• CHAPTER ONE •

Basic Elasticity and Viscoelasticity

In the physically stressful environment there are three ways in which a material can respond to external forces. It can add the load directly onto the forces that hold the constituent atoms or molecules together, as occurs in simple crystalline (including polymeric crystalline) and ceramic materials—such materials are typically very rigid; or it can feed the energy into large changes in shape (the main mechanism in noncrystalline polymers) and flow away from the force to deform either semipermanently (as with viscoelastic materials) or permanently (as with plastic materials).

1.1 Hookean Materials and Short-Range Forces

The first class of materials is exemplified among biological materials by bone and shell (chapter 6), by the cellulose of plant cell walls (chapter 3), by the cell walls of diatoms, by the crystalline parts of a silk thread (chapter 2), and by the chitin of arthropod skeletons (chapter 5). All these materials have a well-ordered and tightly bonded structure and so broadly fall into the same class of material as metals and glasses. What happens when such materials are loaded, as when a muscle pulls on a bone, or when a shark crunches its way through its victim's leg?

In a material at equilibrium, in the unloaded state, the distance between adjacent atoms is 0.1 to 0.2 nm. At this interatomic distance the forces of repulsion between two adjacent atoms balance the forces of attraction. When the material is stretched or compressed the atoms are forced out of their equilibrium positions and are either parted or brought together until the forces generated between them, either of attraction or repulsion, respectively, balance the external force (figure 1.1). Note that the line is nearly straight for a fair distance on either side of the origin and that it eventually curves on the compression side (the repulsion forces obey an inverse square law) and on the extension side. With most stiff materials the extension or compression is limited by other factors (see section 1.6) to less than 10% of the bond length, frequently less, so that the relationship between force and distance is essentially linear. When the load is removed, the interatomic forces restore the atoms to their original equilibrium positions.

It is a fairly simple exercise to extend this relationship to a material such as a crystal of hydroxyapatite in a bone. This crystal consists of a large number of atoms held together by bonds. The behavior of the entire crystal in response to the force is the summed responses of the individual bonds. Thus one arrives at the phenomenon described by Hooke as *ut tensio, sic vis*, "as the extension, so the force." In other words,
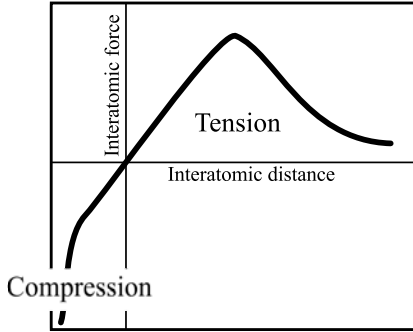
Figure 1.1. Stress–strain curve at the atomic level for a "perfect" material. The origin represents the equilibrium interatomic distance. On either side of the origin the curve is nearly straight.

extension and force are directly and simply proportional to each other, and this rela-tionship is a direct outcome of the behavior of the interatomic bond. However, when one is dealing with a piece of material it is obvious that measurements cannot conve-niently be made of the interatomic distance (though they have been made using X-ray diffraction, which confirms the following). What is actually measured is the increase in length of the whole sample or a part of the sample (making the verifiable assump-tion that in a homogeneous material one part will deform as much as the next). This difference is then expressed as a function of the starting length called the *strain, ε*. Strain can be expressed in a number of ways, each offering certain advantages and insights into the processes of deformation. The most commonly encountered form is conventional, nominal, engineering, or Cauchy strain, which is the increase in length per unit starting length:

$$\varepsilon_C = \frac{\Delta l}{L_0}.$$
[Eq. 1.1]

This estimate of extension works well if the material is extended by no more than a tenth of its starting length. Strain is expressed either (as in this text) as a number (e.g., 0.005) or as a percentage (e.g., 0.5%).

The force acting on each bond is a function of the number of bonds available to share the load. Thus if the area over which the force acts is doubled, then the load carried by each bond will be halved. It is therefore important, if one is to bring the data to the (notionally) irreducible level of the atomic bond, to express the force as a function of the number of bonds that are responding to it. In practice this means expressing the force as force divided by the area across which the force is acting, which is called the *stress, σ*:

$$\sigma = \frac{f}{A_0}.$$
[Eq. 1.2]

However, just as with strain, this simple equation is suitable only for small extensions.

In SI units, the force is expressed in newtons (a function of mass and the accelera-tion due to gravity: one newton is approximately the force due to 100 g, which can be produced by an average apple falling under the influence of gravity), the area in

square meters. One newton acting over an area of one square meter is a pascal (Pa). Other units are in use in many parts of the world. For instance, in the United States the unit of force is the dyne (the force exerted by one gram under the influence of gravity), and the unit of area is the square centimeter. One dyne per square centimeter is one hundred-thousandth (10^{-5}) of a pascal. Traditional engineers in Britain often use pounds and square inches as their measures of "force" and area.

The slope of the straight, or Hookean, part of the curve in figure 1.1 is characteristic of the bond type and is a function of the energy of the bond. For the same reason, the ratio of stress to strain is a characteristic of a material. This ratio is the stiffness or Young's modulus, E:

$$E = \frac{\sigma}{\varepsilon}.$$
[Eq. 1.3]

The units of E are the same as for stress, since strain is a pure number. Graphs showing the relationship between stress and strain are conveniently plotted with the strain axis horizontal and the stress axis vertical, irrespective of whether the relationship was determined by stretching the test piece in a machine and recording the developed forces or by hanging masses onto the test piece and recording the extension. Do not be surprised if it takes a long time for the mental distinctions between stress and strain to become totally clear. Not only are the concepts surprisingly difficult to disentangle, but the confusion is compounded by their uncritical use in everyday speech.

One other characteristic of Hookean materials is that they are elastic. That is to say, they can be deformed (within limits) and will return to their original shape almost immediately after the force is removed (*almost* immediately because the stress wave travels through the material at the speed of sound in that material. Thus when you pull on the brake lever of your bicycle, the brake blocks begin to move a short time later, the time dependent partly on the speed of sound in the steel cable and partly on the length of the cable). This use of the word *elastic* must not be confused with the use of the term as in "elastic band," where "elastic" is taken to mean highly extensible.

Young's modulus is a measure of stiffness in simple extension or compression. There are ways of deforming a material that have different effects on the interatomic forces and therefore different effects on the material. Such a mode of deformation, frequently met, is shear. (Another mode of deformation—volume change, from which is derived the bulk modulus—is ignored here.) As with Young's modulus, the shear modulus is defined as the ratio of stress to strain. The shear stress, τ, is defined as (figure 1.2)
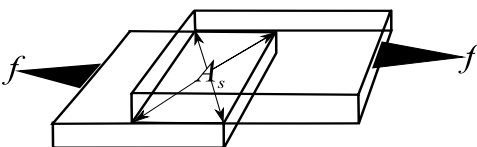
$$\tau = \frac{f}{A_S}.$$
[Eq.1.4]

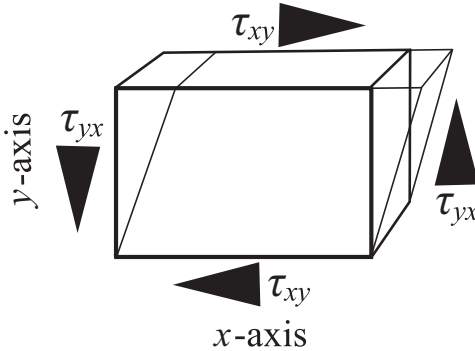Figure 1.2. Conditions for the definition of one-dimensional shear stress (Eq. 1.4).

Figure 1.3. Conditions for the definition of two-dimensional shear stress (Eq. 1.5).

The shear strain is defined somewhat differently (figure 1.3). The strain, y, is measured in radians, and the shear modulus, G, is given by

$$G = \frac{\tau}{y}. \qquad\qquad [Eq.1.5]$$

The simple picture given here is for isotropic materials whose structure and, therefore, mechanical response, is the same in all directions. Young's modulus and the shear modulus in an isotropic material can be related to each other by the expression

$$G = \frac{E}{2(1 + \nu)}, \qquad\qquad [Eq.1.6]$$

where ν is Poisson's ratio. This important ratio is discussed at greater length in section 4.3. A material that is Hookean in extension is usually Hookean in shear. The mathematics for high strain shear deformation is not considered here and, indeed, remains to be established!

1.2 Non-Hookean Materials and High Strains

With greater deformation, another form of strain—true or Hencky strain—is a better indicator of what is going on in the material. With true strain, each small extension is expressed as a fraction of the immediately preceding or instantaneous length. It is slightly more cumbersome to calculate,

$$\varepsilon_H = \ln\left(\frac{L_0 + \Delta l}{L_0}\right) = \ln(1 + \varepsilon_C), \qquad\qquad [Eq.1.1a]$$

and has the curious property that the sample does not "remember" its strain history. True strain is an instantaneous measure of strain. Figure 1.4 compares true and conventional strain, showing that the mutual deviation is far greater in compression.

At larger strains (greater than 0.1 or so), Poisson's ratio effects in an isotropic material (section 4.3) will cause the sample to become narrower, reducing the area over which the force is being transmitted. This will cause the true stress to increase at a
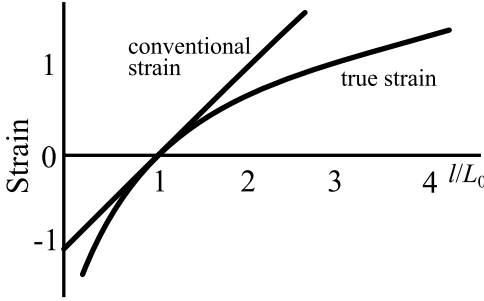
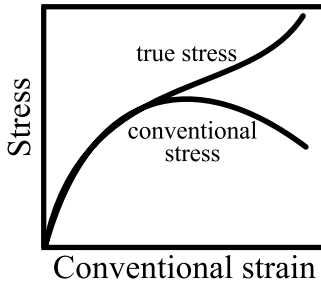Figure 1.4. Comparison of true and conventional ("engineering") strain plotted against extension ratio.

Figure 1.5. Comparison of true and conventional ("engineering") stress plotted against conventional ("engineering") strain.

higher rate than the conventional stress (figure 1.5). However, since, as will be seen, Poisson's ratio frequently varies with strain, especially with soft biological materials that are complex, extensible, and fibrous, it is not possible to give a universal formula for calculating true stress from the starting conditions. The cross-sectional area has to be measured at the particular strain for which the stress is to be calculated. To give you a feel for the relationship between engineering stress and true stress, assume that Poisson's ratio varies in the same way as a rubber, that is to say, the volume of the material remains constant (for many biological materials a doubtful assumption). Thus if the cross-sectional area at any time is A, and A_0 the area at zero strain (L_0), then

$$A(L_0 + \Delta l) = A_0 L_0, \qquad \text{[Eq.1.2a]}$$

so

$$A = \frac{A_0 L_0}{(L_0 + \Delta l)} = \frac{A_0}{(1 + \varepsilon_C)},$$

so that

$$\sigma_H = \frac{f}{A} = \frac{(1 + \varepsilon_C)f}{A_0} = (1 + \varepsilon_C)\sigma_C,$$

which relates the true stress to the apparent or engineering stress.

Both true and conventional methods of expressing stress and strain are used; in the small-strain elastic range, the conventional measures are more usually used and are more convenient, though not strictly accurate. However, at the strains that soft

biological materials reach, true stress and strain are the proper indicators of what
is happening in the material, although these parameters are seldom used. When the
material starts to yield (section 1.6), even true stress and strain are inadequate, since
neither is uniform across the yield zone. Even so, it seems reasonable to use these
measures rather than the conventional parameters, although finite element modeling
is probably the preferred compromise. However, since all biological materials show
some form of relaxation (section 1.4), an estimate of cross-sectional area for the
calculation of true stress has to be made instantaneously. In practice, where such data
are required, it is often found that the best technique is to record the test with a num-
ber of cameras using split-screen video and to make the necessary measurements of
the specimen after the test is completed. This sort of complexity at the practical level
goes a long way to explaining why there are so few data on biological materials in
which true stress has been measured. When it is measured it is often found to be dis-
tributed nonuniformly, so that the assumption of affine (i.e., average or distributed)
deformation is not valid.

1.3 The Energy Approach

It is often easier to consider elasticity not as stress and strain but as their product—
that is to say, energy. When material is deformed (stretched, perhaps), energy (usu-
ally referred to as *strain* energy) is stored in the deformation of its bonds, and it is
this energy that brings the material back to its original shape—or perhaps not, since
that energy can be dissipated in a number of ways, such as heat, sound, surface
energy, plastic deformation, or kinetic energy. With a Hookean material the strains
are relatively small, and all the energy is stored in stretching the interatomic bonds,
termed the *internal* energy. However, if the material is made of relatively long and
unrestrained molecules, the energy can also be stored in changes in their shape and
mobility, termed the *entropic* energy. This is typical of the long-range elasticity ex-
hibited by rubbers, which can stretch up to six times their original length. When a
rubbery material is deformed, its molecules lose mobility, and the energy that has
been powering their random movements is dissipated as heat. If you stretch a rubber
band while you hold it to your lip (very sensitive skin), you will detect an increase in
temperature. Relax the band, and the molecules resume their motions, taking energy
from their surroundings, and you will feel the band go cool. The entropic component
can be characterized by this exchange of heat. In simple terms,

$$A = U - TS,\qquad\qquad\text{[Eq. 1.7]}$$

where A is the Helmholtz free energy, U is the internal energy component, and $-TS$
is the entropic component, made up of temperature, T, and entropy, S. Note that the
entropic component is negative. If you increase the temperature of a material that
relies on internal energy for its elastic behavior, it will expand, but an entropic-based
material will contract. This is the corollary of the experiment with the rubber band.

   To introduce external work (i.e., your stretching the material) we have to introduce
force, f, multiplied by change in length, dl:

$$dA = -P\,dV - S\,dT + f\,dl. \qquad\qquad \text{[Eq. 1.8]}$$

This formulation can now be developed to give the basis for measurement of the mechanical properties of a material at a variety of temperatures, which yields the relative contributions of the internal and entropic components of the elastic restoring force. But we need to see what this measurement means in terms of molecular interactions, since this is the starting point for biology. How does the entropic component work at the molecular level?

A technical rubber is composed of very long chains (molecular weight of about 10^5) of one or more monomer units, with each unit more or less freely jointed into the chain, so that each joint allows a wide range of movement. This motion is called "free rotation" about the bonds of the backbone and is what distinguishes a rubbery polymer from a crystalline one: in a crystalline polymer (or in areas of crystallinity) the units cannot move freely because they are packed so closely, and rubbery behavior is impossible. In fact, it takes more than one monomer unit or residue to make a freely rotating unit or "random link," because the monomer units are of a finite size and shape and so cannot move with absolute freedom without hitting their neighbors ("steric hindrance"). With paraffin chains with a tetrahedral valence angle it takes three C — C links to make up a freely rotating or equivalent random link; with *cis*-polyisoprene units, as in rubber made from the latex of *Hevea brasiliensis*, the number of monomer units per random link is 0.77, since there are four bonds to each isoprene unit (Treloar 1975). Under the influence of Brownian motion the free rotation of the equivalent random links about the backbone of the polymer allows the chain to assume a random conformation. In other words, there is no pattern to the angles that each link makes with its neighbor other than a statistical one. The fact that the molecules are in Brownian motion also leads to the concept of kinetic freedom, which is a way of saying that the chains are free to writhe in any direction. Brownian motion is temperature dependent—as the temperature increases, the movement of the molecules and their subunits becomes more and more frenetic. Conversely, as the temperature decreases, the activity of the molecules slows until, finally, at a temperature dependent on the particular rubber in question, it ceases altogether and any force that is exerted on the rubber meets the resistance of the covalent bonds linking the atoms, probably bending rather than stretching them. A rubber at the temperature of liquid nitrogen is Hookean and is said to be glassy. The temperature at which this phenomenon occurs is called the *glass transition temperature*.

At normal temperatures the rubber chains are writhing in Brownian motion. It is this writhing that produces the tension. Imagine that you hold one of these writhing molecules by the ends and try to pull it straight. You are trying, by doing work on the molecule, to decrease its entropy. If the temperature increases and the molecule writhes more violently, it opposes your efforts with greater force. As we shall see, short stretches of molecules of biological elastomers demonstrate this behavior. The rest of the molecules support and isolate these small sections. Thus biological elastomers are only partly rubbery (implying that lengths of their molecules are capable of random movement; the remainder are organized and stiff). Biological elastomers (resilin, elastin, abductin, gluten, and doubtless others) have about the same stiffness

(1 MPa) as rubber made from the latex of *Hevea brasiliensis*, but their ultimate strain is only about a fifth that of the cured (cross-linked) latex (depending on the degree of cross-linking). Biological elastomers are more complex still, since they are associated with water, which itself seems to contribute to the elastic mechanism.

1.3.1 VISCOELASTICITY: STRESS, STRAIN, AND TIME

Many biological materials contain crystalline components. A few contain rubbers that are sufficiently well cross-linked to be analyzed in terms of rubber elasticity. But by far the greatest number, if not all, biological materials are viscoelastic to a greater or lesser extent. They have a viscous component. Thus although the mechanical properties of crystalline materials and "ideal" rubbers, at constant temperature, can be described in terms of stress and strain, the mathematical description of viscoelastic materials involves the introduction of a new variable—time.

Viscoelasticity and related phenomena are of great importance in the study of biological materials. Just as strain can be measured in more than one way, so the related rate of strain (i.e., the amount of strain per unit time) can be measured in a number of different ways (Ward 2004). Cauchy strain rate is given by $dl/L_0\,dt$; Hencky strain rate by $dl/l\,dt$. In each expression, dl is the infinitesimally small extension achieved during the short time dt, L_0 is the length at zero time, and l is the length just before the present extension.

Viscosity, η, is defined as the ratio of shearing stress to velocity gradient (Newton's law). Its equivalence to the shear modulus can be seen in Eq. 1.9; its definition is

$$\eta = \frac{F/A}{dv/dy},$$ [Eq. 1.9]

which can be compared with the expression for the shear modulus, G:

$$G = \frac{\tau}{y} = \frac{F/A}{dx/dy}.$$ [Eq. 1.5a]

"Newtonian" viscosity is independent of strain or shear rate. This means that if the force applied to a Newtonian fluid is doubled, the shear rate will also be doubled. Non-Newtonian fluids are those that respond with a more or less than doubled shear rate, depending on whether they show shear thinning or shear thickening. Most biological materials show shear thinning, so that doubling the force will more than double the shear rate, thus making deformation of the material relatively easier at higher shear rates. The units of viscosity are kg m^{-1} s^{-1} or Pa s.

At this point it is necessary to point out that viscoelasticity is not plasticity, with which it is often confused. A viscoelastic material will return to its original shape after any deforming force has been removed (i.e., it will show an elastic response) even though it will take time to do so (i.e., it will have a viscous component to the response). A plastic material will not return to its original shape after the load is removed. In metals, plasticity is called *ductility*. It is, if you like, the converse of elasticity in that the energy of deformation is not stored but is entirely dissipated. A material can show a combination of elasticity and plasticity, in which case although
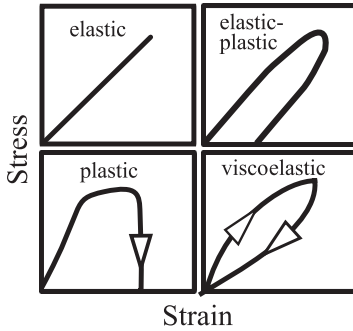
Figure 1.6. Stress–strain curves illustrating different types of behavior.

it partly returns to its original shape on removal of the load, some permanent deformation or "set" remains owing to plastic deformation or molecular "slippage" of an irreversible nature (figure 1.6).

Two major types of experiment are performed on viscoelastic materials: transient and dynamic. Transient experiments involve deforming the material (by simple elongation or in shear) and following the response of the material with time. There are two transient experiments. In one the material is loaded and the change of deformation with time is noted. This is the creep experiment. Under load, segments of the molecules of the material rotate and flow relative to one another at a rate controlled by the viscosity of the material, the stress, the temperature, and the time for which the material has been stressed. Figure 1.7 shows how the strain varies with constant (engineering) stress over a wide range of times after loading. The parameter J, obtained by dividing the strain by the stress, is the *compliance* (roughly the inverse, or opposite, of stiffness) and is here further defined as the creep compliance, $J(t)$. A compliant (or pliant) material is a nonstiff or soggy material. As a practical example, a retro-vinyl buff will choose a pickup cartridge with a "high-compliance" stylus mounting; the stylus presents minimum resistance to being moved by the irregularities that constitute the signal on the groove of the record. The molecular origin of the various regions of the compliance curve is discussed in section 1.5.

The other transient experiment is the stress-relaxation experiment, in which the material is deformed, and the force required to maintain the deformation at a constant value is measured as, with time, the molecules of the material move relative to one
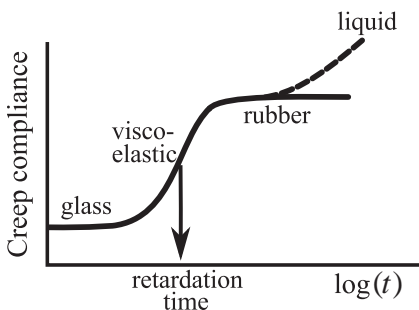

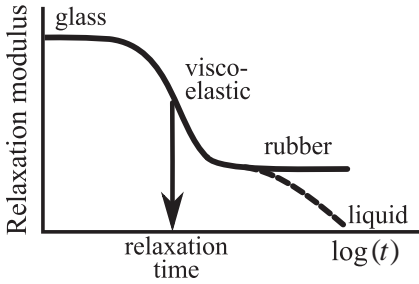Figure 1.7. Creep compliance, $J(t)$, as a function of time, t. The characteristic or retardation time is ρ.

Figure 1.8. Relaxation modulus, $G(t)$, as a function of time, t. The characteristic or relaxation time is ρ.

another. Thus the stress required to hold the material at constant deformation dies away with time and is said to relax. Figure 1.8 shows how the stress varies with constant (engineering) strain (the relaxation modulus, $E(t)$ for simple extension, $G(t)$ for shear) in a manner analogous to that for creep compliance. Note that the two transient experiments are possible because there are three variables—stress, strain, and time. It is therefore possible to plot a three-dimensional surface showing how these variables are interrelated.

The other major type of experiment is the dynamic one, in which either stress or strain (usually strain) is varied cyclically (usually sinusoidally for mathematical convenience) with time, and the response is measured at various different frequencies of deformation.

Transient experiments are usually easier to understand and will be described first. The assumptions made about the mechanical response of the material are similar for both transient and dynamic experiments.

There are three major ways of describing viscoelastic behavior, all interrelated. The first starts with the Boltzmann superposition principle and is sometimes called the *integral* representation of linear viscoelasticity because it defines an integral equation. The second way, which leads to a linear differential equation and is therefore called the *differential* representation, uses assemblages of (Hookean) springs and (Newtonian) viscous elements (dashpots) as models. The third method is based on assumptions about the molecules themselves. At this point you may find it easier to read the section on the behavior of the molecules of viscoelastic materials and then come to the phenomenological approach that follows. Either way you will need to read the following sections several times to see how all the different measurements and ideas fit together.

I cannot emphasize too much that both the integral and differential models are only models and are not explanations. A number of papers on biological materials interpret the behavior of the material solely and finally in terms of springs and dashpots, as if that were an answer. The models are like the hangman's noose—they serve to concentrate the mind but not much more. With the use of mathematical expressions derived from consideration of the models it is possible to derive constants that can be used as a basis for comparison or prediction; but it is highly unlikely that a biological material can be described in terms of a single spring-and-dashpot unit. The other major caveat in the theory of viscoelasticity that follows is that both the models and their mathematical representations rely on linearity of response of both elastic

and viscous components. This is normally considered to be attainable only at strains of less (usually much less) than 0.01, but nearly all biological materials (and most artificial polymers) are not only nonlinear in response but normally function at high and extremely high (0.5+) strains. The models for viscoelasticity are not valid under these conditions. This is a severe limitation and one that is not commonly recognized. Thus much work on artificial and natural polymers is of dubious value, because it applies linear, small-strain models to nonlinear, large-strain materials. That such data may well often be internally consistent is no argument for the acceptance of the linear interpretation; it may merely be coincidence. The mathematics of viscoelasticity at large strains remains to be worked out.

1.3.2 LINEAR VISCOELASTICITY

1.3.2.1 THE INTEGRAL MODEL

The Boltzmann superposition theory may be stated as follows:

1. The creep in a specimen is a function of the entire loading history.
2. Each increment of load makes an independent and additive contribution to the total deformation.

(For creep, substitute stress-relaxation to cover all circumstances.) The first condition could be called the memory function: the response of the material is influenced by what has happened to it so far, so that it is "remembering" deformations long past and allowing them to influence its present behavior. The second condition states that if a specimen is loaded and is creeping under load, then the addition of an extra load will produce exactly the same additional creep as if that total load had been applied to the unloaded specimen and the specimen allowed to creep for the same amount of time. This is said to be a linear (i.e., directly additive) response. The second condition also implies that when the load is removed, the recovery in length of the specimen will follow the same time course as, and be identical with, the initial creep response. The importance of Boltzmann's principle to the study of viscoelasticity is not so much that it provides any explanations as that it provides a starting point for mathematical models that can be tested against reality and refined to give a better fit. For instance, many papers have been written in which the effects of different sequences of stressing or straining have been calculated according to the Boltzmann principle and the results tested by a variety of experiments on real materials. The mathematical formulation of viscoelastic behavior derived from the Boltzmann principle is illustrated by figure 1.9.

The total strain at time t is given by

$$\varepsilon(t) = \Delta\sigma_1 J(t - \tau_1) + \Delta\sigma_2 J(t - \tau_2) + \Delta\sigma_3 J(t - \tau_3), \qquad \text{[Eq. 1.10]}$$

where J is the compliance of the material, and $J(t - \tau_n)$ is the creep compliance function and is the first explicit introduction of time into these equations as an extra variable. This equation can be generalized to give

$$\varepsilon(t) = \int_{-\infty}^{t} J(t - \tau_n) d\sigma(\tau_n), \qquad \text{[Eq. 1.11]}$$
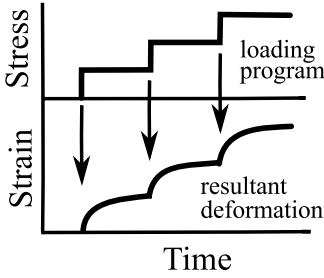
Figure 1.9. Creep behavior of an ideal linear viscoelastic solid.

which is usually rewritten—the immediate elastic response, ε, is removed—allowing mathematicians to rewrite the integral in what they find a more acceptable form:

$$\varepsilon(t) = \left[\frac{\sigma}{G_u}\right] + \int_{-\infty}^{t} J(t - \tau_n)\frac{d\sigma(\tau_n)}{d\tau_n}\, d\tau_n, \qquad \text{[Eq. 1.12]}$$

where G is the immediate or unrelaxed stiffness.

This operation divides the equation into a time-independent and a time-dependent (the integral) function. The stress-relaxation modulus can be calculated in the same manner to give

$$\sigma(t) = [G_r \varepsilon] + \int_{-\infty}^{t} (t - \tau_r)\frac{d\varepsilon(\tau_r)}{d\tau_r}\, d\tau_r. \qquad \text{[Eq. 1.13]}$$

Notice in particular the pattern and symmetry of Eqs. 1.16 and 1.17. This implies that there is probably a formal relationship between the two expressions, but not only is this relationship rather too simple and generalized to be of much use when dealing with biological materials, it is more easily approached from a different starting point!

1.3.2.2 THE DIFFERENTIAL MODEL

Probably the best starting point, and certainly the one most easily appreciated by most biologists, is that of mechanical models using springs (elastic elements) and dashpots (viscous elements)—the differential approach. The springs are Hookean and the dashpots Newtonian. The Maxwell model (figure 1.10) has two elements:

$$\sigma_1 = E_m \varepsilon_1 \text{ for the spring, and} \qquad \text{[Eq. 1.14a]}$$

$$\sigma_2 = \eta_m \, d\varepsilon_2 \, dt \text{ for the dashpot.} \qquad \text{[Eq. 1.14b]}$$

Equation 1.14a can be divided on both sides by dt and rewritten as

$$\frac{d\sigma_1}{dt} \cdot \frac{1}{E_m} = \frac{d\varepsilon_1}{dt},$$

and Eq. 1.14b can be rewritten to give

$$\frac{\sigma_2}{\eta_m} = \frac{d\varepsilon_2}{dt},$$

The two elements are in series, so that $\sigma_1 = \sigma_2 = \sigma$. Also, the total strain on the model, ε, is the sum of ε_1 and ε_2. Thus Eqs. 1.14a and 1.14b can be added to give

$$\frac{d\sigma}{dt} \cdot \frac{1}{E_m} + \frac{\sigma}{\eta_m} = \frac{d\varepsilon_1}{dt} + \frac{d\varepsilon_2}{dt} = \frac{d\varepsilon}{dt}. \qquad \text{[Eq. 1.15]}$$

In a stress-relaxation experiment the length is held constant, so

$$\frac{d\varepsilon}{dt} = 0 \quad \text{and} \quad \frac{d\sigma}{dt} \cdot \frac{1}{E_m} + \frac{\sigma}{\eta_m} = 0. \qquad \text{[Eq. 1.16]}$$

Then, simple rearrangement gives

$$\frac{d\sigma}{\sigma} = -(E_m/\eta_m)dt. \qquad \text{[Eq. 1.17]}$$

At the start of the stress-relaxation experiment, $t = 0$, and $\sigma = \sigma_0$ is the initial stress. Integrating the last equation, we obtain

$$\sigma = \sigma_0 \exp(-E_m/\eta_m). \qquad \text{[Eq. 1.18]}$$

In other words, the stress decays exponentially (i.e., logarithmically) with a characteristic time constant $\tau = \eta_m/E_m$, so that

$$\sigma = \sigma_0 \exp(-t/\tau). \qquad \text{[Eq. 1.19]}$$

The Kelvin or Voigt model (Figure 1.10) models the creep test and, using arguments similar to those with the Maxwell model, gives rise to the expression

$$\varepsilon = \varepsilon_0 \exp(-t/\tau). \qquad \text{[Eq. 1.20]}$$

Note again the extreme symmetry between the two expressions for stress-relaxation and creep.

But why use two models? The Maxwell model is no use for modeling creep, since under constant load the dashpot will allow viscous flow, and the spring will be in constant tension. All that will then be observed is the Newtonian nature of the fluid in the dashpot. This does not accord with observation of real creep experiments, so the Maxwell model is inappropriate for their description. An even more serious objection arises against the use of the Voigt model for stress-relaxation experiments, since under such conditions the model behaves as an elastic solid. We can overcome these objections by combining the two models into a standard linear solid model (figure 1.10). However, one can go on for ever with more and more complex combinations
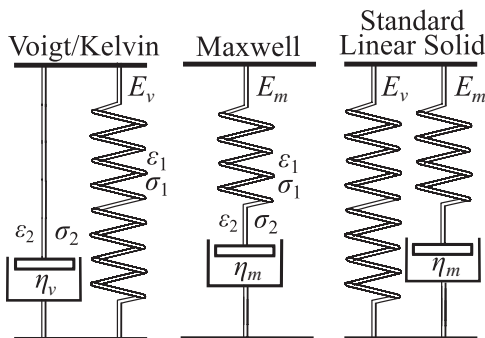

Figure 1.10. Simple spring and dashpot models.

of units that do not produce any more unifying concepts. Further developments along
these lines are ignored here.

The most profitable approach with spring-and-dashpot models, at least in the mod-
eling of artificial polymers, has been found to be that of combining numbers of Max-
well or Voigt elements (not mixing them) to obtain a spectrum of time characteristics.
If the course of relaxation of a single Maxwell element is plotted, it is found to have
the general shape shown in figure 1.11. If the slope of this curve is plotted against log
(or ln) time, a curve of shape similar to a skew log-normal distribution is obtained
(figure 1.12). The vertical axis of figure 1.12 is labeled $-H(\tau)$, which is known as
the *relaxation spectrum function*. The relaxation spectrum is the skew normal curve,
and the relaxation time, τ, of the Maxwell element that generated it is given by the
mode of this curve. If more Maxwell elements, each with a different time constant,
τ, are arranged in parallel, it is not difficult to see that the decay of stress will be
spread over a longer period as a result of a broader spread of relaxation times (figures
1.11 and 1.12). The peculiar usefulness of this relaxation spectrum is that it can be
derived from different types of experiments and so is a convenient transform for gen-
eral comparisons between materials and tests. The other usefulness of the relaxation
spectrum (and, it should be added, the related retardation spectrum calculated in a
similar manner from creep data) is that it gives some idea of the number and nature
of the relaxation processes going on while the stresses are relaxing. This is because
each process has its own characteristic relaxation time (section 1.5). In general, bio-
logical materials have a very broad relaxation/retardation spectrum, but the mesoglea
of two sea anemones, *Calliactis parasitica* and *Metridium senile*, has a retardation
spectrum dominated by a process having a retardation time of $10^{3.4}$ s (figure 4.16).
The model used was a single Voigt element working at strains much greater than 0.01
(3, in fact), so at least in terms of strain, the model was probably inappropriate (Al-
exander 1962). Bill Biggs (unpublished) reworked Alexander's data and found that
they were better fitted by a five-element model with retardation times in the range
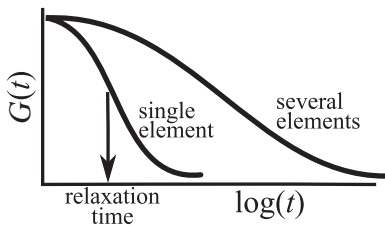

Figure 1.11. The time course of relaxation of a
single Maxwell element and of several elements
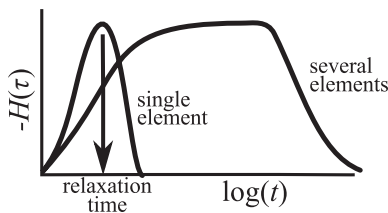with a range of relaxation times.

Figure 1.12. Relaxation spectrum function,
$[-H(\rho)]$ derived from the curves in figure 1.11
using the Alfrey approximation (Eq. 1.26).

of 1 to 10^4 s. But on plotting $-H(\tau)$ against $\ln t$, Biggs found that the compliance is dominated by processes that have a retardation time of $10^{3.5}$ s. Although even here, where definitely more than one process is involved, a single process is dominating the response. This is very unusual with biological materials; therefore, anemone mesoglea could be a good model medium for investigating the mechanisms controlling the retardation spectra of biological materials.

The relaxation spectrum is clearly an important measure of viscoelastic behavior. Its mathematical derivation is as follows: for stress-relaxation at constant strain a single Maxwell element gives $\sigma(t) = E_m \varepsilon \exp(-t/\tau)$. For a number of such elements joined in parallel, all at strain ε, the stress is

$$\sigma(t) = \varepsilon \sum_{n}^{n} E_n \exp(-t/\tau_n), \qquad \text{[Eq. 1.21]}$$

where E_n and τ_n are the stiffness and relaxation time, respectively, of the nth element. Equation 1.21 can be rewritten as

$$\sigma(t) = [G_r \varepsilon] + \varepsilon \int_0^\infty f(\tau) \exp(-t/\tau_n) d\tau. \qquad \text{[Eq. 1.22]}$$

The term $G_r \varepsilon$ is the instantaneous stress. The integral represents the way in which the stress dies away with time, to give $a(t)$. The function $f(\tau) d\tau$ replaces E_n and defines the concentration of Maxwell elements with relaxation times between τ and $(\tau + d\tau)$. The relaxation modulus is then given by

$$G(t) = G_r + \varepsilon \int_0^\infty f(\tau) \exp(-t/\tau) d\tau. \qquad \text{[Eq. 1.23]}$$

The "relaxation time spectrum" $f(\tau)$ is replaced by $H(\tau)$ on a logarithmic time scale (simply because a log time scale is more convenient to handle). Then

$$G(t) = G_r + \int_0^\infty H(\tau) \exp(-t/\tau) d(\ln \tau). \qquad \text{[Eq. 1.24]}$$

In other words, the modulus at time t after the imposition of the strain is the sum of the initial modulus (initial stress divided by the [constant] strain) and of a function that describes how τ varies with time after the start of the experiment. Because τ is the ratio of the (Newtonian) viscosity to the stiffness of the individual elements, the integral can be considered as a function of modulus with time and describes the way in which the modulus changes (diminishes) with time. To calculate $-H(\tau)$ simply, the Alfrey approximation is used. This assumes that $\exp(-t/\tau) = 0$ up to time $t = \tau$, and $\exp(-t/\tau) = 1$ when τ is greater than t, and thus replaces a set of exponentials with a set of step functions. Equation 1.20 can then be rewritten as

$$G(t) = [G_r] + \int_{\ln \tau}^\infty H(\tau) d(\ln \tau), \qquad \text{[Eq. 1.25]}$$

so that

$$H(\tau) = -\left[\frac{dG(t)}{d\ln t}\right]_{t=\tau}, \qquad \text{[Eq. 1.26]}$$

which is the negative slope of a plot of relaxation modulus against $\ln$ (or $\log$) t.

1.3.2.3 THE MOLECULAR MODEL

The third major approach to understanding viscoelasticity is the molecular one. It is probably more convenient in this approach not to use stress-relaxation experiments or creep experiments but, rather, to use dynamic tests . This is not to say that transient experiments are of limited use. Far from it. Their versatility can be increased through variations in the temperature, and this variation will be referred to again once it has been dealt with in conjunction with dynamic tests. The strengths of stress-relaxation and creep tests are their ease of execution—measurement and experimental apparatus are very easily managed—and their immediate applicability to the life of the animal or plant. But rather more can be accomplished with dynamic testing, since it is more versatile and covers a wider range of conditions. The theory is also applicable to a wide range of test rig geometries.

Once again the argument is for linear viscoelastic solids. (The usual subterfuge if you have a nonlinear solid [as are most biological materials] is to say that if you deform the material by a sufficiently small amount, then the material will give a linear response.) Dynamic testing is particularly suitable for tests under such limitations. The sample is subjected to strain varying sinusoidally with time at a frequency w. If the material being tested is Hookean, then the stress will be proportional to the strain—figure 1.13 shows stress and strain plotted against time; figure 1.14 (left) shows stress plotted against strain, which is a straight line. But if the material is viscous and has no elastic component, the stress in the material will be highest at the highest strain rate. Because the strain is varying sinusoidally about zero, the highest strain rate will be at zero strain. Stress in the material will be lowest at the lowest strain rate, which will be the point at which the strain is highest. The resulting stress–strain Lissajous figure will be a circle. Looked at another way, the stress in a viscous material induced by sinusoidal strain is proportional to the changes of accelerations
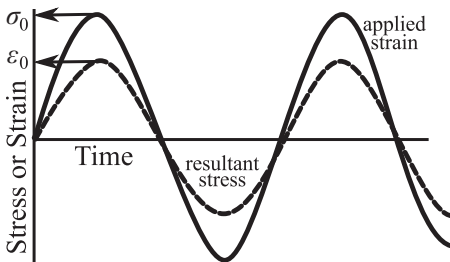

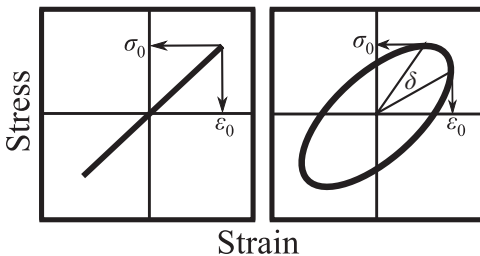Figure 1.13. Sinusoidal strain and resulting stress induced in an elastic material.

Figure 1.14. Simple Lissajous diagrams showing (left) an elastic material (as figure 1.13) and (right) a viscoelastic material (as figure 1.15).
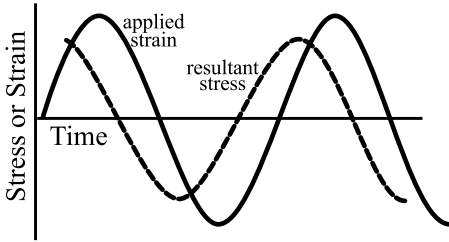
Figure 1.15. Sinusoidal strain and resulting stress induced in a viscoelastic material.

in strain and is therefore the first differential. Thus if $y = \sin x$, then $dy/dx = \cos x$. But $\cos x$ lags $\sin x$ by 90°, so the Lissajous figure is a circle.

A viscoelastic material has a response that is partly viscous and partly elastic, so its response to a sinusoidally varying strain (figure 1.15) will be a combination of the preceding two extremes (figure 1.14, right). The problem is how to extract the information from the Lissajous figures. Essentially it is possible to measure a "modulus" at the highest strain and the highest strain rate—these moduli are the elastic (or real, or storage) modulus and the viscous (or imaginary, or loss) modulus. Because the modulus, G^*, is the ratio of maximum stress to maximum strain, then

$$\sigma_0 = \varepsilon_0 G^* \sin(\omega t + \delta).$$ [Eq. 1.27]

This expression can be expanded to give

$$\sigma_0 = \varepsilon_0 (G^* \cos\delta)\sin\omega t + \varepsilon_0 (G^* \sin\delta)\cos\omega t = \varepsilon_0 G' \sin\omega t + \varepsilon_0 G \cos\omega t,$$ [Eq. 1.28]

where G' is the elastic modulus, and G'' is the viscous modulus. There is thus a simple relationship between G^*, G', G'', and (δ), which is summarized as a vector diagram (figure 1.16). The same argument can be used to extract a complex compliance, J^*, and to resolve it into its components.

The viscosity, η, can, as would be expected, be extracted from the preceding analysis very simply:

$$\eta' = G''/\omega;$$ [Eq. 1.29a]
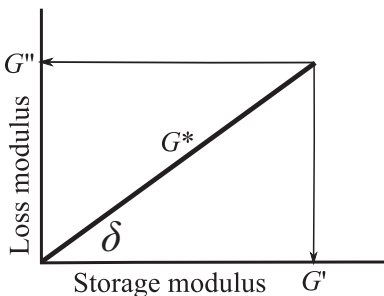
$$\eta'' = G'/\omega.$$ [Eq. 1.29b]

Figure 1.16. Geometric resolution of the complex modulus, G^*, into its component real (G') and imaginary (G'') moduli, and δ.

The in-phase or real component of the viscosity, η', is often called the dynamic viscosity.

Another form of dynamic experiment that has been very useful in the investigation of biological materials will be mentioned in passing. Using either a torsion pendulum or a beam of material mounted at one end only (Ward 2004) it is possible to subject the material to oscillations that decay freely with time. In such free decay the amplitude of the oscillations decreases exponentially with time. Thus if θ_n is the amplitude of the nth oscillation,

$$\frac{\theta_n}{\theta_n + 1} = \exp \Delta, \qquad \text{[Eq. 1.30]}$$

where

$$\Delta = \pi G''/G' = \pi \tan \delta \qquad \text{[Eq. 1.31]}$$

Δ is also known as the logarithmic decrement and is an extremely useful experimental handle for the simple determination of G' and G'', especially at low frequencies up to about 10 Hz in a number of different test geometries.

1.3.3 SPECTRUM OF VISCOELASTIC BEHAVIOR

It is now necessary to draw together the mathematical descriptions of viscoelasticity to show how they are interrelated, how they can complement one another in the investigation of biomaterials, and how their results relate to the structure of the biomaterials at the molecular level. The response of a polymer to dynamic oscillations is probably the easiest to understand from the molecular point of view. The most basic variable with dynamic experiments is time. In this case it's the frequency of oscillation, which is inversely proportional to time after loading in a transient experiment. The polymer molecules are in Brownian motion, just as described for rubber. The backbones are constantly changing their shape, rapidly at short range but with the entire length of the molecular writhing more slowly representing a long-range average of the short-range motions. Any side groups are wagging and twisting. To a first approximation the proportion of molecular displacements that are in phase with the externally applied oscillations represents energy storage; the proportion that are out of phase represents energy dissipated as heat. The material can exhibit the mechanical properties of a glass either if it is cooled (the amount of cooling required depends on the material and is typical of it) or if oscillations are applied at such a high frequency that essentially no backbone motions occur during the period of oscillation. The effect of either of these treatments is to restrict the amount by which the molecule can respond to the external forces by changing its shape, and so the forces are concentrated onto the backbone of the molecule. Under these conditions the molecule behaves much as a Hookean solid.

Rubbery behavior is typical of the plateau zone (figures 1.8 and 1.9). When the polymer molecules are excited at these intermediate frequencies, they become entangled very easily, much as a ball of wool carelessly handled, and the entanglements act as labile cross-links, effectively transmitting the forces.

Between these two zones—the glassy and rubbery plateaus—exists a viscoelastic transition zone. As the frequency of imposed oscillation increases from the rubbery state, the configurational changes in the network strands fail progressively to adjust themselves in the time allowed by the frequency of the oscillations. The long-range motions, being of lower frequency, are the first to run short of time in which to adjust, leaving increasingly shorter range motions to respond as the frequency rises. The strain in response to the applied stress gradually diminishes, and G' increases from the rubbery modulus of 1 MPa to the glassy modulus that is nearly 10 GPa.

What is happening to the loss modulus during these changes? G'' is a measure of the energy lost through "viscous processes." Relatively little energy is lost while the period of oscillation is not similar to the characteristic times that describe the rates of molecular processes involved in mechanical deformation. In the rubbery and glassy zones the oscillation period is different from these molecular resonances, so losses, and G'', are relatively small. But in the transition zone the period of oscillation is similar to that of one or other of the molecular movements; the molecular movements lag the imposed oscillation, dissipating large amounts of energy and giving a high loss modulus, thus contributing a greater viscous component. Obviously, if there are several distinct molecular movements, then there will be distinct discontinuities or secondary transitions, and the curves of G' and G'' will be rather more sinuous.

At the other end of the frequency range—the terminal zone of the modulus curve—entanglement slippage can occur within the period of oscillation, and the molecules can assume any and all possible shapes. There is thus little restraint on the material, and if it is not cross-linked, it will behave as a liquid of high viscosity. The terminal or flow zone will not appear if the material is cross-linked, and the modulus recorded will be the equilibrium modulus of a stress-relaxation experiment: the relaxation modulus, $G(t)$, is approximately a mirror image of G' reflected in the vertical axis. The appearance of the zones is also affected by the molecular weight: if the molecular weight is low (below 10 kDa), the plateau zone is absent, and the transition and terminal zones blend directly. Highly crystalline or glassy polymers will have a relatively high modulus over the entire frequency range, although there are still changes in the modulus that can give much information.

You will remember from the models of transient experiments that the relaxation and retardation times fall in the zone between the rubbery and the glassy states (figures 1.11 and 1.12; Eqs. 1.24–1.29). The characteristic relaxation processes are the same as those occurring in the transition zone of dynamic experiments. In other words, the relaxation times can be associated with the various modes of motion of the molecules. It is this basic association of molecular and mechanical properties that makes t such an important and general constant and that makes the relaxation spectrum, $H(\tau)$, such a useful form of comparison between tests, whether they be transient or dynamic. Using the Alfrey approximation to derive $H(\tau)$ from $G(\tau)$, we can derive $H(\tau)$ from G' or G'' by the following relationship:

$$H(\tau) = \left(\frac{dG'}{d\ln\omega}\right)_{1/\omega=\tau} = \frac{2}{\pi}G''_{1/\omega=\tau}. \qquad \text{[Eq. 1.32]}$$

These transitions can be detected by another parameter, tan δ. Because G'' increases relative to G' in the transition regions, tan δ will also increase in these regions. However, since G' and G'' both increase with frequency, it is not very easy to compare them visually. The ratio between the two gives a much more sensitive comparison, amplifying the differences and making them very obvious. Tan δ is therefore a much-used indicator of the presence, position, and relative magnitude of transitions. As would be expected, τ and tan δ are closely related. For the Maxwell model, tan $\delta = 1/\omega\tau'$; for the Voigt model, tan $\delta = \omega\tau$. These relationships are, however, too simplified and formal to be of practical use in most instances.

So far, temperature has not been mentioned except in reference to the glassy state. A polymer tends to the glassy state either as the temperature is reduced or as the experimental time gets shorter. Thus the high-frequency parts of the dynamic experiments and the first parts of transient experiments (the first fraction of a second—assuming that the loading is instantaneous) produce results equivalent to lowering the temperature in experiments with a longer time constant. Conversely, higher temperature is equivalent to longer times in transient and dynamic experiments, bringing the polymer into a region of lower modulus. For this reason, applying heat to a glassy polymer (e.g., Perspex) softens or melts it, and the plastics-molding industry is made possible. This equivalence of time and temperature has been enshrined in the WLF (Williams, Landel, Ferry, the authors of the paper in which it was derived) equation of time–temperature equivalence or superposition. Its mathematics is beyond the present scope. This relationship allows experiments performed at different temperatures with different time constants to be related to a continuous spectrum of response, which has several implications. The first is that although the different types of transient and dynamic tests are limited in the time ranges over which they are most effective, these ranges can be extended by judiciously varying the temperature. Although the practical range of temperature for biological materials is little more than from 0°C to 40°C, even this amount can extend the time range by four to five orders of magnitude. Thus, although the experimentally convenient time scale for a transient test is about 10^0 to 10^3 s, use of the time–temperature interchangeability allows the range to be extended, from 10^{-2} to 10^5 or so, allowing the analysis of a system with values of t between 10^0 and 10^3 s. Thus the power of the transient experiment in practice becomes greater, which is a Good Thing, because transient tests on the whole require less capital outlay in equipment. However, a note of caution should be sounded. Time–temperature superposition theory has proved its usefulness with artificial polymers but has not yet been adequately tested with the much more complex biological materials. In addition, time–temperature superposition is valid only when no new relaxation processes are made possible by the change in temperature. It is possible, for instance, that a particular relaxation process could occur only above a particular temperature as a result of chemical change due to temperature.

In addition, most biological materials are hydrated. The interactions of these materials with water will change with temperature, so the WLF equation cannot be used directly. The time–temperature relationship highlights another problem that has scarcely been touched on. If a change in temperature, such as might be expected in nature, can change the reaction of a material to mechanical stimuli to the extent

suggested previously, then one might expect mechanical tissues to be adapted to the temperature at which they function. This might be true for collagen, and it is certainly true for elastin.

Water has other effects. It is a plasticizer that swells and softens biological materials. This softening can be attributed to the increase of free space within the material, which allows the polymer molecules greater kinetic freedom. Another way of saying that a material has been softened is to say that G' and G'' have been shifted to higher frequencies or lower temperatures and that the brittle dry materials that become soft and pliable on wetting are brought out of the glassy phase as the glass transition temperature is lowered by the addition of water, which acts as a diluent. The water in proteins has a much greater effect than merely altering the glass transition temperature: being a polar substance, water also greatly influences the conformation of proteins. Hydrophobic amino acids will tend to clump into zones that exclude the water, making the protein globular. Such hydrophobic interactions are largely beyond the scope of this book but are very important in controlling conformation and molecular mobility. In general, the role of water in the mechanics of biological materials has not received very much attention and could do with much more investigation.

1.4 Yield and Fracture

One of the principal characteristics of biological materials and structures is their toughness and resistance to rupture. Skin and wood are as tough as the best man-made materials, although only for wood do we have any idea why this should be so. Toughness is an important requirement for most biological materials—if bladders went pop with the regularity that boilers do, we should all be in deep trouble. There are two main ways in which a material can react when it is extended beyond its safe elastic limit. It can break immediately, or it can undergo plastic deformation, which is known as *plasticity* (in metals, *ductility*). Just as the mechanical properties of a viscoelastic material vary with temperature and strain rate, so do yield and fracture (figure 1.17). Brittle failure is characterized by low strain and rupture that occurs at the highest stress reached. Common brittle materials are biscuit, dessert jello, high-carbon steel, and the membrane around a hen's egg just beneath the shell. Brittle materials are not common in nature. Also shown in figure 1.17 is ductile failure
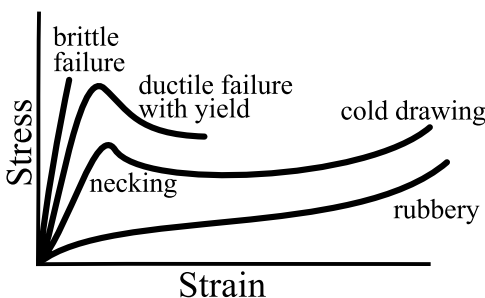

Figure 1.17. Range of yield and fracture behavior of a polymer at different temperatures. As the temperature increases, the initial modulus drops and the material stretches farther.

with yield just before failure. The yield involves plastic deformation. In the curve
illustrating necking and cold drawing, the material has yielded, after which the cross
section is reduced quite abruptly. The material continues to extend, with the polymer
molecules reorientating themselves in the necked region at more or less constant
force (although the stress is really increasing, because the section area of the sample
is decreasing). With a polymer, this process of "cold drawing" produces a mate-
rial, with its molecules now in a preferred orientation, that is much stiffer than the
amorphous material from which it was derived and which is therefore called *strain
hardened*. The strain hardening contributes to the stress–strain curve at a point later
on in the curve where all the material has been cold drawn, causing a final upturn
before brittle failure. A curve of this shape is also produced by hair, and the explana-
tion is somewhat similar. Rubbery behavior is illustrated for comparison. Yield and
plastic deformation are thus associated with molecular transitions; indeed, necking
is itself a transition.

Most biological materials have resistance to necking and yield built into their
mechanics. This property can be demonstrated as follows: the true stress (section
1.2) in a strained specimen is higher than the engineering stress based on the cross-
sectional area of the unstrained specimen. The true stress–strain curve can give the
ultimate tensile stress using the Considère construction (figure 1.18). The tangent to
the stress–strain curve drawn from −1 on the strain axis gives the maximum stiff-
ness of the material. Beyond the point at which the tangent touches the curve, the
true stress is dropping and the material is failing. Figure 1.18 shows this process for
a material with a convex stress–strain curve. Nearly all biological materials have a
concave stress–strain curve to which such a tangent cannot be drawn. Thus there is
no possibility of yield within the working strains of the material, which would not
be so were the stress–strain curve more like that of rubber (figure 1.19). This means
that even if the materials are working near or in their transition zone the energy fed
into the material as it is extended will be spread evenly throughout, and there will be
little possibility of local increases in stress that, as shall be explained, can lead to the
failure of the material at low overall loads.

The fracture behavior of both polymers and the more complex biomaterials is rel-
atively unknown and unexplored, especially since it usually occurs at high strains,
which are mathematically difficult to characterize.The fracture of brittle materials,
the foundations of which were laid by the theory of Griffith (1921), is much better
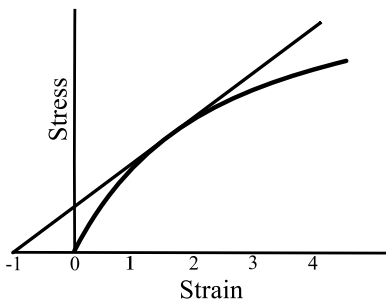

Figure 1.18. The Considère construction as a
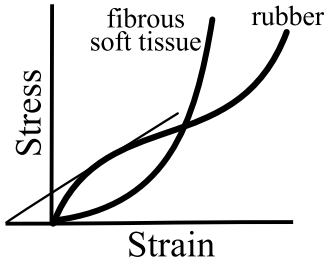criterion of liability to yield.

Figure 1.19. Comparison of the stress–strain curves of
rubber and a typical fibrous soft biological material.

understood. Although the theory is not difficult, it is minimally applicable to polymers
in the viscoelastic and rubbery regions of their behavior and is of limited use with
complex composites such as bone or wood. Essentially, Griffith said that a fracture
results in the formation of two new surfaces on each side of the crack and that the
formation of these surfaces requires energy. This energy is stored as changes in bond
length throughout the rest of the material as it is stretched. The process of fracture
then involves the transmission of this energy to the fracture surfaces, at the same time
relaxing the strain in the area from which energy has been released. If the crack is
considered to be linear, traveling at right angles to the direction of the applied stress
(figure 1.20), then it is reasonable to suppose that the energy that the crack is absorb-
ing comes from an area defined by the crack as the diameter of a circle. It is fairly
obvious that as the crack (length L) extends (by an amount Δl), the amount of energy
available for the propagation of the crack will increase at a greater rate, since it is
proportional to the square of the crack length. Up to a certain crack length the energy
released from this area is not enough to propagate the crack, but after this point (the
critical or *Griffith* length) more energy is released than is required, so the crack is
propagated (figure 1.21). This mechanism, plus the capacity to transmit the stresses
to the crack tip where fracture is occurring, accounts for the fracture of brittle (non-
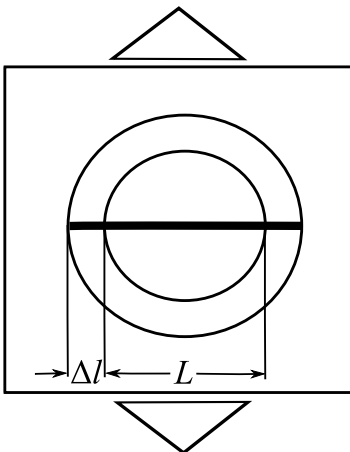ductile) materials. But ductility or plasticity (i.e., irrecoverable deformation) can use

Figure 1.20. Model for deriving the general
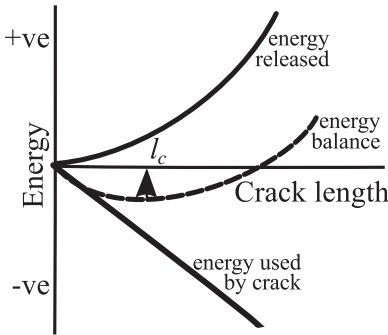conditions for the propagation of a crack.

Figure 1.21. The energy conditions associated with the propagation of cracks showing the derivation of the critical crack length, lc. The term "+ve" indicates strain energy released by the crack; "−ve" indicates energy used to make new surfaces as the crack progresses.

up energy before it can reach the crack tip to contribute to the new fracture surface. In metals, ductility can account for far more energy than is needed for propagation of the crack, and this has also been shown to be true for brittle fracture in polymers. In more compliant polymers and complex materials such as skin, this behavior is expressed in motions such as the alignment of the polymer molecules ahead of the crack tip.

The realization that two criteria have to be fulfilled for a piece of material to break arose largely from Griffith's work. The strength of the material (the critical stress intensity, K_C) must be reached, and there must be sufficient elastic strain energy (work of fracture) available at the tip of the crack to propagate it. In engineering, the critical stress intensity is known as the *toughness*: engineers are, on the whole, more interested in how to resist the initiation of a crack, whereas biological materials sometimes seem to encourage the formation of a crack and then control its propagation through work-of-fracture mechanisms.

In most of the attempts to produce a theory of fracture, one of the aims has been to produce a material parameter that does not depend on the shape of the specimen or the orientation of the crack. This is a reasonable aim that has substantially been achieved for linear materials, in which all the strain energy is involved in fracture and in which fracture occurs at relatively small strains. But one of the properties of polymers at large strains that has already been mentioned is the tendency for the molecules to become oriented in response to the deforming forces. This orientation can, in some materials, lead to strain crystallization somewhat akin to the strain hardening of a cold-drawn polymer. Such crystallization can be reversible and so be a function of strain. Thus for such materials the problem arises that its morphology changes with strain. And just as the properties of the strain-hardened polymer are different from those of the random polymer, so, too, the properties of any polymer may be considered to change with extension. In fact, a cross-linked polymer in a state of large static strain, at equilibrium, may be considered as a new anisotropic material whose linear viscoelastic properties can be studied. It should therefore be no surprise that there is (probably) no unique work of fracture for high-strain polymers.

The other complication is that energy can be dissipated at sites remote from the fracture surface. (Oliver Wendell Holmes's "The Deacon's Masterpiece; or, 'The Wonderful One-Hoss Shay,' a Logical Story" (Holmes 1907) demonstrates this property

to perfection. All the knocks of everyday life were accumulated within the perfectly balanced and adjusted structure and kept away from any fracture surface until, exactly a hundred years after it was made, the shay collapsed into a heap of dust because it could absorb no more strain energy. A few early fiberglass car bodies were rather like this.)

Some indication of this dissipation can be obtained from hysteresis tests, where the hysteresis results from energy lost within the bulk of the material. Should one take account of this loss when calculating the fracture constants? For filled rubbers, in which the amount of filler (carbon black) is varied, it is found that not only is this loss directly correlated with the degree of hysteresis but that higher hysteretic losses are associated with greater toughness. And in the compressive failure of wood across the grain, the size and distribution of large vessels control the distribution of weakness and orchestrate the uniform failure of the material (Hepworth et al. 2002); a similar mechanism seems to be invoked in bone by microcracks (Reilly and Currey 2000). Work with self-healing materials (Trask, Williams, and Bond 2007) shows that a limited amount of damage actually makes material tougher. Additionally, it may well be that mechanisms for energy dissipation also vary with strain and strain rate. A further factor is the transmission of strain energy to the crack tip, where it is needed to supply the energy for crack propagation. The effectiveness of the transmission of stresses to the developing crack tip is a function of shear modulus, which itself changes with shear rate. Thus the speed at which the crack is fed with energy will depend on the rate of change of shape in the material around the crack.

Toughness can be increased by a number of mechanisms, all of which increase the amount of energy required for fracture and all of which can be present in a tough material. The following are some of them:

1. The strain energy is unable to reach the crack tip. For instance, it can be dissipated by plastic yield and failure of the material remote from the crack. It is quite possible that viscous effects within the material will slow down the rate of delivery of energy to the crack tip, so that the crack can be propagated only slowly and with difficulty. Transfer of fluid from one site to another within the material falls into this category and seems to be a mechanism for toughening teeth (Fox 1980) and, very probably, other biomaterials. The strain energy may not be transmitted at all if the shear stiffness of the matrix material is too low (evidenced by a J-shaped stress–strain curve, common in soft tissues) (Mai and Atkins 1989).

2. The total energy required for cracking is raised. For example, the fracture surface is very convoluted and therefore of large area, or the material at the crack tip deforms plastically.

3. The stress at the crack tip is defocused by, for example, increasing its radius of curvature or by the Cook-Gordon effect (see figure 5.29). The sharpness of the crack tip governs the stress intensity. It focuses the strain energy onto the next susceptible bond. At high strains in unidirectional extension, the crack tip rounds off into a semicircle. In rubbers (and probably in other high-strain materials) there is evidence of strain crystallization at the crack tip (now a semicircle) that further strengthens this most vulnerable area.

4. As the crack opens, fibers or filaments extend across it, dissipating energy by their own deformation (Fantner et al. 2005) or by friction as they pull out from the bulk of the material (Pisanova et al. 2001).

5. The material is prestressed in the sense opposite that in which it is most likely to be loaded (e.g., in compression if the most likely loads will be tensile), so that a crack cannot start until this prestress is paid off.

6. The entire structure is so small that the strain energy necessary for fracture cannot be stored.

In practical terms, fracture toughness can be measured and calculated in a large number of different ways. For each test geometry there is a specific mathematical solution that makes a number of assumptions about the material and the test and allows calculation of toughness from a number of more or less simple measurements. General information is available from the work of Atkins (Atkins and Mai 1985). However, biological materials frequently transgress these assumptions, being anisotropic or very stretchy or inhomogeneous or oddly shaped. There is a pragmatic way of coping with these problems: use a test in which the crack grows in a stable fashion such that the test piece can be unloaded (i.e., returned to its original length or shape, or until the recording device shows that no load is still being applied) before it breaks into two pieces. An example is the double cantilever beam (figure 1.22). If a sample of this shape is displaced as indicated, it will, when it reaches breaking load, start to fracture (figure 1.23). If the fracture is allowed to propagate with increasing displacement, and the test piece is unloaded before the specimen breaks in two, the force–extension curve will follow the path *MNO*. Notice that this is not a stress–strain curve: the lines
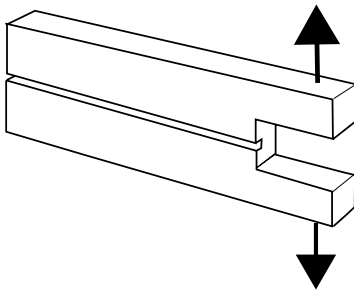

Figure 1.22. Notched double cantilever beam used for the generation of the data in figure 1.23.
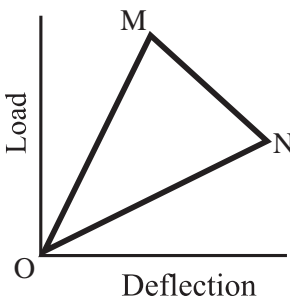

Figure 1.23. Load deflection curve for the estimation of fracture toughness using the Gurney work-area approach.
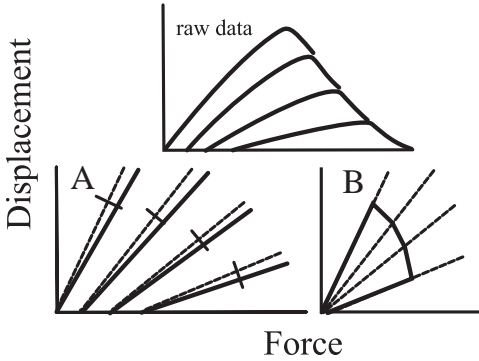
Figure 1.24. Process for deriving work-area curves from experimental data. The slopes and point at which fracture starts (the force drops) are extracted and plotted (A), and a machine stiffness correction is applied (dotted lines). The plastic component of fracture is taken out by moving all the lines to the origin of the graph (B), and the elastic fracture energy is given by the new enclosed areas (Jackson, Vincent, and Turner 1988).

MO and *NO* differ in slope because the cross section of the test piece has been reduced by the cracking. Notice also that because the test piece has been unloaded and returned to zero load, the area within the triangle *OMN* represents the work done, or energy absorbed, in fracturing the specimen. In a viscoelastic or elasto-plastic material there are additional losses due to hysteresis or permanent deformation on loading and unloading (figure 1.24). Corrections for extrinsic factors such as the stiffness of the test machine can be made graphically (Jackson, Vincent, and Turner 1988).

This type of test can be performed with morphologies other than the double cantilever beam: so long as the crack propagates in a stable fashion and can be arrested at the will of the experimenter, it will give results that can be analyzed in this fashion. Atkins and Mai (1985) give the criteria for controlled cracking in a number of morphologies. Alternatively, if the test is performed at a sufficiently slow rate of extension such that the force falls to zero just as the crack finishes traveling across the specimen and it breaks completely, then there can be no strain energy left in the material, and the area under the curve represents only the energy used for fracture. This is a rather risky trick and implies that you can talk to your sample in its own language! Either way, any elastic strain energy is discounted from the final reckoning, and the energy that the force–deformation curve encloses is that required to propagate the crack. The great advantage of this graphical technique is that it is entirely independent of any mathematical model and the assumptions involved in generating such a model. As such, it is particularly useful for biological materials, which are so complex that there frequently isn't a respectable mathematical theory to describe or analyze their fracture processes. This general approach is known eponymously as the *Gurney work-area method* (Gurney and Ngan 1971). (I was once giving a lecture on the fracture mechanics of food, a class of material that comes in so many difficult shapes and sizes [mostly small] that this method is one of the few that works, since it has no prerequisites. At the end of the lecture an elderly gentleman who had been sitting in the front row of the lecture theater came up to me and said, "I'm so pleased you like my ideas"!)

There is another problem. Any piece of tissue will have a number of imperfections (scratches, nicks, notches, and cuts) whose size, nature, and distribution are difficult to control or predict. Depending on the nature of the material, these imperfections

can affect, or even direct, the mechanisms of failure; for instance, they can initiate a crack. This is because any imperfection has the effect of concentrating stress around it, more especially at sharp corners (Gordon 1976). Smooth corners and edges are important in controlling fracture. One strategy is to confine the deformation to a very small area, effectively limiting failure to a small zone, as is achieved in the "trouser-tear" test and by techniques involving cutting or wedging. Another strategy is to introduce an imperfection larger than any of those already in the test piece. This is commonly done by notching or cutting the specimen. This option is not available with brittle materials, in which the starter crack need be only a few micrometers long and thus more or less uncontrollable. In that case a statistical approach is necessary, such as that of Weibull (1951). This approach assumes that the strength of the material is distributed more or less normally about a mean value and provides the mean and the deviation (the Weibull modulus). This has proved useful in the analysis of the fracture properties of potato crisps/chips (Rojo and Vincent 2008, 2009).

The problems involved in fracture of biological materials when they are stretched in two directions at right angles to each other (biaxial straining—the preceding discussion has been concerned with uniaxial straining only) have not been investigated much, even though they are of greater relevance to organisms. It seems likely that fracture of a biaxially strained specimen will be more "brittle" than fracture of a uniaxially strained specimen. Compare the way in which a balloon pops when pricked with a pin with the way a piece of rubber from the same balloon reacts to the same stimulus when stretched by the same amount but uniaxially. The "toughness" of rubber is dependent on the way in which that toughness is measured. Thus it is quite possible that the toughness of skin will vary with its position on the body in accordance with the direction and magnitude of likely strains.

1.5 Adhesion

Adhesion will be mentioned only briefly in that the proper tools for its measurement have been discussed in general terms already. There are a number of different mechanisms of adhesion (Kinlock 1987), which can be summarized as mechanical interlocking (micrometer-sized roughness); diffusion of one component into the other (which has been observed when two artificial rubbers are pressed together); electron transfer (an arcane and mostly insignificant theory and effect); and adsorption (by van der Waals forces and hydrogen or chemical bonds). In general, the strength of an adhesive bond is probably best thought of as a problem in fracture, so it can be measured with the caveats of fracture mechanics in mind. As an example, the adhesion of the gecko's foot relies on its being peeled off the substrate rather like a piece of sticky tape pulled up from one end. Gordon (1976) gives a good discussion of adhesive joints—both their formation and strength or toughness—in his treatment of how wooden aircraft were built and maintained.