nonspherical particles* interacting with a *flat plate*. Although the volumes of the *particles* are the same, the area over which the *nonspherical particle* can interact with the *surface* (*A*) is greater

In summary, depletion interactions are attractive forces due to the osmotic pressure gradient in the region close to the surface that is depleted of the fine particles. The strength of the attraction depends on the number concentration of the depleted particles so the effects of depletion interactions can often be reversed by dilution.

Hydrophobic Interactions Particles with hydrophobic surfaces dispersed in water will attract one another due to the combination of attractions between hydrophobic molecules (Sect. 2.10). The force is strongly attractive over a few nanometers and is proportional to the interfacial tension between the hydrophobic group and water. Coating hydrophobic particles with an emulsifier will reduce the exposure of hydrophobic groups to the aqueous phase and the magnitude of this force. A second, longer range component to hydrophobic interactions has been measured and is postulated to depend on dissolved gas (Francis et al. 2006). In this mechanism, the gas violently comes out of solution as tiny small bubbles when the hydrophobic surfaces approach one another generating an attractive force with a much longer range. Interestingly, degassing a hydrocarbon in water mixture makes it much easier to make a stable emulsion without adding any surfactant. This mechanism remains a topic of active research.

In summary, hydrophobic interactions between hydrophobic particles suspended in water are important when the surface is inadequately covered with emulsifier. In protein-stabilized dispersions, denaturation of the interfacial protein can lead to the exposure of hydrophobic groups and generate an interparticle attraction.

This small set of colloidal forces provides a sense of how particles can interact with one another over distances of nanometers. There is a large set of other potentially important interactions discussed in more detail in the resources in the bibliography (e.g., capillary forces, Sect. 5.7). Importantly, the strength of the forces depends on the size and shape of the particles. The formulations given here are for spherical particles and typically are greater for larger particles. Nonspherical particles can often present larger surfaces to one another and so the net force on each particle is greater (Fig. 8.20). Each of these interactions can individually be expressed in terms of the properties of the system and then summed to get a sense of how changes to the formulation will make the particles more or less prone to sticking together—a key factor in their stability.

8.4 Dispersion Stability

Dispersions are thermodynamically unstable because of the large surface excess-free energy and will eventually phase separate. There are various mechanisms of phase separation but they can be broadly characterized as destabilization due to destabilization due to diffusion (i.e., Ostwald ripen-

ing), gravity (i.e., creaming and sedimentation), and destabilization due to attractive forces between particles forces (i.e., flocculation and coalescence).

Destabilization due to Molecular Diffusion As we saw in Sects. 5.7 and 6.5, the contents of small particles can spontaneously diffuse to large particles driven by a difference in surface curvature, that is, Ostwald ripening. Ostwald ripening leads to a characteristic change in droplet radius with the cube root of time and is most important when the dispersed phase has some solubility in the continuous phase and the particles are small (e.g., ice crystals in ice cream, flavor oil emulsions in water).

Destabilization due to Gravity If particles are less dense than the continuous phase, they will tend to float and form a particle-rich cream layer at the surface (Fig. 8.21, e.g., creaming in salad dressings, foam on the top of a glass of beer). On the other hand, if the particles are denser than the continuous phase, they will tend to sink and form a sediment at the bottom of the container (e.g., many insoluble food powders will fall out and form a sludge). Both processes lead to the separation of the initially homogeneous dispersion into a dispersed-phase rich layer and a dispersed-phase depleted layer. It is usually possible to undo creaming/sedimentation by gently mixing the separated layers back together. However, if the particles are allowed to remain in close proximity to one another, they may begin to flocculate and coalesce (see below) which cannot be reversed by mixing.

The stability of a dispersion to creaming/sedimentation can be understood in terms of the forces acting on a single particle (Fig. 8.20a). The gravitational force on the particle depends on the density contrast with the continuous phase and the volume of the particle:

$$F_{gravitational} = -\frac{4}{3}\pi r^3 \Delta \rho g \qquad (8.10)$$

where $\Delta \rho$ is the density contrast (i.e., density of the particle minus the density of the continuous phase), g is the acceleration due to gravity, and
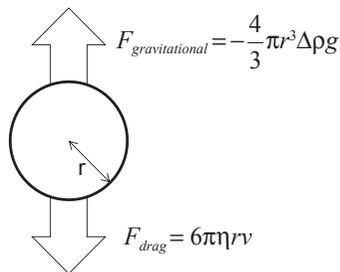
a
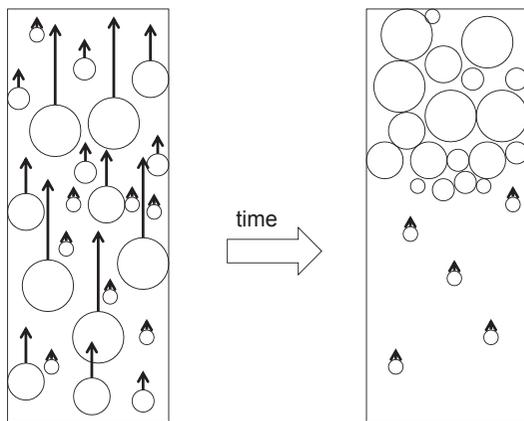
b

Fig. 8.21 a Forces acting on an *isolated particle* in a viscous fluid. The particle is less dense than the continuous phase so the density contrast ($\Delta\rho$) is negative and $F_{gravitational}$ is positive and the particle moves upwards. Other constants are defined in the text. **b** The net effect of these forces is the separation of particles to form a cream layer at the top of the samples. *Larger particles* move faster so there is often a distribution of particle sizes within the creamed layer with larger nearer the top. Similarly the bottom of the tube may not be completely clear because the *smaller particles* cream much more slowly

r is the particle radius. If $F_{gravitational}$ is negative, the particle will tend to move downwards and vice versa. The particle will accelerate in response to this force and as it does it will experience an opposing drag force proportional to its size and to its velocity (v) as well as the viscosity of the continuous phase (η):

$$F_{drag} = 6\pi\eta r v \qquad (8.11)$$

As the particle moves faster, the magnitude of the opposing drag force will increase until eventually it matches the gravitational force. At this point

there is no net force acting on the moving particle and so it will continue to move downwards at the same constant velocity. We can calculate this velocity (v_{max}) as the value where the gravitational and drag forces on the particle become equal. Combining Eqs. 8.10 and 8.11 gives:

$$v_{max} = \frac{-2gr^2\Delta\rho}{9\eta} \qquad (8.12)$$

In practice, the droplet will accelerate to its terminal velocity almost instantaneously so Eq. 8.12, the Stokes equation, gives a measure of the speed of droplet movement. Dickinson (1992) suggests a creaming rate of less than 1 mm/day can be considered adequately stable, so accepting this rule we can quickly estimate the stability of any system. Perhaps more importantly, Eq. 8.12 provides a sense of which factors could be expected to slow sedimentation, that is:

- Reducing particle size, for example homogenizing the fat in milk.
- Increasing continuous phase viscosity, for example adding a viscous water-soluble polysaccharide to a salad dressing.
- Reducing density contrast, for example using a brominated vegetable oil to increase the density of a flavor oil in an oil/water emulsion or removing sugar to lower the density of the aqueous phase (see example below).

Although the Stokes equation gives a useful sense of the parameters controlling the rate of creaming and sedimentation, its quantitative predictions are likely to be unreliable for real food dispersions as many of its assumptions are violated. Firstly, the continuous phase is unlikely to be a Newtonian fluid and the actual viscous drag affecting the particle will depend on the rate of strain imposed by the droplet. Most biopolymer solutions are shear thinning (i.e., high viscosity at low shear rates, see Sect. 7.8), and the strain rates imposed by particles moving under gravity are very small ($\sim$ mPa). Consequently, a dilute polysaccharide solution may seem very fluid in our everyday use (fast flow, relatively large forces) but at the same time be much more viscous or even effectively solid to fine particles moving through it.

Secondly, the drag term neglected any interactions between particles (i.e., assumed a dilute system). In more concentrated dispersion the continuous phase must move past a mass of particles moving in the opposite direction increasing the effective drag. At the same time the movement of larger, faster particles is blocked by slower-moving particles (Fig. 8.22). The rate of gravitational separation will eventually decrease to zero as the volume fraction approaches close packing as the particles are "jammed" (see Fig. 8.21b, the particles at the top of the container stop moving as there is no space for them to move into while the smaller slower particles at the bottom are still free to move).

Finally, the particles are modeled as isolated spheres. Nonspherical particles cream more slowly because their greater surface area means more drag. Furthermore, when particles flocculate they will cream or sediment as a unit and the floc itself becomes the effective particle. The floc has a larger radius than an isolated particle but a lower density contrast (because some continuous phase is trapped within its structure, see inset in Fig. 8.19). The size effect tends to dominate at least initially and particle flocculation increases the creaming rate up to a certain point. Very extensive flocculation can gel the sample and stop all gravitational separation. The extensive aggregation of uncharged soy particles and the depletion flocculation of flavor oil emulsion droplets was responsible for the enhanced rate of creaming in the examples in Sect. 8.3 (i.e., Fig. 8.13 and Fig. 8.19 respectively).

Example: Creaming of Flocculated Emulsions
Chanamai and McClements (2000) used the back-scattering method described above to measure the creaming rate of fine (radius 0.86 μm) oil droplets in water as a function of particle volume fraction. As expected, the more concentrated emulsions creamed more slowly (Fig. 8.23). Next they used surfactant micelles to induce depletion flocculation in the same emulsions and measured the creaming rate. The more dilute flocculated emulsions creamed

Fig. 8.23 Creaming rates in flocculated and nonflocculated oil in water emulsions (radius 0.86 μm). Adapted from Chanamai and McClements (2000)

exceptionally quickly but the rate of creaming decreased more rapidly with volume fraction in the flocculated system. Beyond about 30 wt% dispersed phase, the rate of creaming of both flocculated and nonflocculated emulsions was slow and similar.

Destabilization due to Aggregation Aggregation is commonly seen as one of two processes (Fig. 8.24):

- Flocculation, where multiple particles stick together while retaining their individual identity. Flocculation is sometimes reversible if the forces causing the particles to stick to one another can either be reduced or overcome by shear forces. Flocculation is notable for causing a large increase in viscosity or even forming a gel (see example above). Moderate flocculation can increase creaming rate (see Sect. 8.4), but if the dispersion becomes too viscous, creaming will stop (see Sect. 8.5)
- Coalescence, where multiple fluid particles merge to form a single larger particle. Coalescence is only reversible by re-homogenizing the phases. Coalescence reduces the number of particles and eventually will lead to complete phase separation (i.e., one remaining "particle" of the dispersed phase). Coalescence has less effect on viscosity than flocculation and typically increases the rate of creaming.

The processes of particle aggregation can be seen occurring in three stages. First the particles must encounter one another. Second they must interact

Fig. 8.22 Effect of dispersed phase volume fraction on creaming rate **a** in a dilute dispersion, the *marked particle* is free to move at the Stokes velocity, **b** at finite concentration, the *marked particle* has to move past other slower particles and is itself slowed, and **c** at close packing, the *marked particle* is jammed and cannot cream
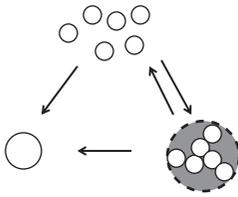
Coalescence
Particle number decrease
Particle size increases
Effective particle size increases
Volume fraction constant
Effective volume constant

Flocculation
Particle number constant
Particle size constant
Effective particle size increases
Volume fraction constant
Effective volume fraction increases

Fig. 8.24 Particle aggregation by flocculation and co-alescence

and come into close contact. Interactions that take place with a deep energy minima at close separations will cause particles to stick together and flocculate (Fig. 8.6d). Finally if fluids inside droplets or bubbles come into contact with one another their contents can mix and they will co-alesce.

Moving particles will occasionally collide and the factors determining the collision rate depend on the dominant mechanisms causing the movement (i.e., Brownian motion, induced flow, creaming/sedimentation). Particles in a static fluid move via random Brownian motion, and their collision frequency, F_B, is given by the Smoluchowski equation:

$$F_B = k_B n^2 = 16\pi Drn^2 \qquad (8.13)$$

where k_B is the second-order rate constant and n is the particle concentration (number per cubic meter). The rate constant is a function of the diffusion constant for the particle in the continuous phase (D) and the particle radius (r). The Smoluchowski equation is more usefully expressed for spherical particles in a Newtonian medium by using Eq. 2.2 to calculate the diffusion coefficient and expressing the number concentration in terms of a volume fraction, ϕ:

$$F_B = \frac{3kT}{2\pi^2} \cdot \frac{\phi^2}{\eta r^6} \qquad (8.14)$$

where η is the continuous phase viscosity. For a 10 % oil in water emulsion with 0.5 μm radius

particles, Eq. 8.14 predicts about 10^{17} collisions/s. Under Brownian motion, collision rate is greater at higher volume fractions for smaller particles in less viscous media. Flow can dramatically increase the particle collision rate beyond that seen in static systems. Under simple shear the particle moving in faster streamlines will tend to catch up to and collide with particles moving in slower streamlines at a rate, F_S:

$$F_S = \frac{3}{\pi} \cdot \frac{G.\phi^2}{r^3} \qquad (8.15)$$

where G is the shear rate. Collision rate increases the shear rate and volume fraction and is greater for smaller particles but continuous phase viscosity is not an important factor. Finally, particles undergoing gravitational separation will move at different rates and the faster-moving will tend to collide with the slower-moving at a rate. The collision rate can be reduced either by minimizing the creaming/sedimentation rate, or by using a monodisperse distribution so all the particles move at the same rate. Of course, extensive creaming/sedimentation will eventually force the particles into close proximity at the top or bottom of the sample where they may begin to aggregate.

As two particles come close to one another they interact through a combination of colloidal forces (see Sect. 8.3) and hydrodynamic forces. There are two main components to hydrodynamic forces. First, the dispersed phase needs to be squeezed out of the gap between the approaching surfaces. Second, the flow of the dispersed phase can drag surfactants along the surface out of the region where the particles are approaching one another and create a surfactant-depleted region. Surfactants will tend to diffuse back to overcome the resulting surface tension gradient and drag dispersed phase along with them back into the gap between particles and pushing the particles apart (i.e., the Gibbs–Marangoni effect, Fig. 8.25). Taken together, hydrodynamic forces tend to oppose the approach of particles.

The presence of a repulsive interaction, either from the equilibrium colloidal forces or the non-equilibrium hydrodynamic forces, means only a fraction of the encounters between particles are "effective collisions," that is, lead to floccula-
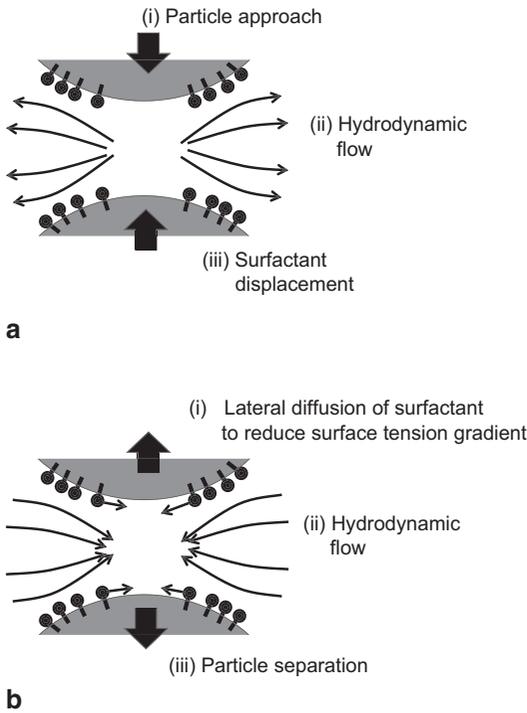
Fig. 8.25 Gibbs–Marangoni effect. **a** Two particles approach one another and the flow of continuous phase out of the gap drags some surfactant along and reduces the surfactant concentration and increases the interfacial tension at the interface close to the approaching particle. The resultant interfacial tension gradient causes the surfactant to flow across the interface to equalize the concentration across the surface of the droplet. The moving surfactant drags some continuous phase back into the gap and forces the approaching particle apart. Note these processes do not happen in sequence, rather the second process tends to oppose the first and acts to stabilize dispersions

tion or coalescence rather than the approaching particles shearing off and missing one another (Fig. 8.26). Combining the collision frequency, *F,* with collision efficiency, *E* (=0 for no effective collisions or 1 if all collisions are effective) gives an expression for the change in particle number with time:

$$\frac{dn}{dt} = -\frac{1}{2}FE \qquad (8.16)$$

Thus the rate of particle aggregation can be reduced by either reducing the collision rate or reducing the collision efficiency.
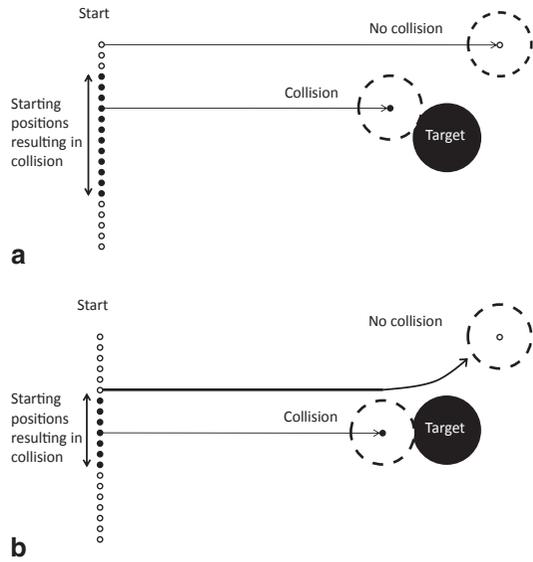
Fig. 8.26 Colloidal and hydrodynamic forces can reduce the number of effective collisions. *Circles* at the start position show the centers of a series of particles moving left to right across the page toward a fixed *target particle*. The *filled particles* are on a trajectory leading to an effective collision while the *open particles* will miss their target. **a** No forces are acting and a higher proportion of the starting positions lead to effective collisions than **b** when repulsive forces act. Strong interparticle attractions could lead to an even higher proportion of the trajectories to be effective

encounter → surface → pore → contents
& interaction   flattening   formation   merging
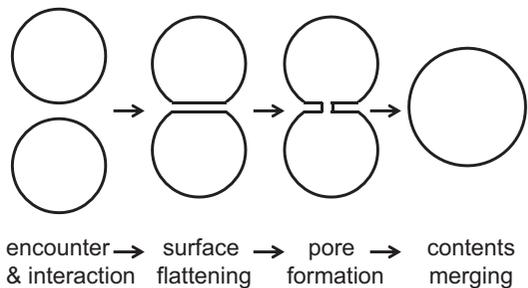
Fig. 8.27 Processes involved in the coalescence of fluid droplets

If flocculation is the encounter followed by sticking of particles, then coalescence is the third step in which fluid particles (i.e., foams and emulsions) merge. The process of coalescence occurs in four stages (Fig. 8.27):

• Droplet encounter and interaction. As described above, the droplets must first encounter one another via Brownian motion, flow induced

collisions, or collisions during gravitational separation. The colloidal and hydrodynamic forces determine if a given encounter will lead to an effective collision. At the end of the interaction process, the particles are close to one another but separated by emulsifier and a thin layer of continuous phase.

- Surface flattening. If the droplets are pressed together by centrifugation or by strong attractive colloidal forces, and if the Laplace pressure is low (i.e., low interfacial tension and large droplet size) then the particle may flatten at the point of contact. This effect is widely seen in the polyhedral bubbles that form in a coarse foam.

- Pore formation. A small pore forms in the flattened plateau border between droplets. The chance of a pore forming is proportional to the area of the plateau border but is often quite low and particles can remain in the previous stage for an extended period as a floc or as a creamed layer. In a protein-stabilized system, pore formation is often due to mixing forces causing the interfacial layer to tear. In surfactant-stabilized systems, pores can form spontaneously depending on the preferred angle of curvature of the surfactant molecules at the oil water interface.

- Contents merging. Once a pore forms, the contents of the two fluid droplets merge to equalize the surface curvature across the surface. This process is typically extremely rapid unless the contents are highly viscous (e.g., imagine two bubbles stuck together as a doublet with a flat plateau border, suddenly they will "pop" apparently instantaneously into a single spherical bubble). One interesting counter example is seen in the partial coalescence of semi-crystalline oil droplets. The crystals from one droplet penetrate the second droplet (i.e., pore formation) but the solid fat network prevents the droplets changing shape and fully merging and instead large aggregates of partially coalesced lipid droplets form. These clumps of fat are the first stage of butter formation during the churning of cream and are also important in the stability of whipped cream and ice cream.

8.5   Dispersion Rheology

Dispersions are always more viscous than their continuous phase and may form gels as the volume fraction approaches close-packing or as a result of extensive flocculation. The rheology of liquid dispersions can be readily understood using modified versions of the theory introduced in Chap. 7 and we will return to the properties of dispersion gels in the next chapter.

In Chap. 7, we understood the viscous properties of polymer solutions by treating the polymer coils as hard spheres. Fine fluid particles resist deformation and so are hard spheres. Larger fluid particles may deform to form nonspherical shapes and solid particles may be nonspherical and resist deformation, but the hard sphere approximation is a good place to start. We can rewrite the expression for polymers (Eq. 7.5) in terms of the volume fraction of the dispersion:

$$\eta = \eta_s \left(1 + [\eta]\phi\right) \qquad (8.17)$$

where η and η_s are the viscosity of the dispersion and its continuous phase and $[\eta]$ is the intrinsic viscosity of the particle. Intrinsic viscosity is a characteristic of particle shape and for spherical particles has a value of 2.5.

Once again, the Stokes–Einstein equation is valid only for very dilute dispersions when the particles do not interact with one another. Viscosity increases at an increasing rate with volume fraction until reaching a critical value when the dispersion stops flowing and starts behaving as a solid (e.g., cream is more viscous than milk because of the higher volume fraction of fat droplets and mayonnaise is a solid because of its very high volume fraction). A good example is the solid-like head on a pint of beer, the dilute foam in the beer itself is liquid but in the closely packed bubbles of the head mean the concentrated foam is a solid.

In a dilute dispersion, viscosity is due to the flow of the continuous phase past the particles but as the concentration increases the particles have to flow past one another meaning more work is needed to achieve the same rate of flow
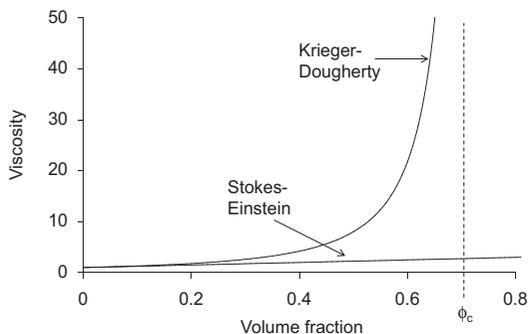
Fig. 8.28 Stokes–Einstein and Krieger–Dougherty models of dispersion viscosity as a function of dispersed phase volume fraction. Viscosity of the continuous phase is 1

(i.e., higher viscosity). Eventually the particles are so closely packed that they jam one another's movement completely and the dispersion behaves as a solid. We used a similar argument in Fig. 8.22 when higher volume fractions were shown to slow creaming rates; here the particles are moving relative to the continuous phase not under gravitational forces but as a result of applied shear forces. The properties of solid materials are treated in much more depth in the following chapter but for now the simple physical picture of a material that cannot flow should suffice.

The viscosity of more concentrated dispersions can be modeled by adding higher-power terms in volume fraction (Eq. 7.5) or using semi-empirical models, for example, the Krieger–Dougherty relation:

$$\eta = \eta_s \left(1 - \frac{\phi}{\phi_c}\right)^{-[\eta]\phi_c} \qquad (8.18)$$

where ϕ_c is a critical volume fraction. At low volume fractions, Eq. 8.18 reduces to Eq. 8.17 (Fig. 8.28).

There are no terms in particle size in Eq. 8.17 or 8.18. According to this theory, the volume of particles determines the viscosity of the suspension but not their size. Thus, we would expect operations such as homogenization and grinding to have no effect on rheological properties. In practice though, smaller particles and nonspherical particles have a greater specific interfacial area,

and thus more energy is lost due to friction between stationary particles and a moving continuous phase and the viscosity is higher, particularly at higher volume fractions. Far more important than the size of individual particles though, is the size of and structure of flocs formed from aggregated particles.

Figure 8.5 shows a set of particles present either free as isolated spheres or aggregated into either a dense or a loose floc. Even though the total volume of the particles is equal in each case, the viscosity of the sample would decrease open floc > dense floc > isolated particles. The reason for the difference is that when the fluid flows, the floc moves as a largely intact object (i.e., a sphere with radius equal to the radius of gyration of the floc, Fig. 8.5). The sphere can then be treated as the "effective" particle for rheological purposes and, because it has a higher volume than the primary particles it is formed from, the viscosity is greater. Furthermore the more open floc entrains more continuous phase with the same number of particles, so its viscosity is the highest of all. We have previously used a similar argument to explain why the same mass of polymer has a higher viscosity in a good over a poor solvent (Sect. 7.7). It is worth noting in passing that if the flocs get very large, they can span the container of the fluid. At this point, the dispersed particles have formed a network that holds all of the fluid continuous phase together and the food will behave as a solid gel. Foods like yogurt (i.e., a network of casein micelles) and butter (i.e., a network of fat crystals) are so-called particle gels formed from the controlled aggregation of dispersed particles and will be discussed in greater length in the next chapter.

Most dispersions are highly non-Newtonian, and usually shear thinning and thixotropic. Shear thinning behavior is associated with the fluid responding to the applied forces by rearranging its structure to present less resistance to flow. Isolated particles in a dispersion can change their shape and, if they are nonspherical, their orientation or form streamlines in response to flow (see Figs. 7.22 and 7.23). However, most solid particles are present as a result of a grinding operation and will only be deformed or broken under
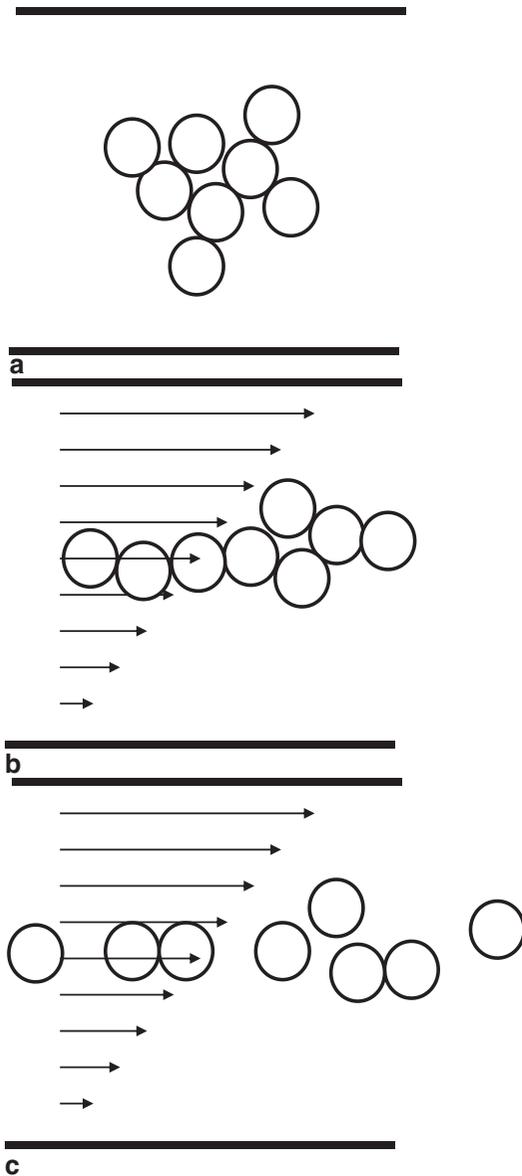
Fig. 8.29 Effect of shear flow on the arrangement of particles in a floc. A large floc under static conditions (**a**) deforms and aligns with the flow (**b**), and may eventually fracture (**c**). The extent of the deformation of the floc depends on its size, and the strength of the bonds holding the structure together as well as the duration and speed of the induced flow

mation is only likely to be responsible for shear thinning at very high shear rates or if the mechanical strength of the particle is very low. Formation of streamlines may contribute to shear thinning in a nonflocculated system, but if flocs are present, their rearrangement under shear is also important. While individual particles tend to be too strong to be deformed in flow, the bonds holding particles together in flocs are often weaker and can rearrange in response to an applied force. The elongation, alignment, and eventual fracture of a floc leads to progressively less resistance to flow and reduced viscosity (Fig. 8.29). When the shear rate is decreased or the flow stopped, flocs may reform to some extent and Brownian motion will move particles out of alignment so viscosity will recover, at least partially, over time.

Example: Effect of Lecithin on the Properties of Fluid Chocolate

Lecithin is widely added to chocolate to reduce the viscosity of the molten product and makes it easier to pump and mold. The lecithin is believed to adsorb on the surfaces of sugar particles and disrupt flocs. Arnold et al. (2013) dispersed finely ground sugar in vegetable oil as a model for liquid chocolate and measured the flow curves with and without added lecithin (Fig. 8.30). The sugar dispersions were
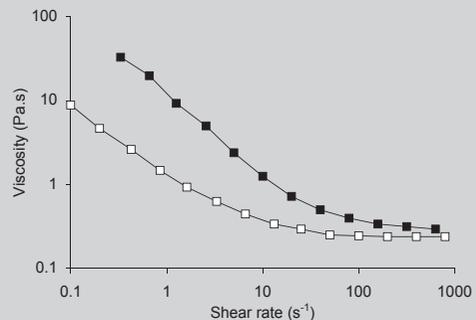
Fig. 8.30 Flow curves of sucrose in soybean oil dispersions (31 wt%) with (*open points*) and without (*filled points*) added lecithin (3.25 mg/m^2 of sugar surface). Adapted from Arnold et al. (2013)

similarly large stresses. Similarly, small fluid particles remain spherical as a result of their surface tension and their surface curvature means the contents are pressurized and will resist deformation under most flows. Consequently, particle defor-
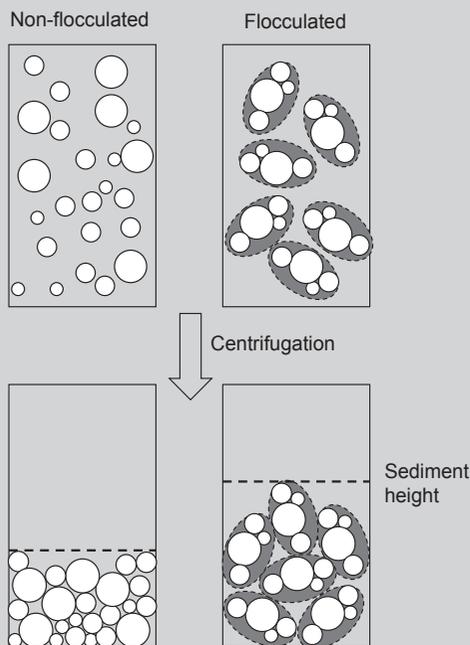
Fig. 8.31 Flocculated dispersions (shown here as *shaded ellipses*) give higher sediment volumes

Fig. 8.32 a Atomic force microscope used to measure the forces acting between two sugar crystals in oil. **b** Sketch of a force distance plot. Adapted from Arnold et al. (2013)

shear thinning, consistent with flocs breaking up under increasing hydrodynamic forces. Lecithin reduced the viscosity but the effect was greater at lower shear rates; presumably the lecithin could disrupt flocs present at low shear rates but after the particles de-agglomerated at higher shear rates the lecithin had less effect. However, this is just one interpretation of the bulk rheological data and one that required more structural evidence.

The structure of flocs can be studied by measuring the volume of sediment formed following centrifugation of the suspension. Highly flocculated dispersions trap continuous phase and so have larger sediment volumes than nonflocculated dispersions or dispersions with weak flocs that can be crushed and broken during sedimentation (Fig. 8.31). In this case, increasing the lecithin concentration decreased the sediment volume consistent with it reducing the forces holding the flocs together. These forces could be measured directly by gluing

a sugar crystal to the tip of an atomic force microscope (AFM) tip and moving it first toward a second crystal fixed in an oil bath and then away (Fig. 8.32). The AFM cantilever will bend if forces act between the two crystals and the bending can be measured using a laser and used to calculate the force as a function of separation. As the crystals were brought together, there was no deformation of the cantilever until they came into contact. At that point, the cantilever bends linearly as the two solid surfaces are pushed together. As the crystals were moved apart again, the cantilever first straightened then bent the other way because the crystals were stuck together. Eventually the strain on the cantilever becomes too great, the crystals detach and the cantilever snaps back to straight. The dynamic measurement of the instrument means force–distance plots are not the same as the equilibrium energy–distance plots in Fig. 8.6 although the net force for detachment should be related to the depth of the energy minimum. This profile shows that there were no attractive forces acting (i.e., the force–distance curve was flat on
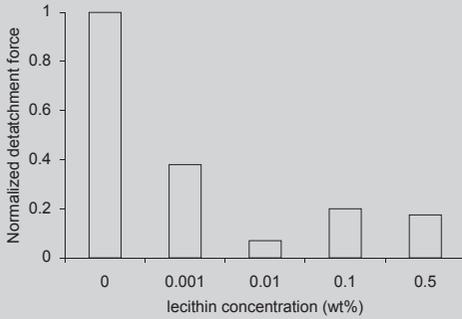
Fig. 8.33 Attachment force between two sugar crystals in oil as a function of lecithin concentration. Adapted from Arnold et al. (2013)

approach) but there was a large force associated with separating the crystals once they had been brought into contact (i.e., the pull-off force in Fig. 8.32). The magnitude of the pull of force decreased with the lecithin concentration (Fig. 8.33). Taken together these data support a picture of lecithin adsorbing to the sugar surfaces and making them easier to deagglomerate during flow.

8.6 Summary

The concepts used to understand the behavior of dispersions have some parallels in our earlier studies of molecules. The stability of a solution depended on the intermolecular forces and entropy: the stability of a dispersion depends on inter-particle and gravitational forces but as the numbers involved are small, the mixing entropy of particles is not significant. The rheology of particles and their aggregates were described using the same picture of spheres and expanded chains in a continuum we used to understand the rheology of polymers. For both polymers and dispersions, high concentrations and strong interactions led to increases in viscosity, non-Newtonian behavior, and eventual solidification. The final chapter will consider the liquid to solid transition more closely as the properties of food gels.

8.7 Bibliography

The properties of colloidal dispersions in general are the subject of several more advanced books. "The Colloidal Domain" (Evans and Wennerström 1994) and "An Introduction to Interfaces and Colloids" (Berg 2010) are quite accessible while Hunter's (2001) "Foundations of Colloid Science" and Russel et al.'s (1992) "Colloidal Dispersions" are more in depth but challenging. The rheology of dispersions is covered in Chap. 10 of Macosko (1994) as well as in most of the other books mentioned in this bibliography.

The two outstanding books in the field of food dispersions are Dickinson's (1992) "An Introduction to Food Colloids" and McClements' (2004) "Food Emulsions: Principles, Practice and Techniques." Especially useful features of McClements book are the use of quantitative equations to describe many phenomena discussed qualitatively here, much more detailed description of colloidal forces, and the discussion of instrumental approaches to measure the structure and properties of emulsions. In Chaps. 9 and 13 of "Physical Chemistry of Foods," Walstra (2003) approaches similar problems from a somewhat different perspective. All of these resources also cover the formation of dispersions, a topic not considered here.

Gels

9

John N. Coupland and E. Allen Foegeding

9.1 Introduction

Cooked egg white, frankfurters, yogurt, jellies, and gelatin desserts are traditional food products that are gels. All start as some liquid mixture of ingredients in water which solidify as a result of processing. Hard-boiled eggs are one of the simplest; by placing whole eggs in boiling water for 5 min, you cause the liquid suspension of proteins in the egg albumen to denature, aggregate, and form a gel network. Changing the cooking time allows some control of the texture of the product. Another interesting egg product is the "thousand-year-old egg." Strong black tea, salt, lime, and wood ash are combined into a paste and buried in soil. After 100 days in a cool dark place, the egg white forms a translucent gel. Different processing conditions convert the same protein into gels with different properties. Other food ingredients form more complex gels, for example, in making cheddar cheese, the initial process involves forming the liquid milk into a soft-solid curd by starter cultures lowering the pH, and by the enzymatic action of chymosin. The gel is weak and fractures with minimal deformation. This initial gel is further processed to remove whey (primarily water) and connect the curds, thus forming cheddar cheese. So what are the important shared structures and properties of gels? Gels have two characteristic features (Flory 1974). First, a gel is a mixture of components not all of which contribute to the solidity. A classic example is a gelatin dessert where as little as 2 % protein renders the 98 % liquid water into composite material that behaves like a solid. The solid-like component forms a network that traps the liquid-like component and gives the overall structure a solid texture. Note, this definition excludes many solid materials (e.g., crystals or glasses) which are one-phase solids but not gels.

The second characteristic feature of a gel is it is a solid. However, there are a range of solid-like behaviors. Cooked egg white is a hard gel while Jell-O™ is much softer. In this context, hard and soft generally refer to a lot or a little force respectively needed for deformation. Cranberry sauce from a can will stand up on a plate for hours and still retain the ridges from the can (at least the "traditional" American version of the recipe will) while a starch-based pudding may initially maintain the shape of the pot but will quickly sag. It is often hard to find the words to differentiate one texture from another but any difference from our expectations will usually be unpleasant in the previous chapters, we were forced to think more deeply about fluid flow and define viscosity as a characteristic property of liquids; to understand the properties of gels; we must think more deeply about the deformation of solids and their characteristic property, elasticity.

J. N. Coupland (✉)
Department of Food Science, Pennylvania State
University, University Park, PA, USA
e-mail: jnc3@psu.edu

E. Allen Foegeding
Department of Food, Bioprocessing and Nutrition Sciences,
North Carolina State University, Raleigh, NC 27695

The narrative of the chapter will bounce backward and forward between these two essential statements of structure and function. We will begin with networks as the essential feature needed for solidification. Next, we will turn to rheology to describe and measure the properties of the networks. We will then try to explain the rheological properties in terms of the microstructure for small nondestructive deformations and large deformations where the gel is broken up.

9.2 Network Formation

A universal aspect of all food gels is that some process is initiated that turns a liquid solution or suspension into a solid. A simple mechanical definition of a solid would be if you pushed or pulled on one side of it, the force would be instantaneously transmitted and the other side would move (Push this book with your finger and the mechanical force is transmitted through the structure of the paper and the whole object moves). A gel is mainly liquid, so it will only behave as a solid if the tiny fraction of solids can act together in some way to transmit force across the bulk. Before gelation, those solids must be dispersed in the liquid phase, then to solidify, they must somehow associate in a way that gives the network its mechanical properties.

In developing models for gel formation, we will consider two extremes of structure for the dispersed solid elements: spherical particles and extended polymers (Fig. 9.1). These models correspond fairly well to polymers in a poor solvent (e.g., the globular proteins in raw egg white) and in a good solvent (e.g., gelatin or starch in hot water), respectively (Chap. 7). The particle model also provides a good framework for gelation in dispersed systems (e.g., emulsion droplets, fat crystals, casein micelles, see Chap. 8). Real foods are more complex, but the simple models are a useful starting point.

The particle model (Fig. 9.1a, left-hand column) begins with isolated particles suspended in a fluid. There are no connections between the particles, so pushing or pulling one side of the container would just cause the liquid to flow.

The total volume of particles affects the viscosity of the liquid (see the discussion on the Stokes–Einstein equation in Sects. 7.7 and 8.5), but it remains a liquid. The next stage is to allow some limited bonds to develop, connecting the particles (Fig. 9.1b). We will return to the physicochemical basis for these connections later, but for now imagine them as some form of chemical bond. A few bonds make the fluid more viscous (see Sect. 8.5 for a discussion of the rheology of flocculated dispersions), but there is still no way for the force to transmit instantaneously through the material. However, adding yet more connections allows a network to form that spans the container and provides a pathway for mechanical force to be transmitted instantaneously through the material (Fig. 9.1c). The open circles show particles that are part of a structure-spanning network; pulling the left-hand wall would pull on this strand of particles until force is transmitted to the right-hand wall. The critical degree of interconnectedness needed to form a solid network is the percolation threshold.

A similar model is presented for polymers in the right-hand column of Fig. 9.1. The noninteracting polymers form a viscous solution (Fig. 9.1a), while limited interactions (Fig. 9.1b) increase the viscosity, and extensive interactions form a solid network (Fig. 9.1c). Once again the chain involved in transmitting force from left to right is highlighted with *thicker lines*.

In both models, the key property of a gelled system is the formation of a network of solid-like particles/polymers that span the container (i.e., percolation). What are the conditions required for molecules to interact and form such a network?

First, there must be a minimum amount of gelling material initially dissolved or dispersed in the solvent. The minimum concentration of compact particles required for gelation is typically larger than for extended polymers, simply because each unit takes up less volume (e.g., gelatin is a disordered coil in hot water and gels at about 2 % protein; egg white is about 12 % globular protein and will not gel if significantly diluted before cooking). Increasing the concentration beyond the minimum will increase the strength of the gel formed. If there are insufficient particles/

Fig. 9.1 Models of a
network from particles
(*left-hand column*) or
polymers (*right-hand
column*) showing **a** no
connections (viscous
liquid), **b** limited
connections (increased
viscosity), and **c** extensive
connections (gel).
Connections are shown as
small open circles. In **c** a
network of particles has
formed (shown as *open
circles*) allowing force to
be transmitted from one
side of the container to
the other. Similarly, in
the polymer model, the
polymers shown with
thick lines have formed
a network that spans the
container. Note that in
both cases, although the
network is supported
by connections other
connections link chemi-
cally identical elements of
the structure that do not
contribute to the network
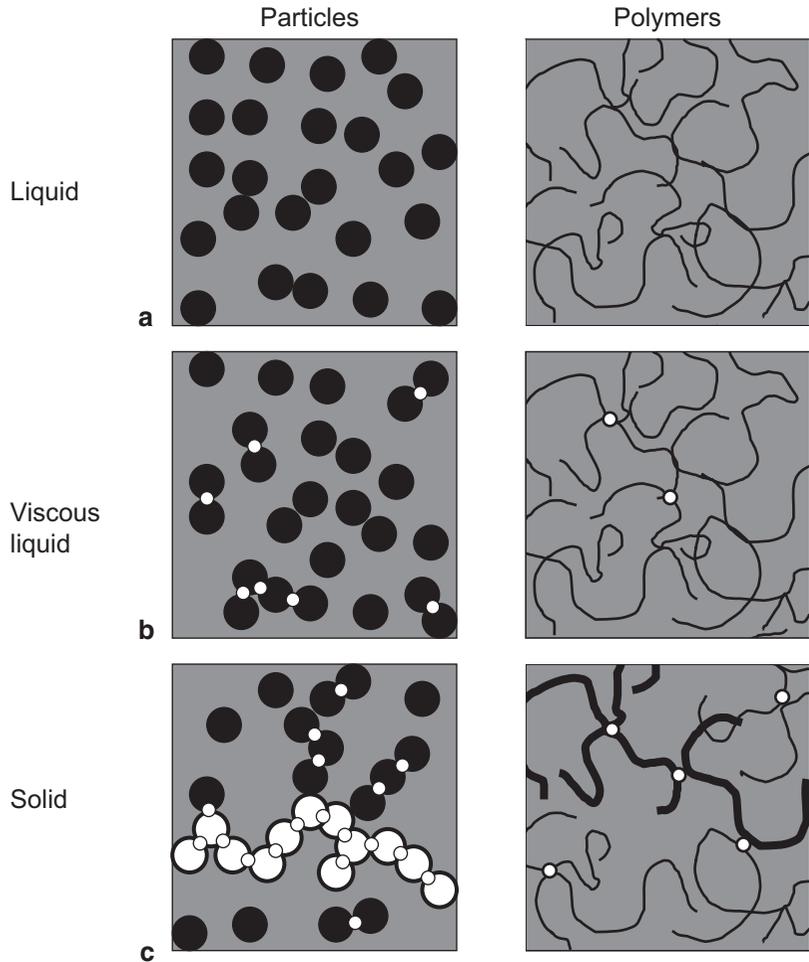
polymers, then their association will lead to a
change in sample viscosity but not to a gelled
network. For example, skim milk is usually con-
centrated from about 9 to 14 % solids before it is
cultured to form a yogurt gel, as the gel from un-
modified milk is too weak and tenuous. In other
cases, gelling polymer may just be insufficiently
soluble to reach the critical concentration for ge-
lation. For example, if you add gelatin powder to
cold water it simply sinks and forms a hydrated
paste. In order to form a gelatin gel, the water
must be hot to allow the polymer to properly dis-
solve so it can gel on cooling.

The second condition required for gel formation
is that, there must be connections formed between
the gelling elements. At least some of the connec-
tions must act as branch points by linking three or
more structural units together. Figure 9.2 shows
examples of particles with different coordination
numbers; only when a few junction points are in-
cluded does a gel form. However, in practice most
food gels are formed without such rigidly defined
stoichiometry and these concerns are more relevant
for synthetic polymer gels supported by covalent
bonds.

The connections must form after the initial
sol is formed, which implies some change in the
intermolecular interactions to form either a cova-
lent or noncovalent "bond."[1] Gels supported by

---

[1] In some cases, the knots and tangles formed in concen-
trated polymer solutions can lead to gel-like behavior
(e.g., shower "gels"). However Clark and Ross-Murphy
(2009) argue that these should be regarded as viscous,

Fig. 9.2 Schematic diagram showing the structures formed between gel elements with **a** zero, **b** one, and **c** two connections with one another. None of these are capable of forming a gel until **d** a few three-connection pieces are added to add as branch points (shown as *arrows*). The gelling elements (represented by *dark circles*) could either represent particles in a particle gel or monomers in a polymer gel. Bonds are represented by *small open circles*
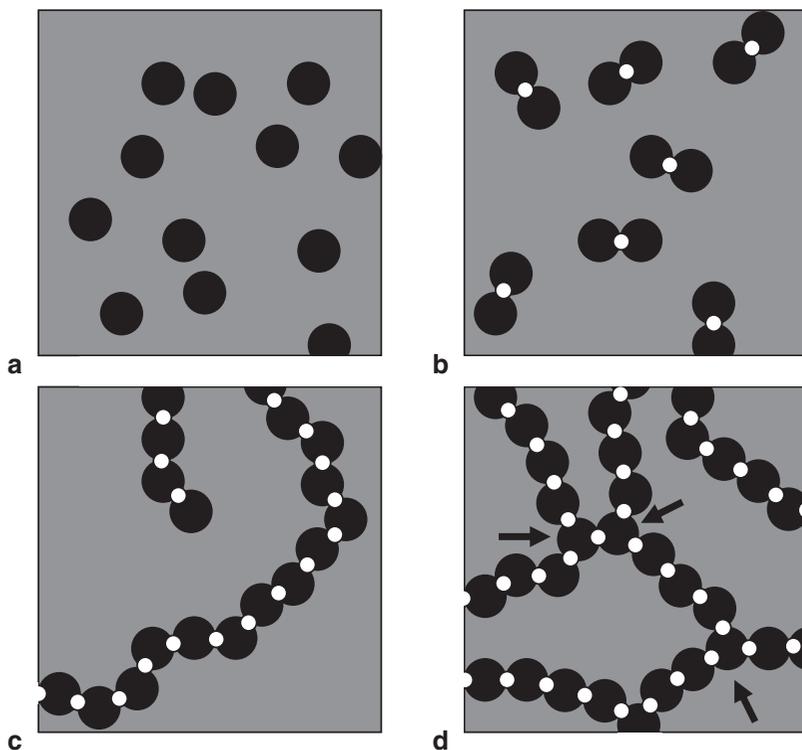
covalent linkages are called chemical or strong gels whereas gels supported by noncovalent attractions (e.g., Van der Waals, electrostatic, hydrophobic) are called physical or weak gels because the bonds are much weaker than the covalent bonds. The terms "strong" and "weak" are used here to describe the strength of the bonds, not necessarily the strength of the gel (i.e., the force per unit area needed to fracture). Because noncovalent bond energies are only a few times the thermal energy, they are readily affected by changes in temperature. Similarly, small changes in pH or ionic composition of the solvent can affect the significance of electrostatic interactions. Therefore, physical gels formed from them tend to be reversible and can be disrupted and re-formed again as changes in solvent conditions favor either solute–solute interactions (leading to gelation) or solute–solvent interactions (leading

to dissolution). Gelatin, agar, and carrageenan gels are examples of physical gels that are formed by heating a solution to unfold the molecules then cooling the solution so the molecules can refold and interact by hydrogen bonding. In contrast, methyl cellulose is a physical gel that forms a gel network when heated, then is disrupted and re-forms a solution when cooled. In this case, hydrophobic interactions, which are favored by elevated temperatures, are responsible for forming a gel network. Most food gels are largely physical gels but are often reinforced by covalent bonds. For example, gelatin-based desserts are physical gels, but what about cooked egg white and hot dogs? If we reheat a hot dog, does it reverse and turn back into a meat protein solution? The obvious answer is no, indicating that it is behaving like a chemical gel. However, proteins also interact by hydrogen bonds and hydrophobic interactions, so a hot dog is best described as a combined chemical and physical gel.

A single covalent bond can be strong enough to be mechanically important in a gel but creating the same mechanical strength with noncovalent

---

non-Newtonian fluids rather than gels, because they will flow, however, incredibly slowly. Other workers disagree (Raghavan and Douglas 2012), and in practice it may not be useful to make the distinction.

interactions requires a series of weaker bonds acting together. Consequently, physical gels tend to be supported by junction zones extending over a significant length of the polymer rather than the single point interactions characteristic of chemical gels (Fig. 9.3). For example, the polymers in starch tend to coil together in double helices supported by hydrogen bonds. In hot water, the helices break down and the polymers behave as disordered coils; but as the solution is cooled the polymers begin to reform their double helices, which act as junction zones in the starch gel. In a chemically homogeneous polymer, there is nothing to stop the continued growth of the junction zones within the gel. The process is slow because the polymers are large and the solution viscous, but potentially continues over the lifetime of the product. For example, as a starch gel ages, it becomes firmer as the junction zones grow. Extensive growth of the junction zones can force the solvent out of the gel as the polymers interact with one another so much they exclude water (i.e., syneresis), for example, the pool of serum that often forms at the surface of a set yogurt after a spoonful has been removed. In other cases, heterogeneity in the polymer composition can limit the spread of junction zones. For example, starch can be chemically modified to add phosphate groups at intervals along the chain. The phosphate groups cannot be readily incorporated into a junction zone to arrest their development. These so-called stabilized starches form long shelf-life gels and are not prone to syneresis.
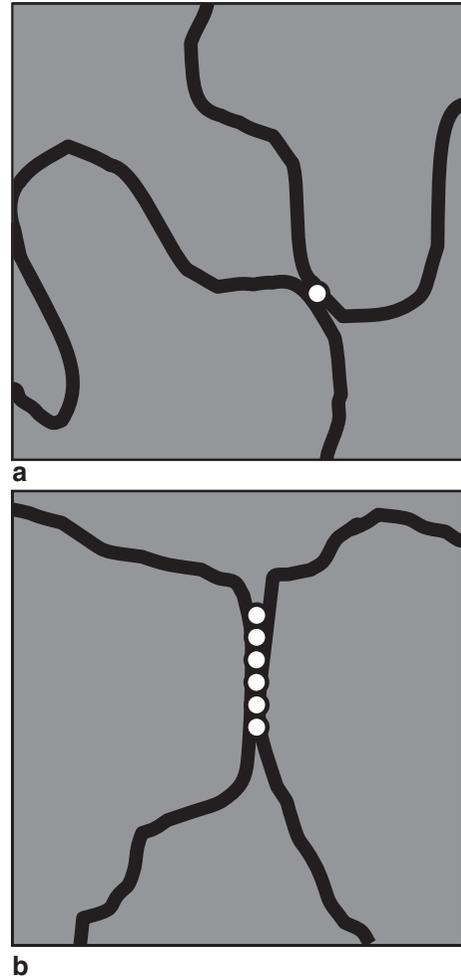
Fig. 9.3 a Covalent and **b** noncovalent bonds (shown as *filled circles*) between polymer molecules. A single covalent bond is strong enough to form a mechanically important junction between two polymer molecules while weaker noncovalent interactions tend to exist over extended regions of the polymer (i.e., junction zones)

9.3   Gel Rheology

Solids testing is a staple of school science labs and many of us have performed experiments where we hung weights from a piece of rubber or wire and measured how much it stretched (Fig. 9.4a). The results depend on the size and shape of the sample as well as its intrinsic properties. A thicker sample of the same material would give less deformation for a given weight and a longer sample of the same material would stretch more. In order to move from the specific, empirical results of a particular test to a characteristic

and fundamental property of the material, we calculate stress as force applied normalized to the cross-sectional area ($\sigma = F/A$, Nm^{-2}, or Pa) and strain as the relative deformation ($\varepsilon = \Delta l/l$, i.e., change in length divided by initial length, dimensionless). This is illustrated for a small element of the sample in Fig. 9.5 and leads to the definition of Young's modulus of the solid, E (Pa):

$$E = \frac{\sigma}{\varepsilon} = \frac{F}{A} \cdot \frac{l}{\Delta l}. \tag{9.1}$$
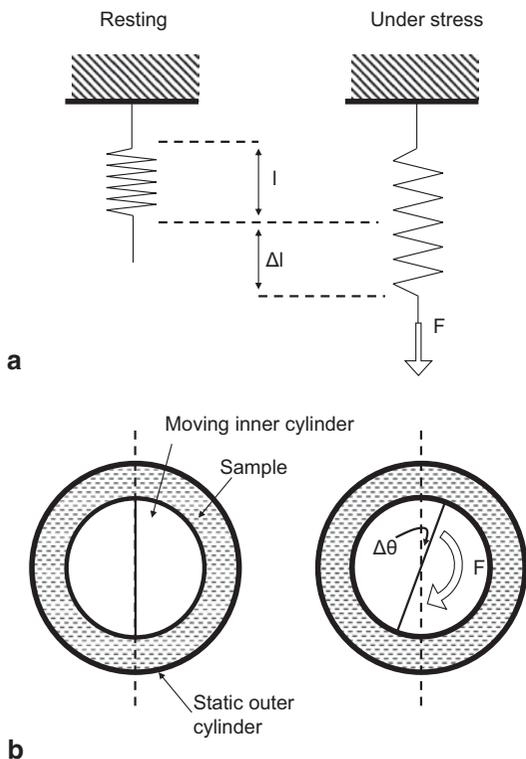
Fig. 9.4 Comparison of **a** the stretching of a spring (*side view*) and **b** the shear deformation of a solid in a concentric cylinder rheometer (*top view*). The applied force (*F*) causes a characteristic change in length (Δ*l*) or rotation (Δ*θ*)

Fig. 9.5 Change in shape of a volume element of an elastic solid in response to an elongational and a shear stress. Note the deformation is assumed small so any change in sample cross-sectional area on elongation is negligible

The elastic modulus is the characteristic property of the solid and is independent of the size and shape of the sample used in the experiment. At small deformations elastic modulus is a constant, but at larger deformations, beyond the elastic limit, it will change as the structure of the material is disrupted prior to fracture (Fig. 9.6 and also see Sect. 9.5).

In most gels testing, it is easier to study shear deformations and use a concentric cylinder rheometer similar to the one we described in for liquids testing in Chap. 7. The design and operation of the rheometer is described in Fig. 7.15, but here we will fill the gap between the cylinders with a gel (Fig. 9.4b). This is experimentally a little more challenging than filling it with a liquid and in practice we might have to add liquid ingredients and induce gelation in situ (e.g., add egg white or gelatin then heat or cool respectively

to induce gelation). In many cases, it might be easier to trap a thin slice of gel between two parallel plates or use any number of other testing geometries, but we will focus our discussion on the concentric cylinder arrangement as it allows us to maintain direct parallels to our earlier discussion of fluids testing.

We can test the gel by applying a rotational force to the inner cylinder and seeing how it moves. If there were a viscous liquid in the gap, then the inner cylinder would spin at the rate inversely proportional to the liquid viscosity, but as we have a solid gel, it will stretch a certain amount then stop. Once again, the results depend on the size and shape of the sample as well as its intrinsic properties. Larger cylinders would allow the same force to be distributed across more gel to produce less rotation, but a larger gap would allow more rotation (i.e., less deformation per unit volume of material). In order to move from the specific empirical results of this particular experiment to a characteristic and fundamental property of the sample, we calculate stress as force applied normalized to the cross-sectional area (σ, Nm^{-2}, or Pa) and strain as $\Delta x/l$
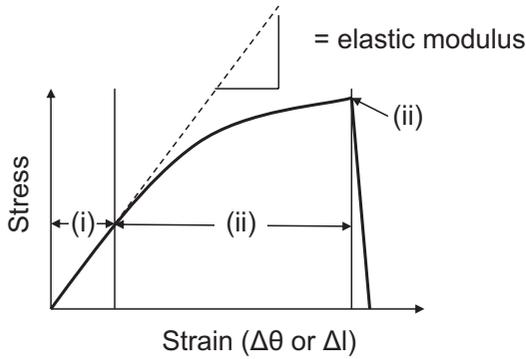
Fig. 9.6 Typical stress–strain relationship for an elastic solid. (*i*) At small deformations, stress is proportional to strain and the slope is constant and equal to the elastic modulus. (*ii*) At larger deformations, the stress–strain relationship is nonlinear and eventually (*iii*) the sample will break

(=tanθ, dimensionless). This is illustrated for a small element of the sample in Fig. 9.5 and leads to the definition of the shear elastic modulus of the solid, G (Pa):

$$G = \frac{\sigma}{\gamma} = \frac{F}{A\tan\theta}. \qquad (9.2)$$

The rotation of the inner cylinder generates a shear deformation in the gel and so measures the shear modulus, distinct from the elongational deformations described above, which determine Young's modulus (Eq. 9.1). The different elastic moduli arise from the stress–strain relationships under different modes of deformation and, while their values are different, all are characteristic properties of the solid.[2] The shear modulus, like Young's modulus, is constant at small deformations, at larger deformations beyond the elastic limit, it will again change as the structure of the material is disrupted prior to fracture (Fig. 9.6, Sect. 9.5).

Although the shear deformations of a solid (this section) and a liquid (Chap. 7), both require measurement of applied force and deformation, there are three important distinctions that should be stressed:

---

[2]A third elastic modulus, the bulk modulus, can be measured as the change in volume under uniform external pressure.

1. The defining relationship for a liquid is between force and rate of rotation (or stress and rate of strain). The defining relationship for a solid is between force and deformation (or stress and strain).
2. The deformation of a liquid will continue to increase as long as force is applied. The deformation of the solid occurs completely and instantaneously as soon as the force is applied and then there is no further movement as long as the force is maintained.
3. If the rotational force applied to the liquid was removed, the inner cylinder would stop moving and stay at whatever position it had reached. If the rotational force were removed during solids testing, the inner cylinder would spring back to its starting position (just as a stretched spring would recover instantaneously and completed once the weight is removed). For this reason, elastic solids are said to store energy, while viscous liquids dissipate energy. The effects of adding and removing a deforming force on an ideal solid and liquid are illustrated in Fig. 9.7.

So can real foods be characterized simply as ideal solids or liquids? In some cases yes: milk is a viscous liquid and tofu is an elastic solid, but in many cases, no. For example, press a piece of dough gently between your fingers and release. It will spring back, but neither instantaneously nor completely (Fig. 9.7c). In some ways, we could describe it as a solid and in other ways a liquid. Characteristically, if we made a measurement a short time after the force was applied (point (i) in Fig. 9.7c), it would be seen as a solid because of the sharp deformation proportional to the force applied. On the other hand, if we made a measurement a long time after the force was applied (point (ii) in Fig. 9.7c) we would see a viscous liquid slowly deforming at a rate proportional to the force applied. Many foods show this combined liquid and solid-like behavior and are described as viscoelastic.

The slow deformation of the sample in Fig. 9.7c is known as creep, and the shape of the curve reveals details of the viscous and elastic aspects of the sample. In a creep experiment, a fixed and constant force is applied instantaneously and
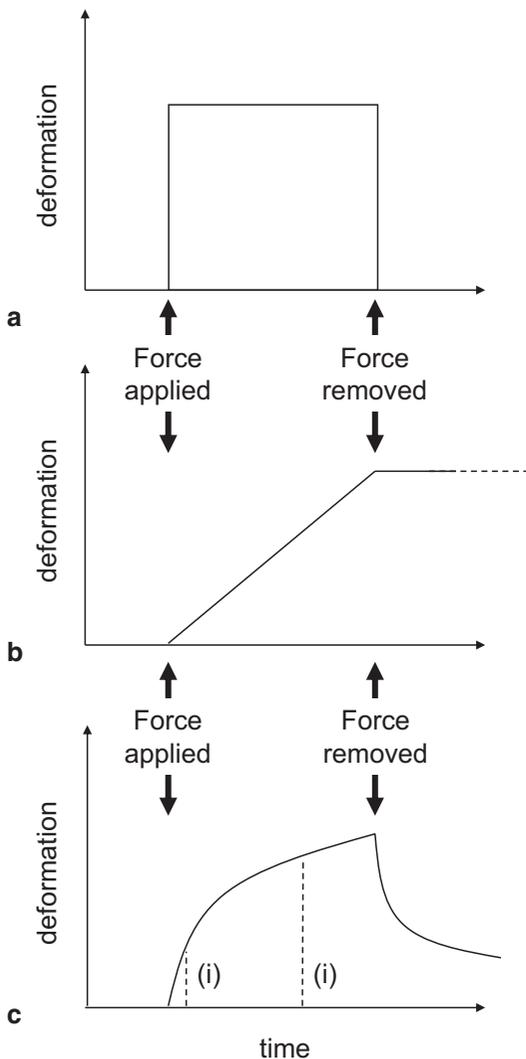
Fig. 9.7 Force applied to **a** a solid will lead to an instantaneous deformation but complete recovery once it is removed, **b** a liquid results in a constant rate of deformation but no recovery once it is removed. **c** A viscoelastic material simultaneously showing behavior characteristics of a solid and a liquid. A characteristic of viscoelastic materials is they behave as liquids over long timescales and solids over short timescales

the deformation measured over time (i.e., controlled stress). Alternatively, a stress relaxation experiment where the sample can be deformed by a fixed amount and the force required to maintain that deformation is measured can reveal similar information (i.e., a controlled strain experiment). However, in both cases, because the timescales of the processes of interest vary widely, it is often necessary to both capture the shape of the curve

at a high resolution in the initial period to capture the fast-moving structural changes and then maintain the experiment for a long time to allow the very slow relaxations to occur. Both requirements are challenging as it requires a delicate instrument to be dedicated to the analysis of a single sample for a long time and that sample is likely to dry out or otherwise chemically change during the test. The viscoelastic properties of the sample are more readily measured in a small deformation oscillatory test.

Oscillatory tests can be conducted in a controlled stress or controlled strain mode, but for the purpose of explanation, we will confine our discussion to the former case and conduct them using the concentric cylinder viscometer, which we have described previously. We will program the inner cylinder to sinusoidally oscillate backward and forward by a small amount from its starting position and measure the force required to achieve that movement. The strain at any instant (γ) is therefore a function of time (t):

$$\gamma = \gamma_0 \sin(\omega t) \quad \text{input strain function} \quad (9.3)$$

(note, the frequency f is expressed as the angular frequency $\omega = 2\pi f$) (Fig. 9.9). The maximum strain is γ_0. For an ideal elastic solid, stress is proportional to strain at any instant, so the resultant stress is also a sine wave with the same frequency as the input strain, and reaching a maximum at the same point (Fig. 9.8):

$$\sigma' = \sigma'_0 \sin(\omega t) \quad \text{elastic solid stress response}$$
$$(9.4)$$

The elastic modulus (G') for the solid can be easily calculated from the ratio of maximum stress to maximum strain:

$$G' = \frac{\sigma'_0}{\gamma_0}. \qquad (9.5)$$

For an ideal viscous liquid, stress is proportional to the rate of strain, i.e., the rate of change of deformation with time or the slope of the input strain function. The rate of strain varies from zero at the limits of deformation when the inner cylinder is instantaneously stationary, to a maximum value halfway between maxima. The measured stress
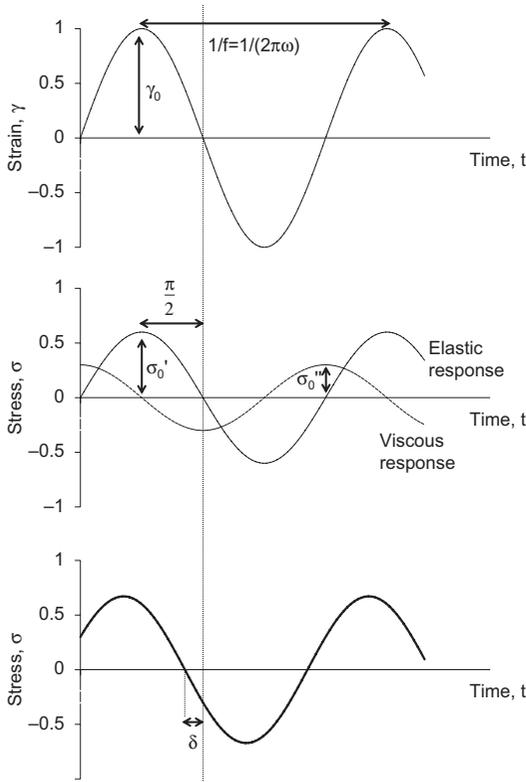
Fig. 9.8 A sinusoidal strain results in an in-phase stress response for a solid and an out of phase stress response for a liquid. The stress response of a viscoelastic material response can be expressed as the sum of a liquid-like (i.e., sine wave) and out of phase (i.e., cosine wave) functions or as a sine wave shifted by δ

for an ideal liquid will therefore also vary sinusoidally at the same frequency as the input deformation but shifted backward (i.e., phase shifted) by 90°, i.e., a cosine wave:

$$\sigma'' = \sigma_0'' \cos(\omega t) \quad \text{viscous liquid stress response.} \tag{9.6}$$

Analogous to the elastic modulus for solids, we can use this wave to calculate a viscous modulus[3] (G'') for liquids:

$$G'' = \frac{\sigma_0''}{\gamma_0}. \tag{9.7}$$

---

[3] The viscous modulus captures the response of a liquid in oscillatory shear but is not the same as the viscosity measured in steady shear. A dynamic viscosity can be calculated by dividing the viscous modulus by angular frequency.
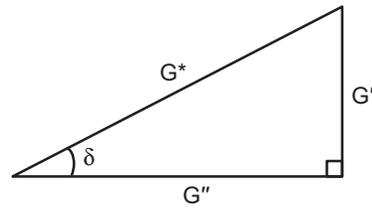
Fig. 9.9 Relationship between the complex modulus (G^*), storage modulus (G'), loss modulus (G'') and phase angle (δ) for a viscoelastic material

A viscoelastic material shows some combination of liquid-like (i.e., viscous) and solid-like (i.e., elastic) properties simultaneously. The stress response can be equivalently expressed as the sum of a sine wave (i.e., elastic response) and a cosine wave (i.e., viscous response) or as a sine wave shifted in time by a certain amount (δ, ranging from zero for an ideal solid to 90° for an ideal liquid):

$$\sigma = \sigma_0' \sin(\omega t) + \sigma_0'' \cos(\omega t)$$

viscoelastic stress response. (9.8)

$$\sigma = \sigma \sin(\omega t + \delta) \quad \text{viscoelastic stress response} \tag{9.9}$$

where $\tan\delta$ is the ratio of the viscous and elastic responses:

$$\tan \delta = \frac{\sigma_0''}{\sigma_0'} = \frac{G''}{G'}. \tag{9.10}$$

From the stress response functions, both an elastic modulus and a viscous modulus can be calculated for the viscoelastic material. The ratio of viscous to elastic modulusis also equivalent to $\tan\delta$, the phase angle. For a purely elastic solid (i.e., $\delta = 0°$) or purely viscous liquid (i.e., $\delta = 90°$), the viscous and elastic moduli, respectively, are zero. Sometimes the condition $G' > G''$ (or $\delta < 45°$) is used to distinguish viscoelastic solids from viscoelastic liquids.

The two moduli can be conveniently expressed as a single complex number, the complex modulus (Fig. 9.9):

$$G^* = G' + iG'' \tag{9.11}$$

where G' (the real part of the number) is the energy stored as the elastic modulus, G'' (the imaginary part of the number) is the energy dissipated as the viscous modulus and $i = \sqrt{(-1)}$. The magnitude of the complex modulus is:

$$G^* = \sqrt{G'^2 + G''^2}. \qquad (9.12)$$

Now that we have seen how a viscoelastic material will respond to oscillatory strain, we can consider the types of measurement that can be conducted to characterize a gel. Just as the elastic modulus in Fig. 9.6 is strain-independent over a limited range, G' is typically constant up to a certain strain then begins to decrease. An essential first experiment in any small deformation measurement is to measure just how small the deformations must be within the linear range, while being large enough to measure. Having determined the linear viscoelastic limit, any smaller strain can be used in further measurements. There are two basic experiments that can be conducted: Gel properties can be measured either as a function of oscillation frequency or as a function of some external variable, usually time or temperature (see example below).

Frequency scanning experiments, sometimes called mechanical spectroscopy, are used to measure how quickly a material can respond to an applied force. Typical spectra for viscous polymer solutions and gels are shown in Fig. 9.10. In a liquid, at low frequencies G'' is greater than G', i.e., the material behaves like a liquid, but increases more slowly with frequency. At a critical frequency, G' exceeds G'' and the solid-like properties begin to dominate and above a certain frequency both moduli reach a plateau. A good example here is the mixture of two parts water to one part granular cornstarch known as Oobleck. If you pour slowly then it flows like a viscous dispersion, but if you pump it quickly it locks up and behaves like an elastic solid.[4] Over long timescales (low frequencies) the material is able to dissipate energy and flow, but over short timescales (high frequencies) the material stores energy as an elastic solid. A high frequency oscillatory

a

b
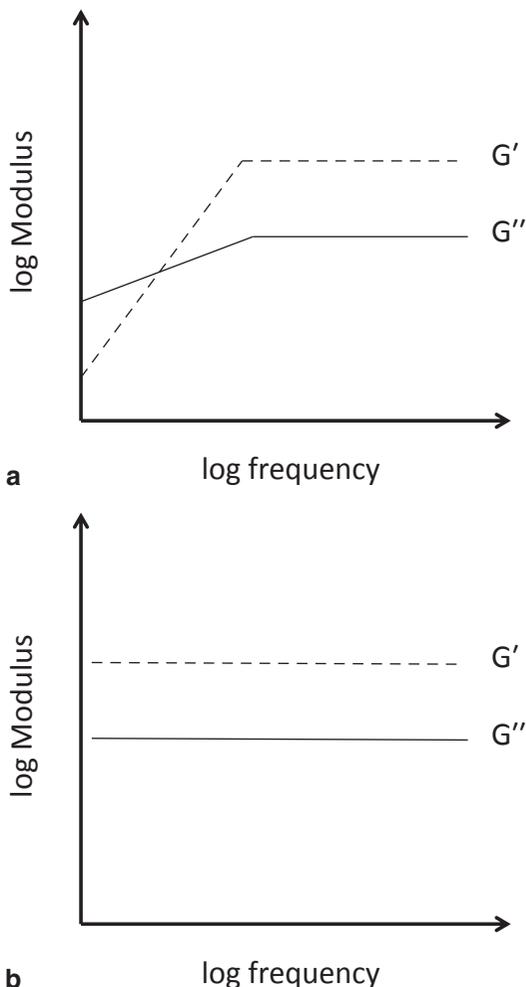
Fig. 9.10 Typical mechanical spectra of **a** a viscous polymer solution and **b** a gel

measurement would reveal similar information as a short timescale creep measurement and vice versa, where the occurrence of transition from solid-like to liquid-like behavior depends on the time it takes for the elements of the material to reconfigure themselves to the applied force and flow (i.e., the relaxation time). In contrast, a "real" gel (Fig. 9.10b), where there are permanent bonds between the polymers or particles that do not reconfigure, whatever the timescale of observation, then both moduli are reasonably constant over a wide frequency range and $G' > G''$. However, depending on the frequency range accessible, the mechanical spectrum of a gel could look a lot like the plateau region for a viscous liquid.

---

[4] A search of videos on the Internet will reveal far more entertaining examples.

Example: Development of Cheese Curd

In cheese making, milk is coagulated to form a curd by the addition of chymosin, a highly specific protease that cleaves a hydrophilic glycopeptide from the κ-casein portion of the casein micelle leaving the micelle surface more hydrophobic and prone to aggregation. Bohlin et al. (1984) used small-deformation oscillatory measurements to follow the formation of the cheese curd. Milk was warmed to 31 ℃ before chymosin was added then the mixture was immediately added to a concentric

cylinder rheometer and oscillatory measurements were made at a small deformation (maximum deformation 0.075) at a fixed frequency of 0.5 Hz. Changes in the viscous and elastic moduli and in the phase angle with time are shown in Fig. 9.11.

At the start of the experiment, the milk is too fluid for the instrument to record any reading, but after about 15 min the storage and loss moduli began to increase. At the time of first measurement, the phase angle is greater than 45° (i.e., storage modulus less than the loss modulus, a liquid-like

Fig. 9.11 a Changes in the rheological properties of milk during coagulation with chymosin. Adapted from Bohlin et al. (1984). **b** Schematic physical model for casein few aggregation. (*i*) After a minute of reaction, some casein has been hydrolyzed leading to limited aggregation but no network formation. (*ii*) After about 15 min at least one aggregate is large enough to span the gap in the rheometer and the first moduli are measured. Casein micelles that serve as elastic connections in the network are shown white. (*iii*) Further proteolysis allows more casein micelles to join the network and contribute to its mechanical properties
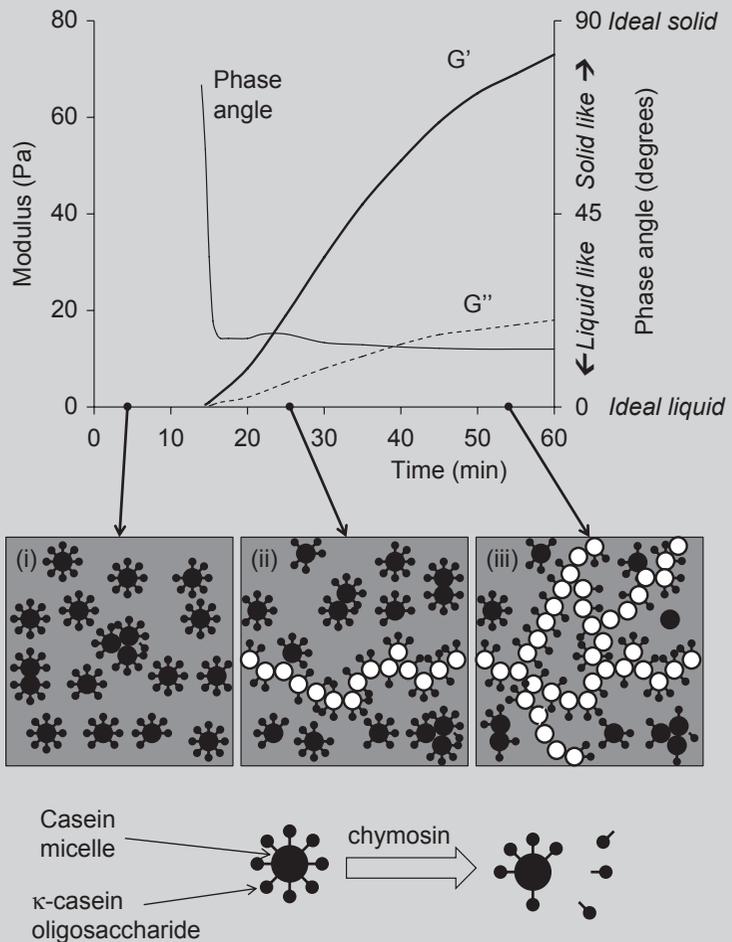
sample) but almost immediately the phase angle decreases and the sample can be said to have solidified. During the later stages of the reaction, both the viscous and elastic moduli continue to increase as more casein join the network. Characteristically for food gels, it never really reaches an equilibrium value and the gel strength is still increasingly slower over the course of the experiment as junction zones continue to form and the strands rearrange.

Because the deformations induced were small, the test can be regarded as nondestructive. Other methods to measure the time it takes for the gel to form (e.g., viscosity measurements in a liquid or texture analysis of a gel) would require a separate sample to be destroyed for each time point.
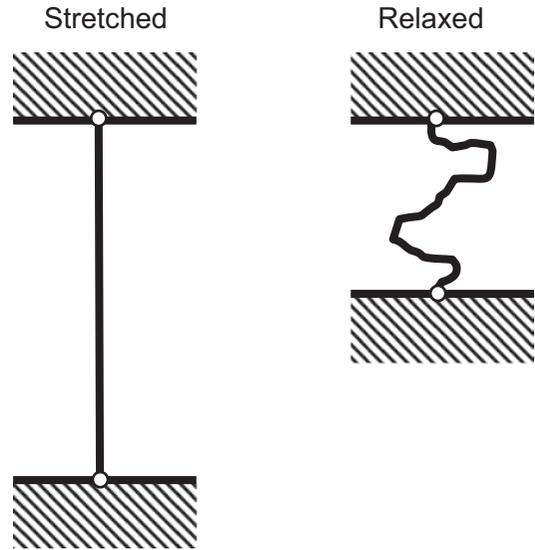
Stretched   Relaxed

Fig. 9.12 A cord of polymer (the section between two junction points in a gel) is stretched to a single conformation when the gel is stretched but has many available conformations when the gel is relaxed. The difference in chain entropy provides the restoring force for gel elasticity

9.4 The Molecular Basis of Elasticity

For small deformations, the elastic moduli describe how the intact gel network stores the mechanical energy of an applied load. Network elasticity is usually understood in terms of two extreme models, one entropic and one enthalpic. Figure 9.12 shows a portion of the network between two junction points, i.e., a cord. The figure is drawn to resemble the polymer gel in Fig. 9.1, but could equally be a "string of beads" between junctions in a particle gel. At rest, the cord is slack and can take on a range of conformations. When the gel is deformed, the cord is drawn tight and is fixed in a linear conformation. Thus, stretching the gel lowers the entropy of the gel and there is a thermodynamic driving force to recoil (see also the discussion of polymer shape around Fig. 7.6, where we argue that the stretched-out form is statistically improbable).

The alternative energetic model for elasticity is based on stretching the bonds connecting elements in the gel. In their resting, unstretched state, two molecules in a critical bond in the gel network have average separation of $s*$ at which

the combination of attractive and repulsive forces gives rise to a minimum free energy (Fig. 9.13, also compare with Fig. 2.19). When the structure is stretched, the bond length increases and the retractive force arises from the drive to return to the energy minimum. In both models, the elastic modulus of a gel is proportional to the number of elastically active chains in the network. To be elastically active, the chain must span the container, for example, in Fig. 9.1c there is one elastically active chain present. Nonelastically active chains can contribute to the viscous properties of the gel.

We can attempt to distinguish between the models in two main ways. Firstly, the amount of deformation before fracture is usually greater in entropic gels. Secondly, the contribution of entropy to free energy increases with temperature ($G = H - TS$, Eq. 1.10), so purely entropic gels get harder at increased temperatures (i.e., the entropy difference between the stretched and unstretched states is the same but multiplied by a larger value of T, the contribution to free energy is greater). Most noncovalent bonds on the other hand be-
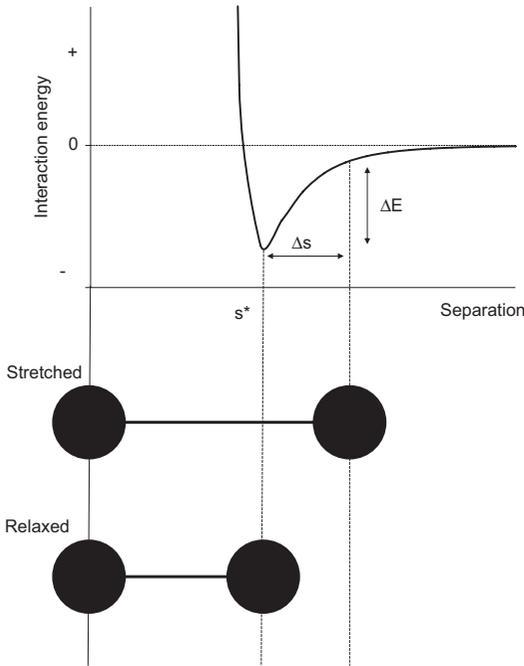
Fig. 9.13 The interaction potential for a bond between two elements of a gel. When the bond is stretched, the intermolecular separation increases from the equilibrium value and the bond energy increases accordingly. The difference in bond energy provides the restoring force for elasticity

Fig. 9.14 A molecular model for viscoelasticity. In the stretched gel, a cord of polymer is pulled tight between two fixed points and held in that position by weak noncovalent interactions. The polymer relaxes over time (i.e., releases the stored energy) by exchanging bonds and "sliding through" the junction zone without the need for contracting

come weaker with increased temperature so we would expect enthalpic gels to soften with increased temperature. An important exception to this rule is gels supported by hydrophobic interactions which are largely entropic in nature and so harden upon heating. In practice, it is difficult to find a food gel that is completely and unambiguously described by either the entropic or enthalpic model and usually some combination of models is required.

An important and unrealistic feature assumed in both models is the bonds themselves are unaffected by stretching. Weak noncovalent interactions and molecular entanglements can reconfigure themselves in a stretched gel to dissipate some of the stored energy without contracting (see Fig. 9.14 for a schematic illustration). The mechanical effect would be that the gel is elastic over short timescales but viscous over long ones, i.e., viscoelasticity. For example, if you pressed a

gel and released immediately, before the internal bonds have had time to reconfigure, it will recoil elastically. On the other hand, if you pressed the gel and held for sufficient time for the entanglements to disentangle and the weak bonds to rearrange to relax the cord, then when you released, the gel would not recoil.

9.5   Larger Deformations, Fracture, and Texture

Food scientists are often interested in physical properties observed beyond the linear viscoelastic region. Indeed, chewing food causes large deformations and subsequent fracture of large particles into small particles. We now move from the field of rheology to consider fracture and start by considering three different stress–strain (force–deformation) relationships for three different gels (Fig. 9.15):

1. An ideal elastic gel with stress proportional to strain up until the point of fracture (i.e., con-
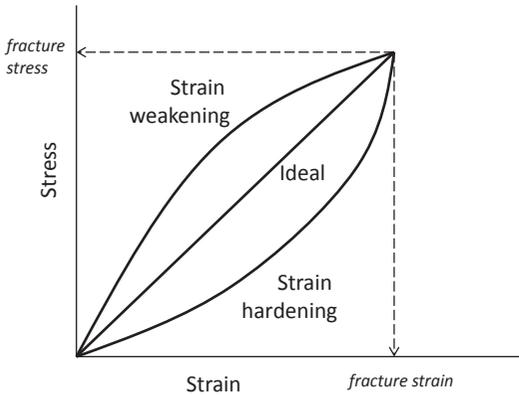
Fig. 9.15 Stress–strain curves for three gels: (i) ideally elastic, (ii) strain-weakening, (iii) strain-hardening. The three gels are shown as having the same stress and strain at fracture but other formulations might break at other points along the curve

stant elastic modulus). Note however that the stress and strain at fracture could be different for different gels even though the modulus is same.

2. A strain weakening gel (e.g., cheddar cheese), the modulus decreases with increasing strain as the structure is weakened by deformation.
3. A strain hardening gel (e.g., gummy candies), the modulus increases with increasing strain as the structural elements become more resistant at greater deformations.

So which gel would be perceived as toughest during chewing and which would be hardest to spread or slice? At the end of deformation, they all have the same fracture stress and strain; therefore, one might conclude they would have similar texture. In reality, the texture of the three gels would be much different and the only way to get a comprehensive fingerprint of the gel would be to test in the linear viscoelastic region, at fracture, and in the zone in-between that we will call the nonlinear region. We should also account for the rate of deformation as most real gels are at least somewhat viscoelastic and will usually appear more solid like (i.e., higher modulus), if deformed quickly.

When an ideally elastic gel is deformed, all of the applied energy (i.e., the area under the stress–strain curve) is stored as deformation of polymer chains or bonds. In a viscoelastic gel, the structure responds to the strain in a time-dependent manner (e.g., by inhomogeneous deformation of particles or chains in the network, or by the flow of water through pores) and some of the applied energy is dissipated. However, once the stored energy exceeds a critical value the gel will fracture, i.e., all the structural elements in the network along a macroscopic plane will break leading the gel to fall into pieces. If the gel is dissipating energy due to viscous flow, then the slower the deformation the greater the loss of energy and the higher the level of deformation required to cause fracture. This often is seen in gels where the primary network is composed of proteins. If dissipation of energy is due to friction, then faster deformation generates more energy loss, and a higher level of strain is needed to cause fracture. Gels with polysaccharide networks tend to show this behavior.

Note that fracture properties do not necessarily relate to the modulus of a gel. Fracture is usually initiated at a macroscopic defect in the gel structure, perhaps an air bubble or an included particle, that concentrates the stress at a point. The stress at this point, rather than the lower overall stress across the entire gel, must exceed the critical value for fracture. The size of the defects is important with fracture initiating at the largest defects present. Van Vliet (1996) suggests the characteristic defect sizes for a range of food gels to be in the scale range of 10–100s of micrometers.

9.6 Summary

The association of polymers or colloids into networks can lead to the formation of elastic gels. The formation of associations depends on changes in molecular or colloidal forces in response to changing solution conditions (see Chaps. 2 and 7 respectively). The initial stages of the liquid to solid transition can be understood in terms of the changing viscosity of polymers or particle sus-

pensions with increasing molecular weight or degree of aggregation (Chaps. 7 and 8), which eventually result in the formation of a deformable elastic network that is the basis of many of the solid foods we eat.

The overall goal of this book has been to describe the physical properties of foods in terms of the interactions between its constituents. We built from molecules to higher order structures and in this chapter allowed them to interact and form

Example: Egg White Gels

Nagano and Nishinari (2001) compared the large deformation and fracture properties of two commercial egg white products. The egg was gelled by heating for different times and the stress–strain relation was measured in compression (Fig. 9.16). In both cases, the curves were slightly shear-hardening and the modulus for product A was greater for product B. The stress and strain at fracture were also higher for A than for B as was the energy for fracture (i.e., area under the curve). Different processing conditions changed the properties of the gel, for example heating to different temperatures increased Young's modulus (initial slope) and the stress at fracture but
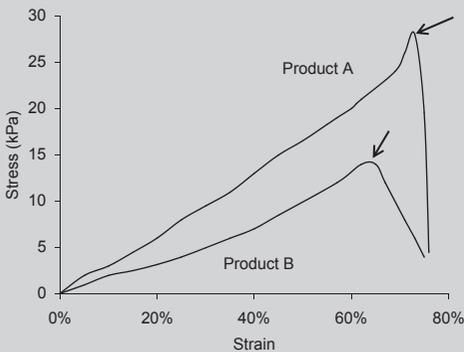
Fig. 9.16 Stress–strain relation for gels of two types of commercial egg white product under compression. Gels were formed by heating for 30 min at 80 °C. *Arrows* show point of fracture. (Adapted from Nagano and Nishinari 2001)
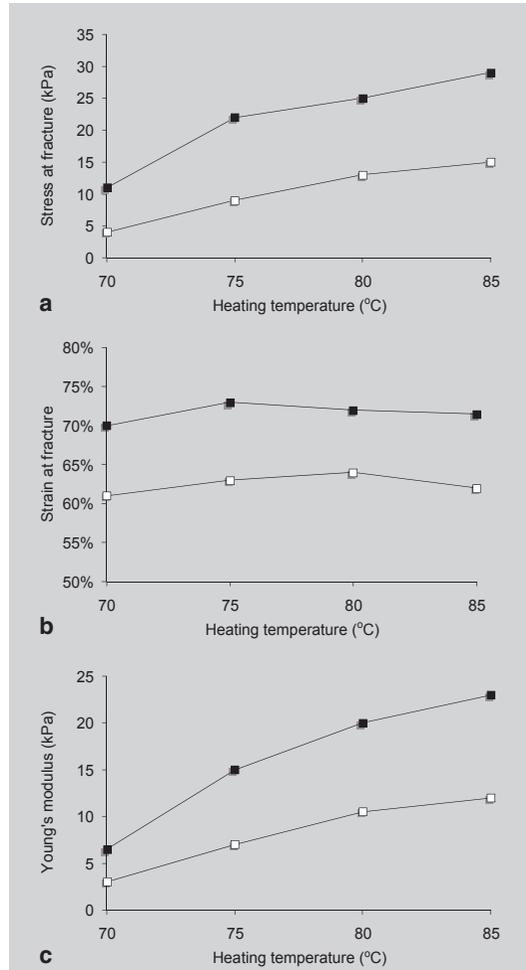
Fig. 9.17 Rheological parameters for gels of two types of commercial egg white product under compression as a function of initial heating temperature (product A, *filled points*; product B *open points*). **a** stress at fracture, **b** strain at fracture, **c** Young's modulus. (Adapted from Nagano and Nishinari 2001)

not the strain at fracture (Fig. 9.17). The different products can be matched on some but not all parameters by adjusting process condition. Whether one product or process would be "better" would depend on which of these parameters had the biggest effect on sensory properties.

something we can actually chew. While we have been able to maintain a causative thread through all stages, the causes have often been qualitative and based on the most idealized molecules and particles. Even as we close, the properties we can explain are only related to our enjoyment of food via correlation. We might be able to understand why one egg white gel has a higher modulus than another but not usefully predict which will taste better or support a firmer soufflé. Still, however cartoonish our explanations, they are explanations rather than just isolated facts and as such provide a framework for thought and hopefully a useful introduction to a profound and fascinating field of further study.

9.7 Bibliography

Chapter 1 and 3 of Mancuso's (1994) "Rheology" textbook give a good description of the properties of solids and viscoelastic materials, while Kavanagh and Ross-Murphy (1998) is very good for rheological methods for gel characterization. Clark and Ross-Murphy (2009) is a thorough review of the formation and properties of biopolymer gels. Another chapter in the same edited book by Van der Linden and Foegeding (2009) is a comprehensive treatment of protein gelation. Chapter 3 in Dickinson's (1992) "Introduction to Food Colloids" also covers the transition between suspension and gels well.

While not considered much in this chapter, the connection between rheological measurements and food texture is an important topic and Bourne's (2002) "Food Texture and Viscosity" is very practically useful.

References

Allen, M. P., & Tildesley, D. J. (1987). *Computer Simulation of Liquids* (p. 385). Oxford: Oxford University Press.

Anema, S. G., & Mckenna, A. B. (1996). Reaction Kinetics of Thermal Denaturation of Whey Proteins in Heated Reconstituted Whole Milk. *Journal of Agricultural and Food Chemistry*, *44*, 422–428.

Arnold, G., Schuldt, S., Schneider, Y., Friedrichs, J., Babick, F., Werner, C., & Rohm, H. (2013). The impact of lecithin on rheology, sedimentation and particle interactions in oil-based dispersions. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, *418*, 147–156. doi:10.1016/j.colsurfa.2012.11.006.

Atkins, P. (2007). *Four laws that drive the universe* (p. 131). Oxford: Oxford University Press.

Atkins, P., & De Paula, J. (2006). *Atkins' Physical Chemistry* (8th ed., p. 1051). New York: W. H. Freeman and Company.

Avery, H. E. (1974). *Basic Reaction Kinetics and Mechanism* (p. 174). London: MacMillan.

Azanha, A., & Faria, J. (2005). Use of Mathematical Models for Estimating the Shelf-life of Cornflakes in Flexible Packaging. *Packaging Technology and Science*, *18*(May), 171–178.

Basaran, T. K., Coupland, J. N., & McClements, D. J. (1999). Monitoring Molecular Diffusion of Sucrose in Xanthan Solutions using Ultrasonic Velocity Measurements. *Journal of Food Science*, *64*(1), 125–128.

BeMiller, J. N., & Huber, K. C. (2008). Carbohydrates. In S. Damodaran, K. L. Parkin, & O. R. Fennema (Eds.), *Fennema's Food Chemistry* (4th ed., pp. 83–154). Boca Raton: CRC Press.

Berg, J. C. (2010). *An Introduction to Interfaces and Colloids: The Bridge to Nanoscience* (p. 785). New Jersey: World Scientific.

Bohlin, L., Hegg, P.-O., & Ljusberg-Wahren, H. (1984). Viscoelastic Properties of Coagulating Milk. *Journal of Dairy Science*, *67*(4), 729–734. doi:10.3168/jds. S0022-0302(84)81362-4.

Bourne, M. (2002). *Food Texture and Viscosity: Concept and Measurement* (2nd ed., p. 416). Oxford: Academic Press.

Chanamai, R., & McClements, D. J. (2000). Impact of Weighting Agents and Sucrose on Gravitational Separation of Beverage Emulsions. *Journal of agricultural and food chemistry*, *48*(11), 5561–5565. doi:10.1021/jf0002903.

Chanamai, R., & McClements, D. J. (2001). Depletion Flocculation of Beverage Emulsions by Gum Arabic and Modified Starch. *Journal of food science*, *66*(3), 457–463.

Chumpitaz, L. D. A., Coutinho, L. F., & Meirelles, A. J. A. (1999). Surface tension of fatty acids and triglycerides. *Journal of the American Oil Chemists' Society*, *76*(3), 379–382. doi:10.1007/s11746-999-0245-6.

Clark, A. H., & Ross-Murphy, S. B. (1987). Structural and Mechanical Properties of Biopolymer Gels. In *Advances in Polymer Science*. Berlin, Heidelberg: Springer-Verlag.

Côté, A. S., Smith, W., & Lindan, P. J. (2014). Democritus. Retrieved from http://www.compsoc.man.ac.uk/~lucky/Democritus/Basic/Contents.html.

Damodaran, S. (2008). Amino Acids, Peptides and Proteins. In S. Damodaran, K. L. Parkin, & O. R. Fennema (Eds.), *Fennema's Food Chemistry* (4th ed., pp. 217–330). Boca Raton.

Dhami, R., Harding, S. E., Jones, T., Hughes, T., Mitchell, J. R., & To, K. (1995). Physico-chemical studies on a commercial food-grade xanthan—I. Characterisation by sedimentation velocity, sedimentation equilibrium and viscometry. *Carbohydrate polymers*, *21*, 93–99.

Dickinson, E. (1992). *An Introduction to Food Colloids* (p. 214). Oxford: Oxford University Press.

Dickinson, E. (2012). Use of nanoparticles and microparticles in the formation and stabilization of food emulsions. *Trends in Food Science & Technology*, *24*(1), 4–12. doi:10.1016/j.tifs.2011.09.006.

Dill, K. A., Bromberg, S., & Stigter, D. (2003). *Molecular driving forces: statistical thermodynamics in chemistry and biology* (1st ed., p. 665). New York: Garland Science.

Esteban, B., Riba, J.-R., Baquero, G., Puig, R., & Rius, A. (2012). Characterization of the surface tension of vegetable oils to be used as fuel in diesel engines. *Fuel*, *102*, 231–238. doi:10.1016/j.fuel.2012.07.042.

Evans, D. F., & Wennerström, H. (1994). *The Colloidal Domain: Where Physics, Chemistry, Biology, and Technology Meet* (1st ed., p. 515). New York: Wiley-VCH.

Feynman, R. P. (1963). Lectures in Physics. Retrieved from http://www.feynmanlectures.info/docroot/I_46. html#Ch46-S5.

Flory, P. J. (1974). Introductory lecture. *Faraday Discussions of the Chemical Society*, *57*, 7–18. doi:10.1039/dc9745700007.

Francis, M. J., Gulati, N., & Pashley, R. M. (2006). The dispersion of natural oils in de-gassed water. *Journal of colloid and interface science*, *299*(2), 673–677. doi:10.1016/j.jcis.2006.02.055.

Funami, T., Kataoka, Y., Omoto, T., Goto, Y., Asai, I., & Nishinari, K. (2005). Food hydrocolloids control the gelatinization and retrogradation behavior of starch. 2a. Functions of guar gums with different molecular weights on the gelatinization behavior of corn starch. *Food Hydrocolloids*, *19*(1), 15–24. doi:10.1016/j.foodhyd.2004.04.008.

Gaonkar, A. G. (1989). Interfacial Tensions of Vegetable Oil/Water Systems: Effect of Oil Purification. *Journal of the American Oil Chemists Society*, *66*(8), 1090–1092.

Gerber, W. (2014). Surface Tension. *gPhysics.net*. Retrieved from http://gphysics.net/index.php/surface-tension.

Ghanbarzadeh, B., Musavi, M., Oromiehie, A. R., Rezayi, K., Razmi Rad, E., & Milani, J. (2007). Effect of plasticizing sugars on water vapor permeability, surface energy and microstructure properties of zein films. *LWT—Food Science and Technology*, *40*(7), 1191–1197. doi:10.1016/j.lwt.2006.07.008.

Ghosh, S., Peterson, D. G., & Coupland, J. N. (2006). Effects of droplet crystallization and melting on the aroma release properties of a model oil-in-water emulsion. *Journal of agricultural and food chemistry*, *54*(5), 1829–1837. doi:10.1021/jf052262w.

Góral, M., Wiśniewska-Gocłowska, B., & Mączyński, A. (2006). Recommended Liquid–Liquid Equilibrium data. Part 4. 1-Alkanol–Water Systems. *Journal of Physical and Chemical Reference Data*, *35*(3), 1391. doi:10.1063/1.2203354.

Grosberg, A. Y., & Khokhlov, A. R. (2010). *Giant Molecules: Here, There and Everywhere* (2nd ed., p. 348). Singapore: World Scientific Publishing Company.

Hagiwara, T., & Hartel, R. W. (1996). Effect of Sweetener, Stabilizer, and Storage Temperature on Ice Recrystallization in Ice Cream. *Journal of Dairy Science*, *79*(5), 735–744. doi:10.3168/jds.S0022-0302(96)76420-2.

Hammes, G. G. (2000). *Thermodynamics and kinetics for the biological sciences* (p. 184). New York: Wiley.

Hartel, R. W. (2001). *Crystallization in Foods* (p. 325). Gaithersburg: Aspen.

Haynes, W. M., Bruno, T. J., & Lide, D. R. (Eds.). (2013). *CRC Handbook of Chemistry and Physics* (93rd Inte.). CRC Press.

Haynie, D. T. (2001). *Biological thermodynamics* (p. 379). Cambridge: Cambridge University Press.

Herrera, M. L., de León Gatti, M., & Hartel, R. W. (1999). A kinetic analysis of crystallization of a milk fat model system. *Food Research International*, *32*(4), 289–298. doi:10.1016/S0963-9969(99)00083-6.

Hiemenz, P. C., & Rajagopalan, R. (1997). *Principles of Colloid and Surface Chemistry* (p. 672). Boca Raton: CRC Press.

Hoffmann, R. (1998). Qualitative thinking in the age of modern computational. *Journal of Molecular Structure (Theochem)*, *424*, 1–6.

Hunter, R. J. (2001). *Foundations of Colloid Science* (2nd ed., p. 816). Oxford: Oxford University Press.

Inaba, H., & Sato, K. (1996). A measurement of interfacial tension between tetradecane and ethylene glycol water solution by means of the pendant drop method. *Fluid Phase Equilibria*, *125*(1–2), 159–168.

Israelachvili, J. N. (1991). *Intermolecular and surface forces* (p. 450). Academic Press.

Israelachvili, J., & Adams, G. (1978). Measurement of forces between two mica surfaces in aqueous electrolyte solutions in the range 0–100 nm. *Journal of the Chemical Society, Faraday Transactions*, *74*, 975–1001.

Jasper, J., & Kring, V. (1955). The isobaric surface tensions and thermodynamic properties of the surfaces of a series of normal-alkanes, C-5 to C-18, 1-alkenes, C-6 to C-16, and of normal-decylcyclopentane, normal-decylcyclohexane and normal-decylbenzene. *Journal of Physical Chemistry*, *2142*(3), 3–5.

Johansson, D., & Bergenståhl, B. (1995). Lecithins in oil-continuous emulsions. Fat crystal wetting and interfacial tension. *Journal of the American Oil Chemists' Society*, *72*(2), 205–211. doi:10.1007/BF02638901.

Jungwirth, P., & Tobias, D. J. (2006). Specific Ion Effects at the Air/Water Interface. *Chemical Reviews*, *106*(4), 1259–1281.

Kavanagh, G., & Ross-Murphy, S. (1998). Rheological Characterization of Polymer Gels. *Progress in Polymer Science*, *23*(97), 533–562.

Kontogiorgos, V., Tosh, S. M., & Wood, P. J. (2009). Phase behaviour of high molecular weight oat β-glucan/whey protein isolate binary mixtures. *Food Hydrocolloids*, *23*(3), 949–956. doi:10.1016/j.foodhyd.2008.07.005.

Koocheki, A., Ghandi, A., Razavi, S. M. A., Mortazavi, S. A., & Vasiljevic, T. (2009). The rheological properties of ketchup as a function of different hydrocolloids and temperature. *International Journal of Food Science & Technology*, *44*(3), 596–602. doi:10.1111/j.1365-2621.2008.01868.x.

Labuza, T. P., & Labuza, P. S. (2004). Influence of temperature and relative humidity on the physical states of cotton candy. *Journal of Food Processing and Preservation*, *28*(4), 274–287.

Le Meste, M., Champion, D., Roudaut, G., Blond, G., & Simato, D. (2002). Glass Transition and Food Technology: A Critical Appraisal. *Journal of Food Science*, *67*(7), 2444–2458.

Levine, H., & Slade, L. (1992). Glass transitions in foods. In H. Schwartberg & R. W. Hartel (Eds.), *Physical Chemistry of Foods* (pp. 83–222). New York: Marcel Dekker, Inc.

Macosko, C. W. (1994). *Rheology: Principles, Measurements, and Applications* (p. 568). New York: Wiley-VCH.

Makki, F., & Durance, T. D. (1996). Thermal inactivation of lysozyme as influenced by pH, sucrose and sodium chloride and inactivation and preservative effect in beer. *Food Research International*, *29*(7), 635–645. doi:10.1016/S0963-9969(96)00074-9.

Malhotra, A., & Coupland, J. N. (2004). The effect of surfactants on the solubility, zeta potential, and viscosity of soy protein isolates. *Food Hydrocolloids*, *18*(1), 101–108. doi:10.1016/S0268-005X(03)00047-X.

McClements, D. J. (2004). *Food Emulsions. Principles, Practices and Techniques* (2nd ed.). Boca Raton: CRC Press.

Miller, D. D. (1998). *Food Chemistry: A Laboratory Manual* (p. 168). Hoboken: John Wiley & Sons Ltd.

Morris, E. R. (2009). Functional Interactions in Gelling Biopolymer Mixtures. In S. Kapsis, I. T. Norton, & J. Ubbink (Eds.), *MODERN BIOPOLYMER SCIENCE* (First Edit., pp. 167–198). Elsevier Inc. doi:10.1016/B978-0-12-374195-0.00005-7.

Mullen, J. W. (2001). *Crystallization* (4th ed., p. 600). Oxford: Butterworth-Heinemann.

Nagano, T., & Nishinari, K. (2001). Rheological studies on commercial egg white using creep and compression measurements. *Food Hydrocolloids*, *15*(4–6), 415–421. doi:10.1016/S0268-005X(01)00053-4.

Naresh, M. D., Subramanian, V., Jaimohan, S. M., Rajaram, A., Arumugam, V., Usha, R., & Mandal, A. B. (2007). Crystal structure analysis of Hen egg white lysozyme grown by capillary method. *Protein Data Bank*. doi:10.2210/pdb2epe/pdb.

Needham Jr., T. E., Patuta, A. N., & Gerraughty, R. J. (1971). Solubility of Amino Acids in Pure Solvent Systems. *Journal of Pharmaceutical Sciences*, *60*(4), 565–568.

Niño, M. R. R., & Patino, J. M. R. (1998). Surface tension of bovine serum albumin and tween 20 at the air–aqueous interface. *Journal of the American Oil Chemists' Society*, *75*(10), 1241–1248. doi:10.1007/s11746-998-0169-6.

NIST. (2014). Thermophysical Properties of Fluid Systems- The NIST Webbook.

Nordmark, T., & Ziegler, G. (2000). Quantitative assessment of phase composition and morphology of two-phase gelatin–pectin gels using fluorescence microscopy. *Food hydrocolloids*, *14*, 579–590.

Painter, P. C., & Coleman, M. (2004). *Painter and Coleman on Polymers: Polymer Science and Engineering* (p. CD–ROM). Lancaster: DEStech Publications.

Peres, A. M., & Macedo, E. a. (1996). Thermodynamic properties of sugars in aqueous solutions: correlation and prediction using a modified UNIQUAC model. *Fluid Phase Equilibria*, *123*(1–2), 71–95. doi:10.1016/S0378-3812(96)90013-8.

Picuell, L., Bergfeldt, K., & Nilsson, S. (1995). Factors determining phase behaviour of multi-component polymer systems. In S. E. Harding, S. E. Hill, & J. R. Mitchell (Eds.), *Biopolymer mixtures* (pp. 13–36). Nottingham: Nottingham University Press.

Raghavan, S. R., & Douglas, J. F. (2012). The conundrum of gel formation by molecular nanofibers, wormlike micelles, and filamentous proteins: gelation without cross-links? *Soft Matter*, *8*(33), 8539. doi:10.1039/c2sm25107h.

Rees, D. A. (1977). *Polysaccharide Shapes* (p. 80). New York: John Wiley & Sons, Inc.

Reid, D. S., & Fennema, O. R. (2008). Water and Ice. In S. Damodaran, K. L. Parkin, & O. R. Fennema (Eds.), *Fennema's Food Chemistry* (4th ed., pp. 17–82). Boca Raton: CRC Press.

Rodríguez Patino, J. M., Sanchez, C. C., & Rodríguez Niño, M. R. (1999). Structural and morphological characteristics of b-casein monolayers at the air–water interface. *Food Hydrocolloids*, *13*, 401–408.

Rousset, P. (2002). Modelling Crstallization Kinetics of Triacylglycerols. In A. G. Marangoni & S. S. Narine (Eds.), *Physical Properties of Lipids* (pp. 1–36). New York: Marcel Dekker.

Russel, W. W., Saville, D. A., & Schowalter, W. R. (1992). *Colloidal Dispersions* (p. 544). Cambridge: Cambridge University Press.

Ryan-Stoneham, T., & Tong, C. H. (2000). Degradation Kinetics of Chlorophyll in Peas as a Function of pH. *Journal of Food Science*, *65*(8), 1296–1302.

Sá, M., Figueiredo, A., & Sereno, A. (1999). Glass transitions and state diagrams for fresh and processed apple. *Thermochimica acta*, *329*, 31–38.

Schneider, E. D., & Sagan, D. (2005). *Into the cool: energy flow, thermodynamics and life* (p. 362). Chicago: University of Chicago Press.

Shaw, D. J. (1992). *Introduction to colloid and surface chemistry* (p. 306). Oxford: Butterworth-Heinemann.

Stoklosa, A. M., Lipasek, R. A., Taylor, L. S., & Mauer, L. J. (2012). Effects of storage conditions, formulation, and particle size on moisture sorption and flowability of powders: a study of deliquescent ingredient blends. *Food Research International*, *49*(2), 783–791. doi:10.1016/j.foodres.2012.09.034.

Talbot, G. (2014). Water Structure and Science. Retrieved from http://www1.lsbu.ac.uk/water/.

Tangsuphoom, N., & Coupland, J. N. (2009). Effect of surface-active stabilizers on the surface properties of coconut milk emulsions. *Food Hydrocolloids*, *23*(7), 1801–1809. doi:10.1016/j.foodhyd.2008.12.002.

Tinoco, I., Sauer, K., Wang, J. C., & Puglisi, J. D. (2002). *Physical chemistry: principles and applications in biological sciences* (4th ed., p. 740). Upper Saddle River: Prentice-Hall Inc.

Van der Linden, E., & Foegeding, E. A. (2009). Gelation: Principles, Models and Applications to Proteins. In S. Kasapis, I. T. Norton, & J. Ubbink (Eds.), *Modern Biopolymer Science* (pp. 29–91). Elsevier Inc. doi:10.1016/B978-0-12-374195-0.00002-1.

Van Vliet, T. (1996). Large deformation and fracture behaviour of gels. *Current Opinion in Colloid & Interface Science*, *1*(6), 740–745. doi:10.1016/S1359-0294(96)80075-6.

Vonnegut, K. (1963). *Cat's Cradle* (p. 224). London: Penguin Classics.

Walstra, P. (2003). *Physical Chemistry of Foods* (p. 832). New York: Marcel Dekker, Inc.

Wille, R. L., & Lutton, E. S. (1966). Polymorphism of cocoa butter. *Journal of the American Oil Chemists' Society*, *43*(8), 491–6.

Wolf, W., Spiess, W. E. L., Jung, G., Weisser, H., Bizot, H., & Duckworth, R. B. (1984). The water-vapour sorption isotherms of microcrystalline cellulose (MCC) and of purified potato starch. Results of a collaborative study. *Journal of Food Engineering*, *3*(1), 51–73. doi:10.1016/0260-8774(84)90007-4.

Young, F. E., & Jones, F. T. (1949). Sucrose Hydrates. The Sucrose–Water Phase Diagram. *Journal of Physical and Colloid Chemistry*, *53*(9), 1334–1350.

Yucel, U., & Coupland, J. N. (2010). Ultrasonic characterization of lactose dissolution. *Journal of Food Engineering*, *98*(1), 28–33. doi:10.1016/j.jfoodeng.2009.12.003.

Zhou, Y., & Hartel, R. W. (2006). Phase behavior of model lipid systems: Solubility of high-melting fats in low-melting fats. *Journal of the American Oil Chemists' Society*, *83*(6), 505–511. doi:10.1007/s11746-006-1233-8.

Index