2
Macromolecular Structures in Tissues

2.1 Introduction

The human body is a miraculous dynamic structure. When we are young it grows in size and changes shape depending on genetic inheritance and environmental influences such as mechanical loading. Perhaps the most obvious change is the increase in height and weight that is experienced during our adolescent years. These changes are influenced by gravity because the size of our musculoskeleton increases in proportion to the effort we exert against gravity. In addition, the muscle mass we gain reflects the environmental influences such as how much weight we must lift or how fast we must be able to run if we are hunters. Evolutionwise, hunters who couldn't run or lift heavy animals didn't survive to evolve into modern man. If we lived on the moon man would not have needed to have these attributes and we could exist in a more compact evolutionary form. The structural materials in tissues have evolved based on the required mechanical demands on cells and tissues. In humans, the primary structural material is fibrous collagen. This protein is not only stiff, but it can store elastic energy during locomotion and in this manner allow for fast efficient running. In species that do not locomote or fly, other structural materials have evolved including chitosan and other sugar-based polymers.

Materials of construction of the human body include proteins, polysaccharides, lipids, and nucleic acids. These macromolecules not only compose the protein networks that form the scaffold of both mineralized and nonmineralized tissues, but they also make up the structural materials within the cell including the organelles that power and control gene expression and protein synthesis. Proteins form the structural materials of extracellular matrix in the form of collagen and elastin; in addition they make up enzymes and cell surface markers to name a few of their functions. In this text we are most concerned with the role of proteins as structural materials. A few examples of protein-containing molecules of significance include collagens, myosin, actin, tubulins, integrins, and proteoglycans. The basic repeat unit in proteins, also termed polypeptides, is the peptide unit that is

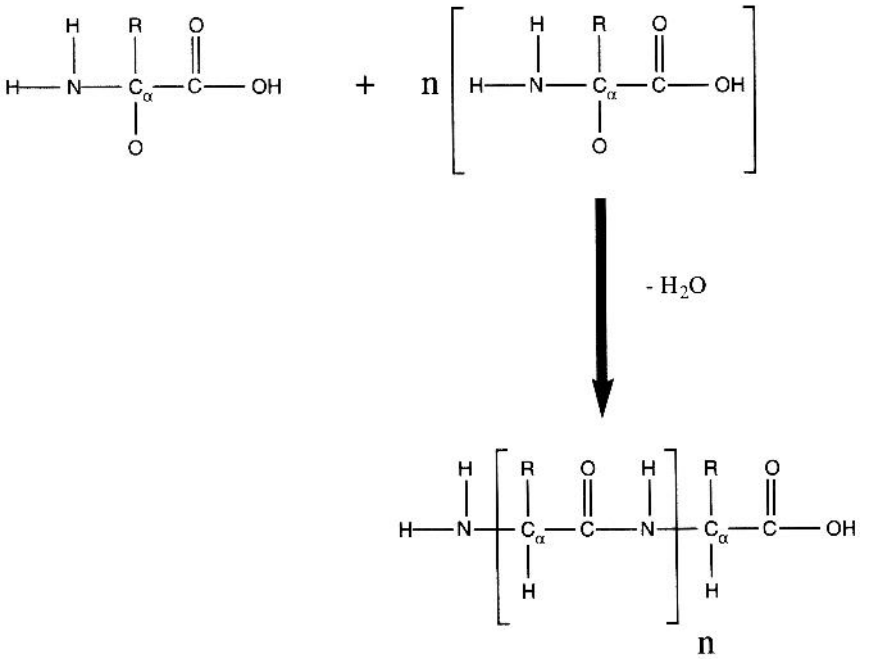


FIGURE 2.1. Formation of a polypeptide. Polypeptides are formed when a condensation reaction occurs between amino acids, releasing water when the amino and acid end groups react.

formed as two amino acids condense to form a dipeptide and water (see Figures 2.1 and 2.2). Proteins are synthesized for transport extracellularly on the endoplasmic reticulum and then transported through the Golgi apparatus for release extracellularly. Proteins are synthesized on ribosomes

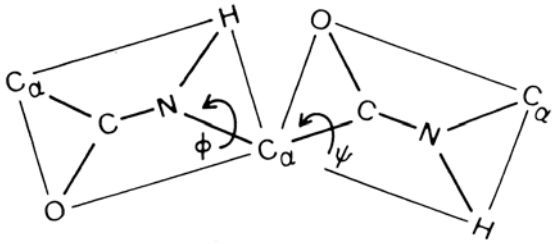


FIGURE 2.2. Diagram of a peptide unit: the peptide unit formed by condensation polymerization is enclosed in a box that is within the plane of the paper. All atoms shown are to a first approximation found within the plane of the paper. The unit begins at the first alpha carbon, C_α, and ends at the second alpha carbon, C_α. Two angles (ϕ, ψ) are sites of free rotation along the backbone of the chain and exist between adjacent peptide units. Both ϕ and ψ are defined as positive for counterclockwise rotation looking from the nitrogen and carbonyl carbon positions towards the alpha carbon between these atoms.

that consist of large and small subunits that translate the genetic code of the cell into a sequence of repeat units.

Polysaccharides are found as sugar polymers that are components of the extracellular matrix (hyaluronan), as carbohydrates (starch), and as energy stores in humans (glycogen). There are numerous repeat units found in polysaccharides but many are derivatives of the simple sugars, glucose and galactose, which are six-membered ring structures (see Figure 2.3). The repeat unit of polysaccharides consists of a six-membered sugar ring linked via an oxygen molecule to the next six-membered ring. Precursors of polysaccharide molecules are synthesized in the cell cytosol and polysaccharides are assembled at the cell membrane or inside the cell depending on whether the polysaccharide is to be released extracellularly or used intracellularly.

Lipids are a heterogeneous group of long hydrocarbon chains that can have one of the following: a free carboxyl group (fatty acids), an ester group

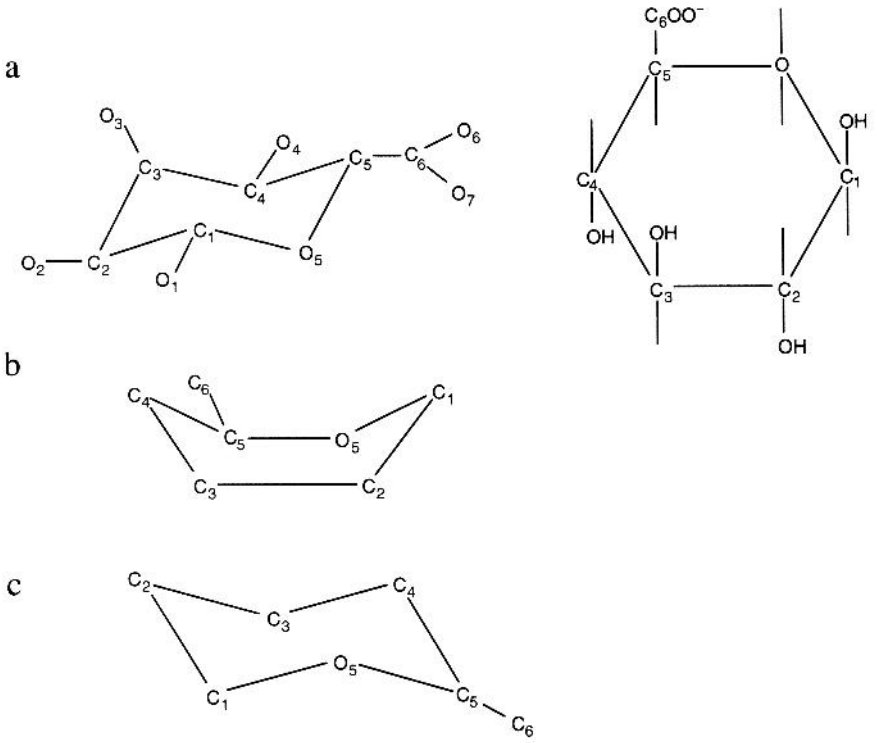


FIGURE 2.3. Boat and chair forms of β-D-glucuronic acid. (**a**) C_1 chair conformation (left) and planar projection (right); (**b**) Boat conformation; and (**c**) ^1C chair conformation of β-D-glucuronic acid. The latter two conformations are energetically less stable than the C_1 chair conformation.

(neutral fats), or derivatives of fatty acids such as membrane phospholipids (see Figure 1.3). Derivatives of phosphatidic acid (phosphatidyl choline, phosphatidyl serine, phosphatidyl inositol) are membrane components that are modified hydrocarbon chains. Rotation around carbon-to-carbon bonds in the backbone gives these molecules great flexibility. The membrane of mammalian cells contains a wide array of polymer structures including lipids, polysaccharides, and proteins. The function of these polymers depends on the size, shape, and flexibility. Polymer molecules of importance in the cell membrane are phospholipids, hyaluronan, cell surface proteoglycans, class I and II histocompatibility markers, and integrins.

The final class of macromolecules to be covered in this book is the nucleic acids. These polymers are composed of a nitrogenous base (either a purine or a pyrimidine), a five-membered sugar ring (a pentose), and phosphoric acid. Deoxyribonucleic acids (DNA) contain the purines, adenine and guanine, the pyrimidines, cytosine and thymine, 2-deoxyribose, and phosphoric acid. DNA is found in the cell nucleus in double-stranded form in the chromosomes. Ribonucleic acid (RNA) contains purines (adenine and guanine), pyrimidines (cytosine and uracil), ribose, and phosphoric acid. RNA is found in the small and large subunits of ribososmes (rRNA), as a copy of the genetic material (messenger (m)RNA) and for adding amino acids to a growing protein chain (transfer (t)RNA). The repeat units and structures of DNA and RNA are shown in Figure 2.4.

2.2 Protein Structure

Proteins make up the bulk of the structural materials found in vertebrate tissues. They comprise the soft and hard ECMs that allow for locomotion, protection from environmental contamination, and for the hair and upper layer of skin that is on the outer surfaces of the body. In addition, they make up the bulk of the structural materials found in the cell cytoskeleton and provide a scaffold on which cells reside in the organs and tissues of the body. For this reason it is important to understand structure–property relationships for proteins.

2.2.1 Stereochemistry of Polypeptides

In the case of proteins the building blocks that are synthesized into a long polymer chain are amino acids. When amino acids are added together they form peptide units that are more easily visualized because they are contained within a single plane (see Figures 2.5 to 2.7) except for the side chain or R group. The sequence of amino acids is important in dictating the manner in which polymer chains behave because the sequence dictates whether a chain can fold into a specific three-dimensional structure or whether the polymer chain does not fold. Specific examples of these prin-

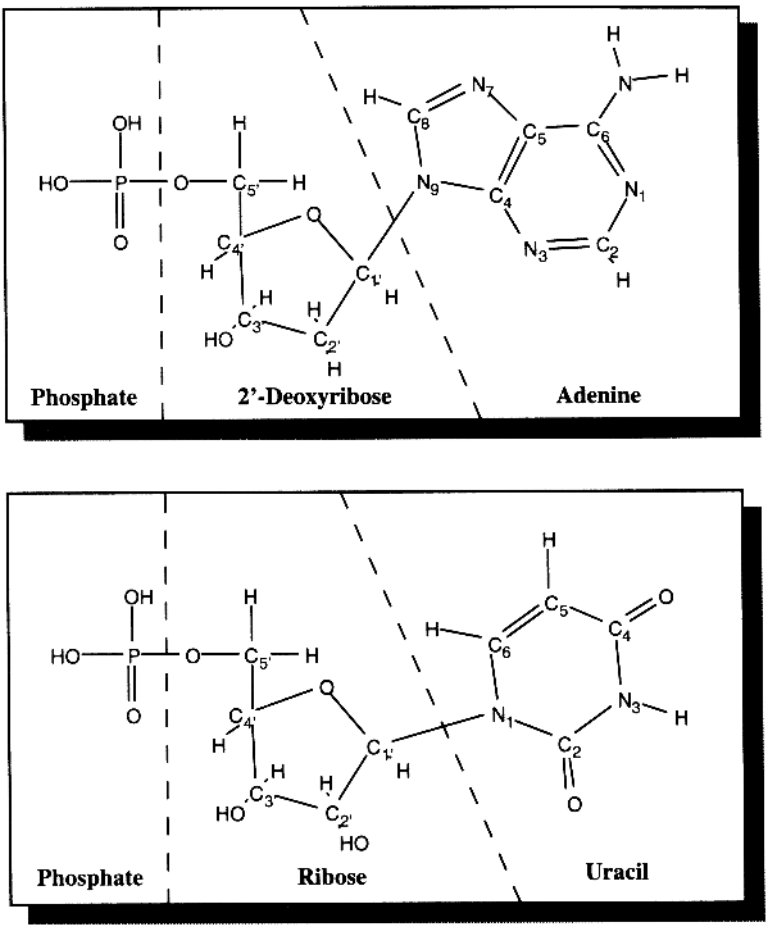


FIGURE 2.4. Structure of nucleic acids. The diagram shows the linkage between phosphate, sugars, and a purine or pyrimidine for DNA (top) and RNA (bottom).

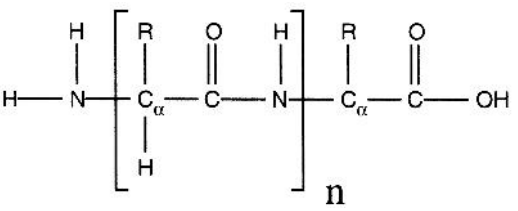


FIGURE 2.5. Chain structure of a polypeptide. The repeat unit of a polypeptide chain consists of amino acid residues, each characterized by a specific amino acid residue designated by R. The size of the polypeptide chain is dictated by *n*, the degree of polymerization.

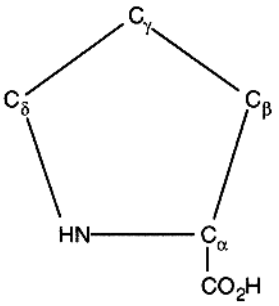


FIGURE 2.6. Chemical structure of proline. The structure of proline differs from that of other amino acids because a ring is part of the polypeptide backbone. The chain backbone includes the amide nitrogen (N), Cα, and the carbonyl carbon (CO_2H).

ciples include keratin that folds into an α helix because of its primary sequence of amino acids and elastin that contains segments with random chain conformations. α helices in keratin pack together to form twisted ropes with very high modulus and tensile strength; these high modulus values are necessary to protect skin from mechanical damage. Elastin forms short segments with random chain structure that result in a material with a much lower modulus and ultimate tensile strength compared to keratin. A relationship between protein structure and mechanical properties exists. Amorphous polymers such as elastin behave as rubberlike materials whereas force-transmitting rigid proteins such as collagen are in extended conformations (i.e., the chain is stretched out in space) and behave as do stiff ropes.

Proteins contain 20 different R groups; actually there are 19 different amino acid side chains or R groups and the 20th, which is proline, is a ring

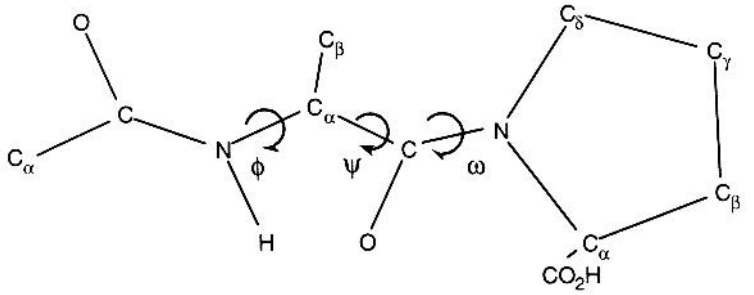


FIGURE 2.7. Polypeptide chain structure containing proline. Chains containing proline lack the flexibility of other peptides because the proline ring has only one available angle for backbone rotation. Rotation occurs around the angles ϕ, ψ, and ω.

that does not have two angles of free rotation. Therefore, the general representation of amino acids in the common side chains found in proteins is shown in Figure 2.5. Proline is the only amino acid found in proteins that is an exception to this rule, inasmuch as the amino acid side chain is part of the backbone as shown in Figures 2.6 and 2.7. The presence of proline and a hydroxylated form of proline, hydroxyproline, are characteristic of collagen and collagenlike polypeptides, which form a triple-helical structure. The primary consequence of the incorporation of proline into the backbone is that it stiffens the three-dimensional structure. Below we explain why this happens.

There are 20 "standard" amino acids that are commonly found in proteins as illustrated by Table 2.1. The sequence of these amino acids dictates the shape of the resulting polypeptide chain, and the presence of large R groups limits mobility of the chain backbone by preventing rotation as diagrammed in Figure 2.7. The structure of a polypeptide and its mobility is very dependent on the location of the alpha carbon (Cα; see Figures 2.8 to 2.10). Also interfering with rotation about the backbone is the possibility

TABLE 2.1. Standard amino acids

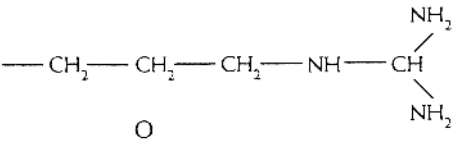
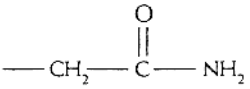
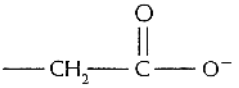
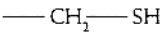
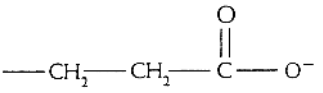
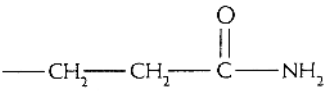
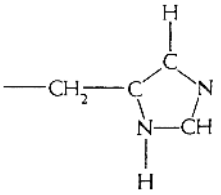
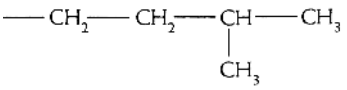
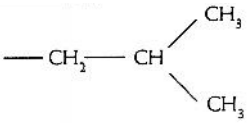
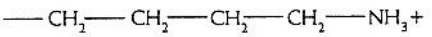
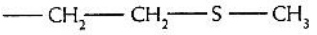
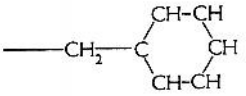
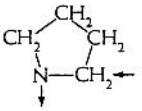
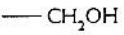
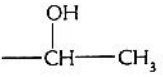
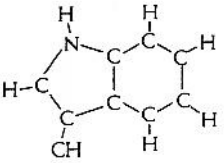
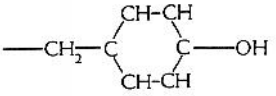
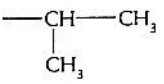
Amino acid	Abbreviation	R group
Alanine	Ala	
Arginine	Arg	
Asparagine	Asn	
Aspartic acid	Asp	
Cysteine	Cys	
Glutamic acid	Glu	
Glutamine	Gln	
Glycine	Gly	

TABLE 2.1. *Continued*

Amino acid	Abbreviation	R group
Histidine	His	
Isoleucine	Ile	
Leucine	Leu	
Lysine	Lys	
Methionine	Met	
Phenylalanine	Phe	
Proline	Pro	
Serine	Ser	
Threonine	Thr	
Trytophan	Trp	
Tyrosine	Tyr	
Valine	Val	

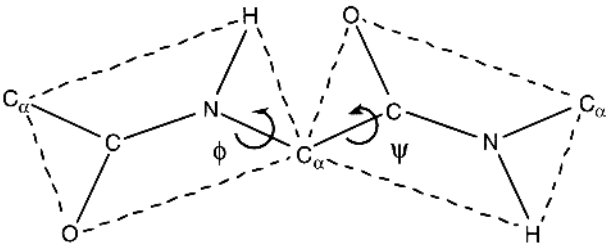


FIGURE 2.8. Diagram of a dipeptide. The diagram shows a dipeptide of glycine in the standard conformation 180°, 180°. To a first approximation, all atoms seen in the dipeptide backbone are found within the plane of the paper. For a dipeptide of glycine, the only atoms not within the plane are the side chain hydrogens. The unit begins at the first Cα and goes to the second Cα. Both ϕ and ψ are defined as positive for clockwise rotation when viewed from the nitrogen and carbonyl carbon positions toward the Cα.

that each amino acid can be in either the D or L form because the carbon to which the R group is attached, the alpha carbon in chemical terms, is asymmetric, that is, has four different chemical groups attached to it. Therefore there are two different chemical forms of amino acids that rotate a plane of polarized light differently (the L form rotates the plane to the left and the D form to the right).

To make a long story short, the structure and properties of the resulting protein synthesized with all D or L amino acids differ. The L form is predominantly found in proteins of higher life species and can be deciphered from the D form by the position of the R group with respect to the backbone. Inasmuch as the alpha carbon (Cα) is attached to three different groups of atoms, the L form cannot be simply rotated around the carbonyl carbon (carbon to which oxygen is attached)–alpha carbon bond. The L form is defined by moving along the peptide chain beginning at the end that contains the free carboxyl acid group (free COOH end of chain) and moving toward the free amino end with the alpha carbon as the point of reference. The amino acid side chain (R group) is on the left as shown in Figure 2.10 (bottom). When the side chain is on the right the D form is

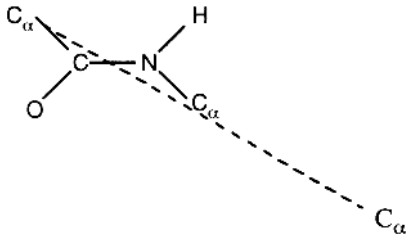


FIGURE 2.9. Location of the second peptide unit. Dipeptides are constructed by knowing the coordinates of atoms in the first peptide unit (using X-ray diffraction) and then translating along the line between the first and second Cα by a distance of 0.37 nm.

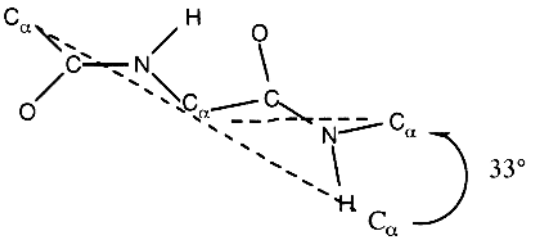


FIGURE 2.10. (Top) Graphical construction of a dipeptide. Once the second peptide unit is located (see Figure 2.12), it is rotated through an angle of 33° clockwise to generate a dipeptide in the standard conformation, with $\phi = 180°$ and $\psi = 180°$.

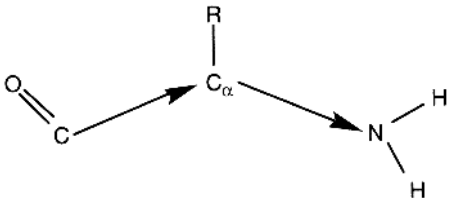


FIGURE 2.10. (Bottom) Difference between L and D amino acids. The two naturally occurring forms of amino acids differ in the position of the R group with respect to the backbone. An L-amino acid has the R group on the left if viewed along the chain from the free carbonyl end to the free amino end as shown. If the R group is on the right, it is defined as the D form. The predominant form found in proteins is the L form, although some amino acids in the D form are present in proteins.

present. This is important because the position of the side chain dictates how a chain can fold in three dimensions.

2.2.2 *Primary and Secondary Structure*

From a structural viewpoint, a polypeptide is composed of planar peptide units as shown in Figure 2.8. The usefulness of considering the peptide unit as opposed to the amino acid is that the peptide unit is almost planar as opposed to the amino acid, which has atoms that are in more than one plane. To illustrate this point, the coordinates of atoms in the peptide unit are given in Table 2.2 and nonbonded atoms cannot be closer than the sum of the minimum atomic distances (Table 2.3). Note that all the atoms from the first alpha carbon (Cα) to the second alpha carbon do not have a *z*-coordinate. These coordinates come from X-ray diffraction studies on proteins and represent the average coordinates found among many pro-

Table 2.2. Coordinates of atoms in peptide unit

Atom	Coordinates (Å) x, y, z
C_α (first amino acid)	0, 0, 0
Carbonyl carbon	1.42, 0.58, 0
Oxygen	1.61, 1.79, 0
Nitrogen	2.37, –0.33, 0
Hydrogen	2.19, –1.31, 0
Cα (second amino acid)	3.70, 0, 0
C_β (first carbon of side chain)	0.51, 0.72, 1.25

teins. The 3-D structure of a dipeptide is shown in Figure 2.8 in the standard conformation.

The standard conformation is obtained by taking the first peptide unit and rotating it about the line between C1 α and C2 α by 180 degrees, translating the rotated peptide unit along the line between C1 α and C2 α by the distance between C1 α and C2 α and then rotation counterclockwise along the extension of the line C1 α to C2 α by 33 degrees as shown in Figures 2.9 and 2.10. If we define the angle of rotation of the bond containing the atoms N–C α as phi (ϕ) and the angle of rotation of bond containing C α and C (carbonyl) as psi (φ), the conformation shown is arbitrarily defined as phi = 180 and psi = 180 degrees.

Using a computer and matrix multiplication techniques the first and second peptide units can be rotated through all possible combinations of values of phi and psi. For each possible set of these dihedral angles, the distances between all pairs of nonbonded atoms can be compared using a set of minimum interatomic contact distances (Table 2.3); if the distance between any set of atoms is smaller than the minimum contact distance then that conformation is not allowed. This boils down to the fact that two atoms cannot be closer than the sum of the van der Waals radii of each atom or electron repulsion occurs. Therefore, by determining the values of phi and

Table 2.3. Minimum interatomic distances for nonbonded atoms

Nonbonded atom pairs	Contact distances normal (minimum), Å
Carbon to carbon	3.20 (3.00)
Carbon to nitrogen	2.90 (2.80)
Nitrogen to nitrogen	2.70 (2.60)
Carbon to oxygen	2.80 (2.70)
Nitrogen to oxygen	2.70 (2.60)
Oxygen to oxygen	2.70 (2.60)
Carbon to hydrogen	2.40 (2.20)
Hydrogen to nitrogen	2.40 (2.20)
Hydrogen to oxygen	2.40 (2.20)
Hydrogen to hydrogen	2.00 (1.90)

psi that result in conformations that are not prevented by electron repulsion, we can determine the number of available conformations (combinations of phi and psi that are allowed).

Conformational maps show values of phi and psi that fall into two regions: the first region contains points that are always allowable and the second region contains points that are sometimes allowed. As shown in Figure 2.11, all points within the inner lines are always allowed and those within the outer solid lines are sometimes allowed. Hence the entropy (or rotational freedom of a peptide) is obtained from the area surrounded by the solid lines on a conformational map (Figure 2.11) and is proportional to the number of allowable conformations of a dipeptide unit. Please note a flexible chain with a lot of rotational freedom is easier to stretch than a rigid chain. We define the flexibility of a polypeptide as the natural logarithm of the number of allowable conformations, #, times Boltzmann's constant (Equation (2.1)).

$$\text{S} = \text{entropy or chain flexibility} = k \ln(\#) \tag{2.1}$$

What Figure 2.11 tells us is that a conformational map for a dipeptide of glycine (the side chain in glycine is very small, just a hydrogen) has mostly allowed or partially allowed conformations and therefore polyglycine is flexible. One question that you might ask is how do we know that the conformational plot for a polypeptide is the same as for a dipeptide? The answer is that because the side chain points away from the backbone for most conformations the atoms in the side chains are separated by more than the sum of the van der Waals radii. However, below we discuss several highly observed conformations of proteins in which the conformational map is an overestimation of the flexibility because of interactions between atoms more than two peptide units apart in space.

A chain of carbon atoms bonded together is ideally flexible in the same manner that polyglycine is flexible. In fact a more practical conclusion of viewing the conformational plot of polyglycine is that poly (ethylene), and lipids are composed of carbon chains bonded together and therefore are ideally flexible macromolecules. Most polypeptides are made up of sequences of amino acids that are not identical and have side chains bigger than glycine. Therefore, protein structure in general is more complicated than that of polyglycine as is discussed next.

Polypeptides, however, are composed of amino acids with side chains that are longer and therefore the area of allowed conformations is reduced when an alanine (Figure 2.12), aspartic acid (Figure 2.13), or a proline (Figure 2.14) is added to the second peptide unit. Finally, the conformational map for a dipeptide of proline–hydroxyproline is dramatically reduced. Rings in the backbone of any polymer reduce the ability of the polymer backbone to adopt numerous conformations and thereby stiffen the structure.

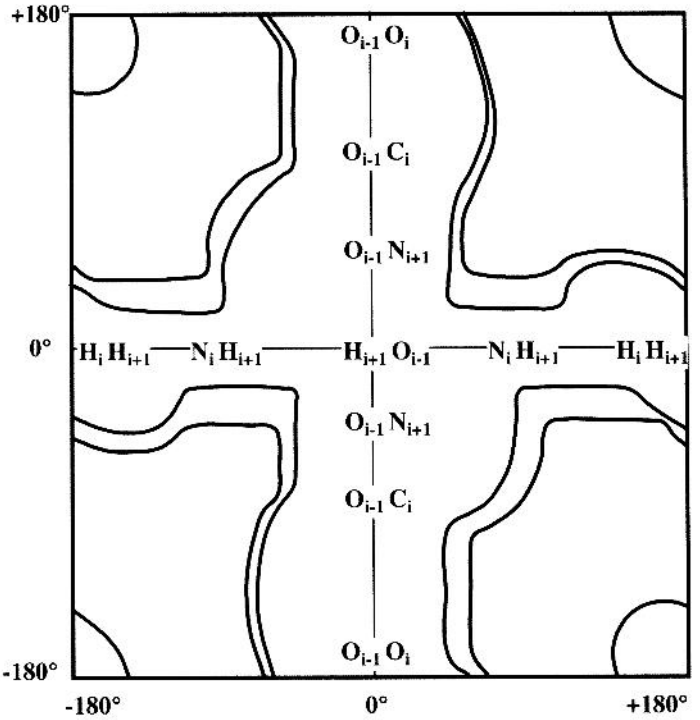


FIGURE 2.11. Allowed conformations for a dipeptide of glycine. Plot of fully (outer solid lines) and partially (inner solid lines) allowed combinations of ψ (vertical axis) and ϕ calculated using a hard sphere model with normal (outer solid lines) and minimum (inner solid lines) atomic contact distances. Dihedral angles (ϕ, ψ) in center of plot (0°, 0°; 0°, 180° and 180°, 180°) are unallowed because of contacts between backbone atoms in the neighboring (*i*th, *i*th + 1 and *i*th − 1) peptide units. The contacts between nitrogen (N), hydrogen (H), oxygen (O), and carbonyl groups (CO_2H) prevent allowed conformations in the center of this diagram. This diagram can be constructed using coordinates of the atoms within the peptide unit and interatomic distances (see Tables 2.2 and 2.3). To construct this diagram, a standard dipeptide unit (see Figure 2.8) is formed by translating the first peptide unit along the line between the α carbons and then flipping the second unit into the *trans* configuration. The values of ϕ and ψ were varied using matrix multiplication, and the interatomic distances were checked for all nonbonded atoms. Pairs of ϕ and ψ that have allowable interatomic distances are found within the inner and outer solid lines shown as the allowable conformations.

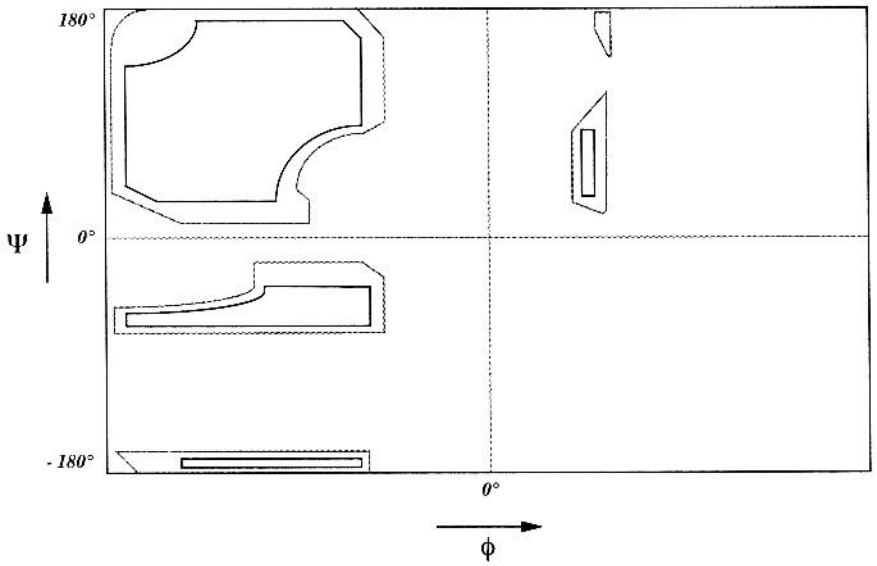


FIGURE 2.12. Conformational plot for glycine–alanine. Plot of allowable angles for peptides containing a repeat unit of glycine and alanine showing totally (outer solid lines) and partially (inner solid lines) allowed conformations determined from normal and minimum interatomic distances.

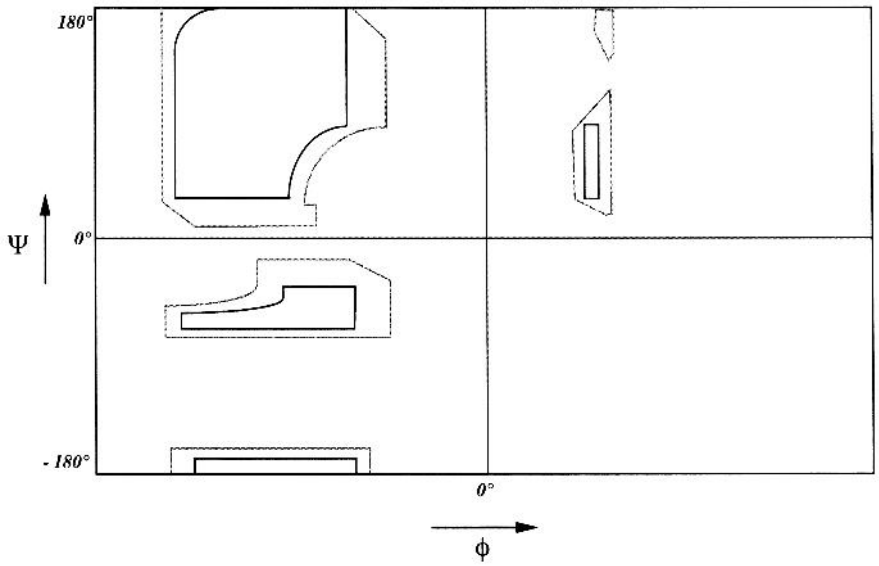


FIGURE 2.13. Conformational plot for glycine–aspartic acid. The allowable conformations are only slightly reduced compared to the plot in Figure 2.12 because the side chain length is increased in going from alanine to aspartic acid.

For polypeptides the addition of a methyl or slightly longer side reduces the number of available conformations to the point where two major regions of allowable conformations stand out (see Figure 2.15). It turns out that when the conformations that are most frequently observed in proteins are tabulated, they fall into these two regions. These conformations include the α helix (α in Figure 2.15; not to be confused with the α chain helix in collagen), β sheet (β in Figure 2.15), and the collagen triple helix (C in Figure 2.15). The regions that these conformations fall into are given by phi between –45 and –120 degrees and psi between 150 and –90 degrees. We learn later that at least the β sheet and the collagen triple helix are highly extended structures and therefore are very rigid and stable; the α helix can be extended into a β sheet and therefore is more deformable.

The mechanical properties of polymer chains that do not exhibit interactions between the side chains and the backbone, or one part of the backbone and another part of the backbone, are related to the number of available conformations and hence the chain entropy. As we discuss later, the stiffness of a polymer chain that does not exhibit bonding with other parts of the chain is related to the change in the number of available conformations. It turns out this refers to random chain polymers of which elastin, poly(ethylene) at high temperatures, and natural rubber are discussed in this text. As we stretch a polymeric chain we reduce the number

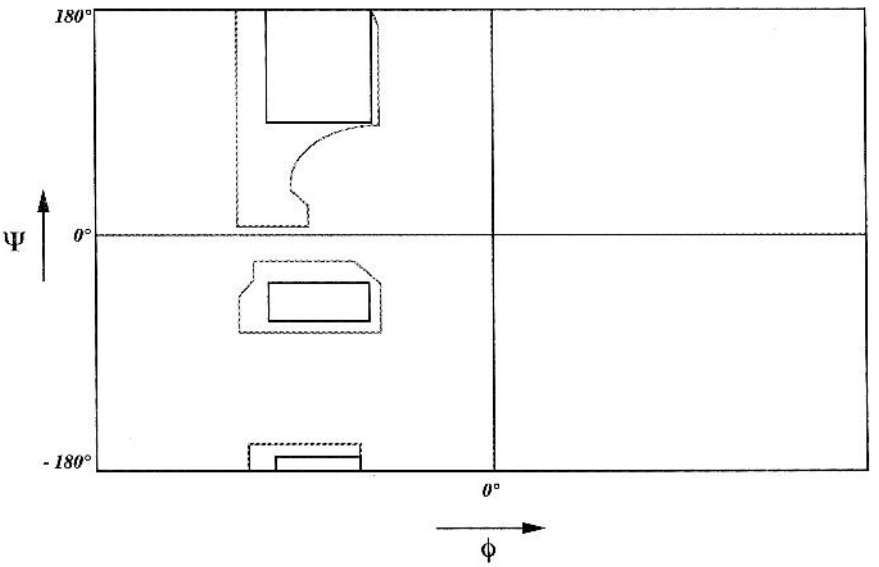


FIGURE 2.14. Conformational plot for glycine-proline. Addition of proline to a dipeptide further reduces the number of allowable conformations when compared to Figures 2.12 and 2.13.

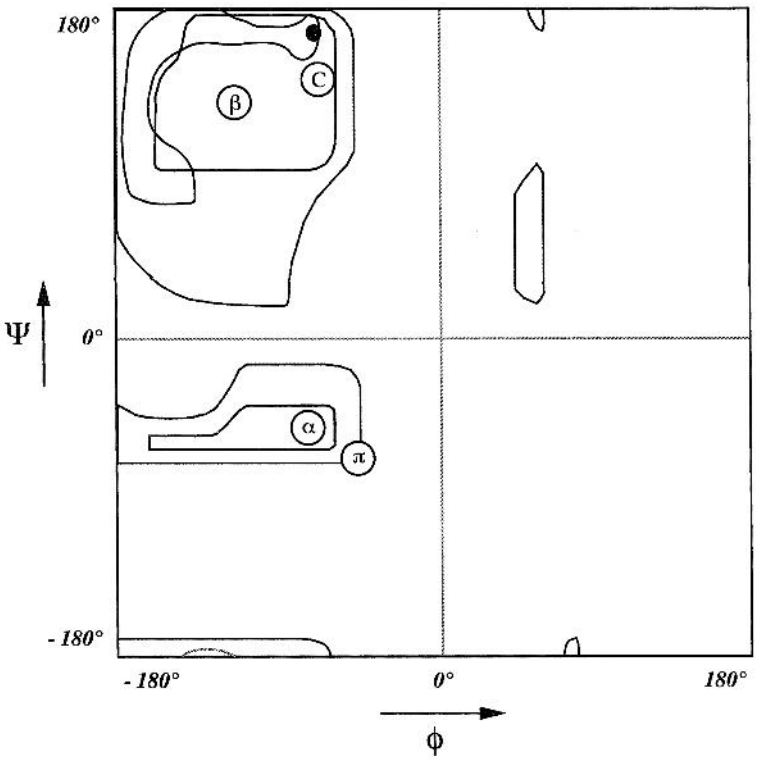


FIGURE 2.15. Conformational plot showing location of α helix, β sheet, and collagen triple helix. The plot shows the localization of the predominant chain structures found in proteins, including the α helix (α), β sheet (β), and collagen triple helix (C). The π stands for a helix that does not occur in nature.

of conformations and the resulting flexibility, thereby increasing the stiffness and resistance to deformation. When a random chain polymer is unloaded, the mobility and flexing of the backbone result in a return of the chain to its original range of rotational motions.

Unfortunately, in general the behavior of proteins is more complex because the α helix, β sheet, and collagen conformations are found repeated in different structures in mammals. As we saw above, a large polymer molecule that has small side chains and no rings in the backbone has the potential to constantly rearrange itself like an eel moving through water. A single conformation is a single set of dihedral angles that characterize the rotational state of each peptide unit that makes up the backbone of the molecule. As a molecule moves by diffusion, it changes its conformation by changing the set of dihedral angles that characterize each dipeptide unit. Folding of a polypeptide chain into the most commonly observed confor-

mations, the α helix, β sheet, and collagen triple helix, occurs because of the inherent flexibility of peptide chains; however, once it folds into a particular structure it is held in place by secondary and other forces.

An interesting perspective is to look at the free energy change associated with transition from a flexible chain into a folded conformation. From thermodynamics we know that the free energy of the transition must be negative for it to occur spontaneously. Let us review a few definitions at this point to refresh your memory. The Gibbs free energy, G, is a measure of the total energy that a system of macromolecules has as well as the entropy (flexibility). Thus for any process to occur spontaneously (within our lifetime) the change in free energy ΔG must be equal to the change in enthalpy ΔH, minus the temperature times the change in entropy $T\,\Delta S$ (see Equation (2.2)). The enthalpy ΔH of a macromolecule is related to the number of covalent Pc, dispersive Pd, and electrostatic bonds Pe, that are formed as well as the translational Et and rotational kinetic energy Er of the chains (see Equation (2.3)).

$$\Delta G = \Delta H - T\Delta S \tag{2.2}$$

$$H = -(Pc + Pd + Pe) + Et + Er \tag{2.3}$$

Therefore the change in enthalpy associated with a transition from a flexible chain to a folded chain involves a change in the number of bonds (the P term) and a change in the flexibility of the chain (S term). We can calculate the change in the S term by a change in the area of allowable conformations (ideally this is k times the natural logarithm of the area under Figure 2.15 divided by one because there is one final conformation). For the process to be spontaneous the change in P must be positive (this means we must form bonds). Formation of a covalent bond lowers the enthalpy by 100 kcal/moles whereas hydrogen and electrostatic bonds lower it by between 1 and 5 kcal/mol. Finally, van der Waals or dispersive forces lower it by 0.01 to 0.2 kcal/mol. Electrostatic and hydrogen bonds occur between atoms with partial charges and can be quantitatively assessed using Coulomb's law, Equation (2.4), where $q1$ and $q2$ are the partial charges on the atoms involved and E is the permutivity of the medium between the charges and rij is the separation distance between charges. Dispersive energy Pd is calculated using the Lennard–Jones 6–12 potential energy function where A and B are two constants specific to the two atoms involved (see Equation (2.5)).

$$Pe = q1q2/(4\pi E\, rij) \tag{2.4}$$

$$Pd = (A/rij^{12}) - (B/rij^{6}) \tag{2.5}$$

In the sections to follow we show that hydrogen bonds are the primary type of bonds that stabilize helical and extended polypeptide conformations.

These structures differ in the different hydrogen bond patterns that occur. Therefore, it is the hydrogen bond pattern that stabilizes folding of polypeptide chains.

2.2.3 Supramolecular Structure

Although protein primary structure (sequence of amino acids) determines how polypeptides fold, proteins that form isolated helices do not possess much mechanical stability. Another way of looking at this is that structural stability of proteins involves transfer of stress between structural units. For this reason the polypeptides that make up our skin, hair, and tendons are assembled into structures with larger diameters and lengths. This implies that tensile-bearing tissues require a higher level of structural hierarchy than do nonload-bearing tissues. Because all structural tissues in vertebrates bear at least tensile loads this suggests that these tissues have levels of structural hierarchy above the primary folding of helices.

Stress transfer is achieved functionally by connecting folded polypeptide chains into continuous networks through electrostatic, hydrogen, hydrophobic, and covalent bonds. An example of the necessity for covalent bonds to stabilize protein structure comes from studying ECMs from animals that have been feed crosslink inhibitors such as beta amino proprionitrile that inhibit collagen crosslinking. The skin and tendons of animals fed this inhibitor tear easily and are rich in collagen molecules that can be removed by immersion in aqueous solvents. For this reason stress transfer and mechanochemical transduction by proteins in the ECM require protein stabilization through chemical crosslinking.

2.2.3.1 Primary and Secondary Structures of Proteins

There are really four classes of protein structures that make up part or all of a biological macromolecule. It should be pointed out that some molecules contain several different structural types connected by sequences that have other structures. The four classes that we discuss include α helix, β sheet, collagen triple helix, and random chain structure. The combinations of backbone angles that are given on the conformational plot shown in Figure 2.15 define these four classifications of macromolecular structures found in proteins. The one exception is the random chain structure that is characteristic of polypeptide sequences such as those found in elastin. In this case the conformational angles are not fixed but allow almost free mobility of polymer chains. However, even in elastin the freely rotating segments are connected via α helical regions and regions with β turns.

2.2.3.2 α Helix

The most commonly occurring helical structure observed in proteins and the first that was worked out is the α helix. The work of Linus Pauling,

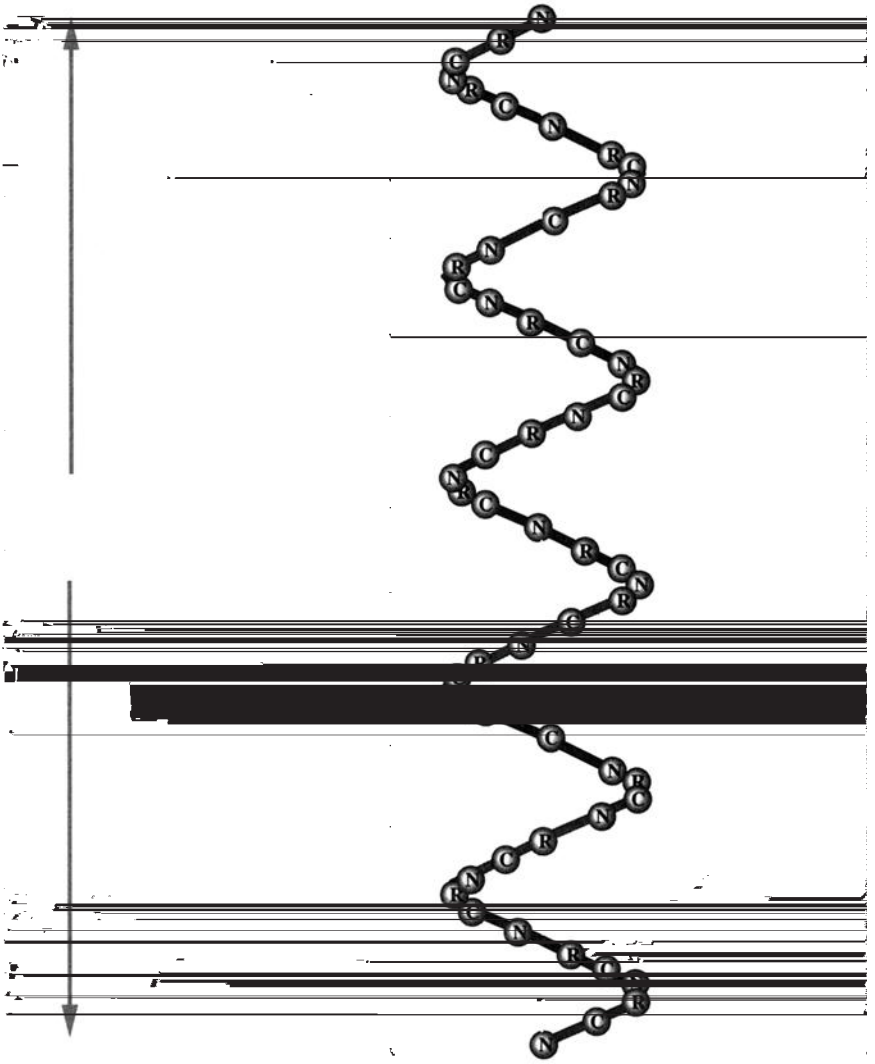


FIGURE 2.16. Structure of a helix. The structure of the α helix is characterized by 3.6 amino acid residues (R groups) per turn of the helix over an axial distance of 0.54 nm. This is consistent with 18 residues per five full turns of the helix.

before he became interested in vitamin C, was instrumental in showing that this helix has 3.6 amino acids per turn, an exact repeat of the helix every 5.4 Å and an axial rise per residue of 1.5 Å as is illustrated by Figure 2.16 (Pauling and Corey, 1951). In the α helix, the amino acid side chains are directed radially away from the axis of the helix. The helix is stabilized by

formation of hydrogen bonds between the carbonyl oxygen of one amino acid residue, which has a slight negative charge and the hydrogen of the amino group of a residue four amino acids farther down the chain, which has a slight positive charge. Of course this requires that the sequence of the amino acids can accommodate a hydrogen bond pattern that is almost perpendicular to the axis of the molecule; protein sequences containing either proline or hydroxyproline cannot form an α helix. α helices are either right- or left-handed (right-handed chains run clockwise as you look from one end to the other whereas left-handed chains are counterclockwise); right-handed forms are most commonly observed. The α helix is very stable because of the numbers of hydrogen bonds formed as well as the linear nature of the bond.

Macromolecules of importance to the biomaterials scientist having some portion composed of α helical structure include hemoglobin, myosin, actin, fibrinogen, and keratin. The α helix is a rather condensed structure because the rise per residue is 1.5 Å and as such is quite different from that of the collagen triple helix and the β structure of silk. The rise per residue in the two latter structures is about twice that found in the α helical structure. For this reason the extensibility of the α helix is greater than that of the collagen triple helix and the β structure and in the case of keratin, tensile deformation of the α helix leads to formation of a β structure.

Although α helices are abundant in proteins, the average length is fairly short, that is, 17 Å long containing about 11 amino acids or three turns. Therefore, α helices are typically found in short domains within proteins and not as continuous stretches. Keratins that make up hair and the most superficial layer of skin contain a central domain with an α helical component of 310 amino acids or about 46.5 nm in length. Keratin is an example of a protein with a fairly long α helix. The amino acid composition that favors α helix formation is fairly broad with the exception of proline and serine.

2.2.3.3 β Sheet

In a similar manner to α helices, extended structures can be held together by hydrogen bonds with the hydrogen bonds running perpendicular to the chain axis. Silk is an example of a protein that is found in the β structure. The amino acid composition of silk is rich in glycine (44.5%), alanine (29.3%), and serine (12.1%) amino acids with small hydrocarbon side chains that form sets of antiparallel hydrogen bonds between molecular chains. Models of polypeptide chains with sequences of poly(gly-ala) and poly(ala-gly-ala-gly-ser-gly) show that the most probable structure contains all the gly residues on one side of the chain and all the ala residues on the other side of the chain, and therefore by packing the chains in

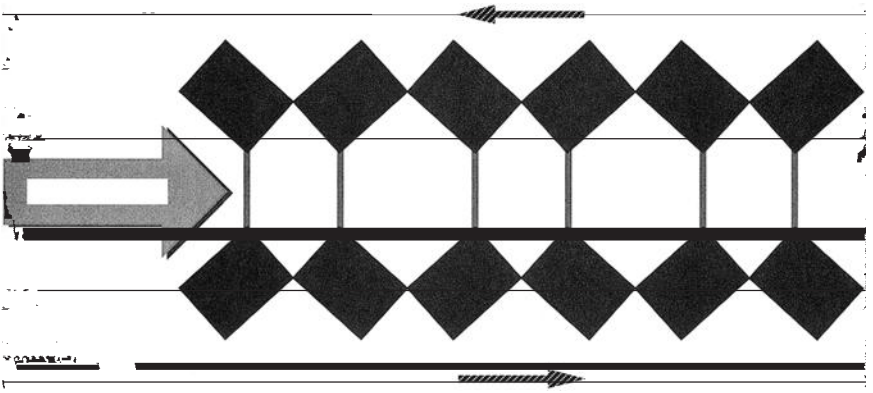


FIGURE 2.17. Hydrogen bonding in antiparallel β sheet. Antiparallel hydrogen bonding between carbonyl and amide groups within the peptide unit stabilizes the β extended conformation.

antiparallel fashion the side chains fit neatly into the empty spaces between the chains (see Figure 2.17). The rise per residue in the β structure is about 3.5 Å.

2.2.3.4 Collagens

The collagens are a family of structural proteins that contain stretches of triple helix that are interrupted by nonhelical regions. Although the fibril-forming collagens found in the ECM are molecules with helical regions about 300 nm long and form fibrils and large fibers seen in ECMs, there are other collagen types that do not form fibrils and have short triple-helical segments. The fibril-forming collagens include types I, II, III, V, and XI; they self-assemble into cross-striated fibrils with the characteristic 67 nm repeat. Fibril-associated collagens, FACIT collagens, are found on the surface of collagen fibrils and appear to connect fibrillar collagens to other components of the ECM. In this text we focus on the fibril-forming collagens that are the structural elements found in vertebrate ECMs and have a characteristic repeat pattern that is observed in the electron microscope (Figures 2.18 and 2.19).

Types I, II, and III collagen form the fibrous network that prevents premature mechanical failure of most tissues and acts to transmit stress to and from cells. The molecular sequences of these collagens are known and they are composed of approximately 1000 amino acids in the form of Gly-X-Y with small nonhelical ends before and after these sequences. All of these collagen types form continuous triple-helical structures that pack laterally

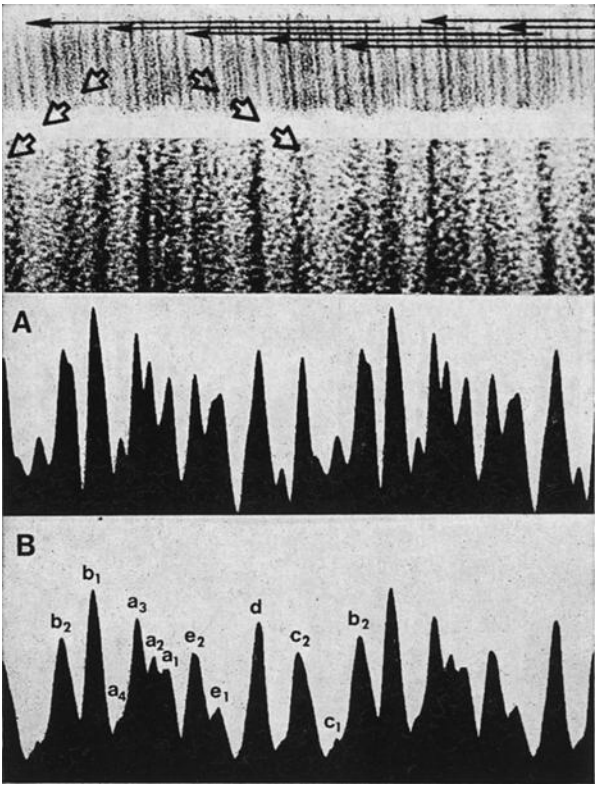


FIGURE 2.18. Positive staining pattern of collagen fibrils: transmission electron micrograph of collagen fibril from rat tail tendon stained with uranyl acetate (top). Uranyl ions electrostatically bind to both negatively and positively charged amino acid residues creating a "positive staining pattern" which represents the location of charged residues along the axis of the triple helix. Arrows indicate that the banding pattern at top is magnified to reveal 12 separate dark lines. Passing a beam of light along the axis of the photographic negative containing this banding pattern results in the peaks shown in (A). The scan shown in (B) was theoretically synthesized using the molecular packing pattern shown in Figure 2.19 and the axial position of charged residues in type I collagen. Band numbers b_2 to c_1 are used to identify the 12 different bands within the repeat period, D, in type I collagen fibrils.

into a quarter-stagger structure in tissues to form characteristic D-periodic fibrils as shown in Figure 2.19. These fibrils range in diameter from about 20 nm in cornea to over 100 nm in tendon. In tendon, collagen fibrils are packed into fibril bundles that are aligned along the tendon axis. In skin, type I and III collagen fibrils form a nonwoven network of collagen fibrils that aligns with the direction of force. In cartilage, type II collagen fibrils

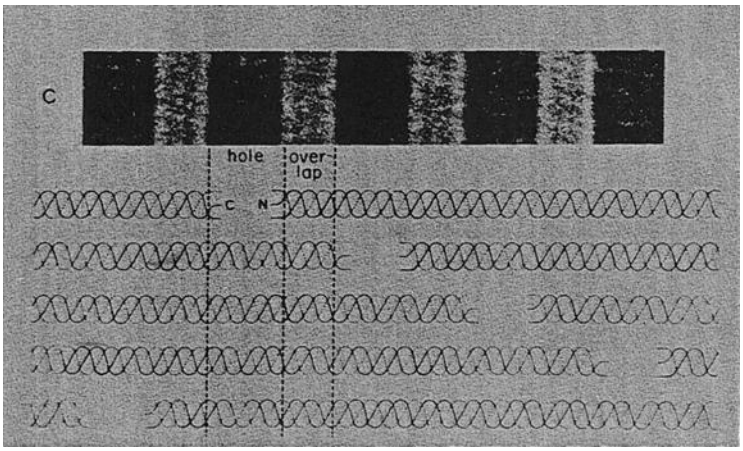


FIGURE 2.19. Packing diagram showing lateral arrangement of collagen molecules in fibrils: collagen molecules observed by electron microscopy in native fibrils are characterized after staining with heavy metals by a repeat pattern consisting of a light region followed by a dark region (top). This repeat pattern occurs in the presence of metal stains such as phosphotungstic acid used for transmission electron microscopy. If short staining times are used, the stain penetrates the holes within collagen fibrils, resulting in a "negative staining pattern." As diagrammed below the staining pattern, this arises because neighboring collagen molecules are staggered laterally by about 22% of the molecular length with a hole of about 13.5% longitudinally between neighboring molecules. This is termed the "quarter stagger" model and is an accurate one-dimensional projection of a type I collagen fibril.

form oriented networks that are parallel to the surface (top layer) whereas in the deeper zones they are more randomly oriented with respect to the surface. The ability of collagen fibers to store and transmit energy is related to the staggered crosslinked structure of the molecules in collagen fibrils.

2.2.3.4.1 Collagen Triple Helix

The other protein structure that we are concerned with is the collagen triple helix. Our knowledge of the structure of collagen comes as a result of early studies on the amino acid composition, the structure of peptides derived from collagen, and the analysis of the X-ray diffraction pattern of collagen fibers. Early compositional studies were important in establishing that collagen was characterized by a high content of glycine, proline, and hydroxyproline. This did not fit in with the established amino acid profile for proteins that form α helices or β sheet structures and therefore a new struc-

ture was proposed for this sequence. However, it was clear that the presence of proline and hydroxyproline would result in an extended structure based on the conformational plot. After cleavage of collagen with acid it was determined that glycine accounted for about 33% of the amino acid residues, and proline and hydroxyproline together accounted for another 25% of the amino acid residues. It was demonstrated by the early 1950s, that the dipeptides that made up collagen were mostly Gly-Pro and Hyp-Gly. This observation led to the hypothesis that every third residue in the collagen structure was probably glycine and the proposed structure must accommodate the rigid proline and hydroxyproline residues.

The other evidence that proved important in unraveling the structure of collagen came from understanding the X-ray diffraction pattern. Prior to 1940 biophysicists recognized that when an X-ray beam passes through a tendon or a tissue containing oriented collagen fibers, spots appear on a photographic plate positioned behind the fiber. Near the meridian (the vertical axis) of the exposed photographic plate, arcs appeared at a position that was equivalent to a spacing of 2.86 Å. These arcs were found in oriented samples of different types of connective tissue, ranging from mammoth tusk to sheep intestine. The repeat of 2.86 Å was thought to be the displacement per amino acid along the axis of the molecule. In 1954, Ramachandran and Kartha (1955) proposed a model consisting of three parallel chains linked to form a cylindrical rod, with the rods being packed into a hexagonal array. A year later Ramachandran and Kartha (1955) modified this model by adding an additional right-hand twist of 36° every three residues in a single chain (see Figure 2.20).

Further refinement of the structure continued for the next 40 years and has led to the understanding that the hydroxyproline has been shown to stabilize the molecule by forming a hydrogen bond between the polypeptide chains via a water molecule as well as a direct hydrogen between the carbonyl group on one chain and amide hydrogen on another chain within groups of three amino acids (Figure 2.21). In addition, it is believed that the molecule can be broken up into rigid and partially flexible regions associated with the gap (four molecules in cross-section) and overlap regions (regions with five molecules in cross-section) within the fibril structure and that crosslinks hold the fibril together (see Figure 2.22). The collagen molecule is found inside the cell in a precursor form termed procollagen (see Figure 2.23). Once the ends of the procollagen molecule are removed within the collagen fibril it has a length of 300 nm and width of about 1.5 nm.

Recently it has been demonstrated, by analysis of the flexibility from conformation maps of dipeptide sequences, that the collagen triple helix contains rigid domains separated by domains with increased flexibility (see Figure 2.23) (see Silver et 2003, for a review). Autocorrelation of the peptide sequences in collagen, using Fourier analysis, demonstrates that there is a period of the flexible domains in both the molecule and

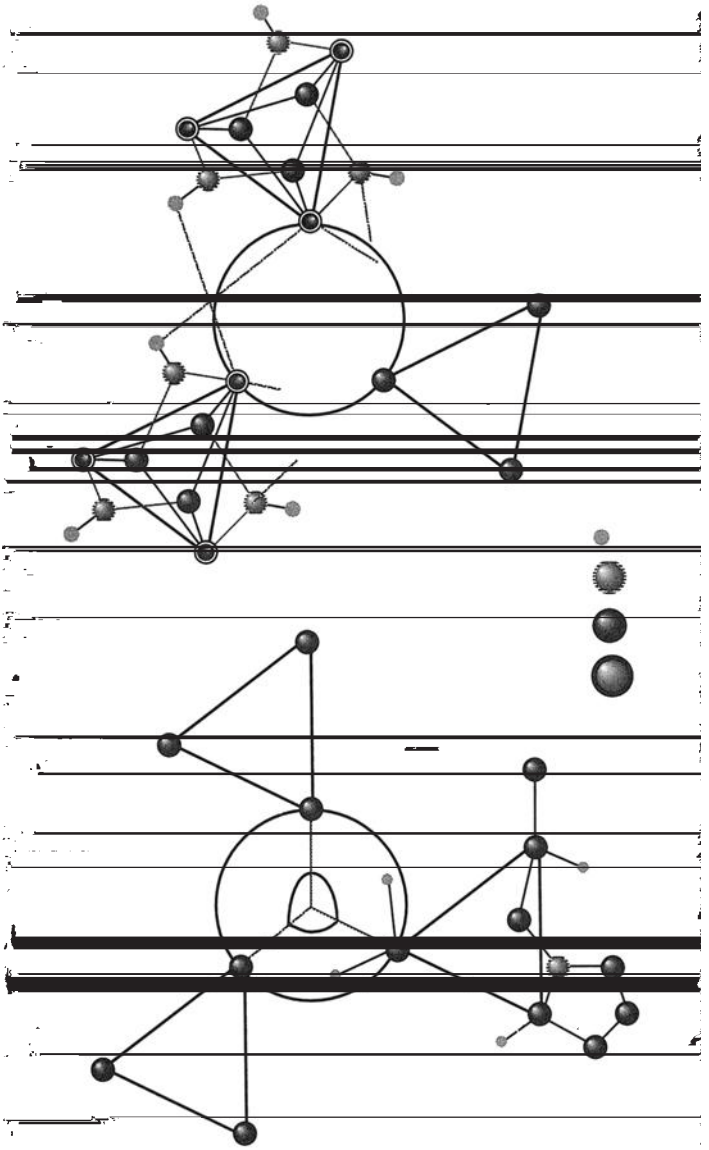


FIGURE 2.20. Models of collagen structure. (**A**) Model of three parallel left-handed helixes of collagen showing the location of Cα (C) for chains A, B, and C. Note all glycines are found in C-1 position because this is the only amino acid residue that can be accommodated at the center of the triple helix. Later studies by Ramachandran and co-workers indicated that the three chains are wrapped around each other (**B**) in a right-handed superhelix. The axial rise per residue is 0.29 nm, and the axial displacement of different Cα atoms is shown in parentheses in angstroms.

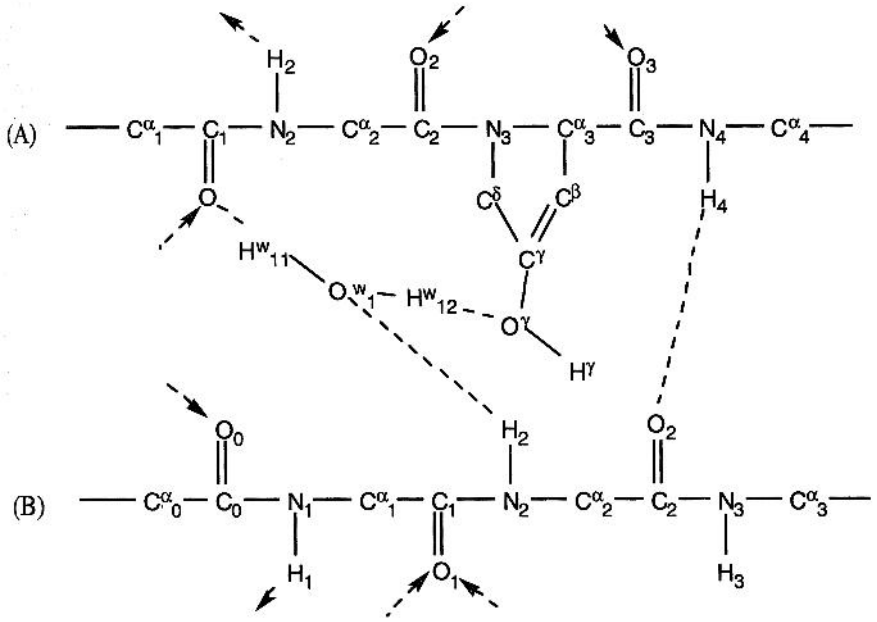


FIGURE 2.21. Stabilization of collagen triple helix. The diagram shows hydrogen bonding between the amide hydrogen in position 4 on chain A and the carbonyl oxygen in position 2 on chain B. A second water-mediated hydrogen bond occurs when hydroxyproline is present in position 3 on chain A between the carbonyl oxygen in position 1 on chain A and the amide hydrogen in position 2 on chain B.

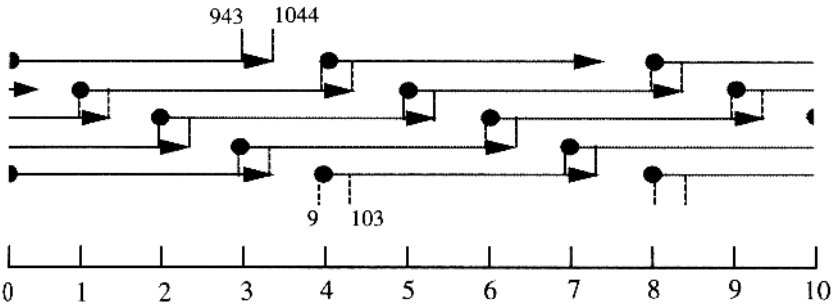


FIGURE 2.22. Cross-linking of collagen molecules in quarter-stagger packing pattern of collagen in fibrils. Each molecule is 4.4-D long (where D is 67 nm) and is staggered by D with respect to its nearest neighbors. A hole region of 0.6 D occurs between the head (circles) of one molecule and the tail of the preceding molecule (arrowheads).

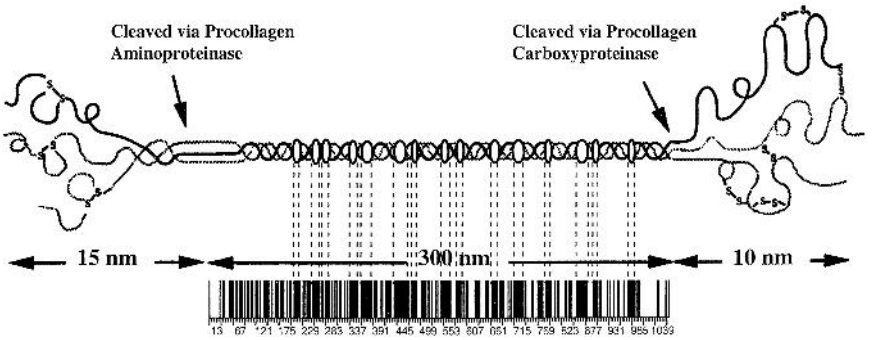


FIGURE 2.23. Diagram illustrating the structure of procollagen, the precursor form of collagen. Procollagen is cleaved to collagen enzymatically during collagen fibril formation. The collagen triple helix is composed of alternating flexible domains (circle) that alternate with rigid domains. The alternation of these domains can be seen as a series of dark (flexible) and light (rigid) regions at the bottom. The collagen triple helix is about 300 nm long and 0.15 nm wide.

collagen fibril. These flexible sequences in collagen are believed to be important in energy storage during mechanical deformation and in attachment to the cell surface through integrin molecules.

Although helical structures dominate the structural hierarchies found in proteins, random chain structures similar to those found in natural rubber also are found in tissues. The most-studied random chain polymer found in vertebrate tissues is elastin.

2.2.3.4.2 Random Chain Coils

Although true randomly coiled chains don't exist in vertebrate tissues, because these structures are susceptible to rapid hydrolysis, sequences that are found in proteins such as elastin are believed to form rubberlike regions that store energy entropically by changing conformational angles during mechanical deformation.

Elastic fibers form the network in skin and cardiovascular tissue (elastic arteries) that is associated with elastic recovery. Historically the recovery of skin and vessel wall on removal of mechanical loads at low strains has been attributed to elastic fibers. Elastic fibers are composed of a core of elastin surrounded by microfibrils 10 to 15 nm in diameter composed of a family of glycoproteins recently termed fibrillins. Fibrillins are a family of extracellular matrix glycoproteins (MW about 350,000) containing a large number of cysteine residues (cysteine residues form disulfide crosslinks). Several members of the family have been described. The common molecular features include: N and C terminal ends with 47 tandemly repeated epi-

dermal growth-factorlike modules separated by a second repeat consisting of eight cysteine residues and other structural elements. Several possible structures for fibrillin have been postulated including unstaggered parallel arrangements and staggered parallel arrangements (Figure 2.24).

Elastin is typically considered as an amorphous protein consisting of random chain sequences connected by α helical regions. The elastin content varies in elastic fibers such as those found in skin. Elastic fibers are termed oxytalan fibers in the upper dermal layer of skin and they are termed elaunin fibers in the deeper dermis where their elastin content is higher. In vessel wall elastic fibers have recently been differentiated based on histological staining patterns suggesting that differences in mechanical properties of different vessel walls may in part be due to differences in elastin

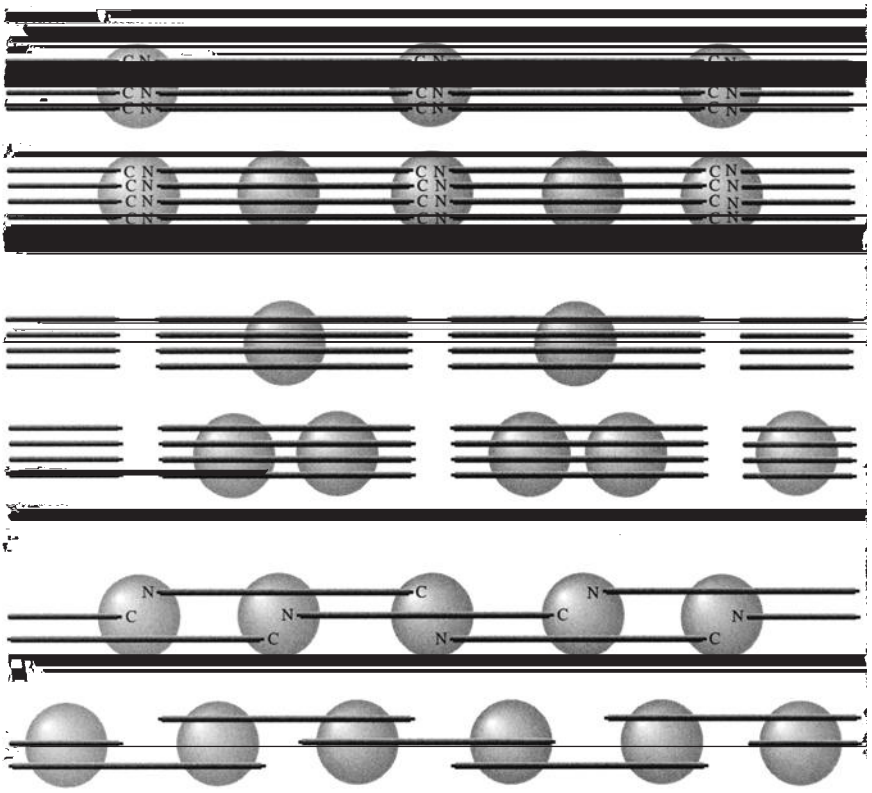


FIGURE 2.24. Structural models for fibrillin in beaded microfibrils. Fibrillin molecules are modeled as being arranged head to tail in parallel arrangements. In some of the models, the molecules are staggered with respect to their neighbors. The circles represent areas where the beads are observed.

concentrations. Mechanically, elastic fibers are in parallel with collagen fibers in skin and vessel wall.

Elastin is a macromolecule synthesized as a 70,000 single peptide chain, termed tropoelastin and secreted into the extracellular matrix where it is rapidly crosslinked to form mature elastin. The carboxy-terminal end of elastin is highly conserved with the sequence Gly-Gly-Ala-Cys-Leu-Gly-Leu-Ala-Cys-Gly-Arg-Lys-Arg-Lys. The two Cys residues that form disulfide crosslinks are found in this region as well as a positively charged pocket of residues that is believed to be the site of interaction with microfibrillar protein residues. Hydrophobic alanine-rich sequences are known to form α helices in elastin; these sequences are found near lysine residues that form crosslinks between two or more chains. Alanine residues not adjacent to lysine residues found near proline and other bulky hydrophobic amino acids inhibit α helix formation. Additional evidence exists for β structures and β turns within elastin thereby giving an overall model of the molecule that contains helical stiff segments connected by flexible segments.

2.2.4 Examples of Other Proteins

Most proteins are a combination of the four structures described above arranged into three-dimensional structures. This results in proteins that have a variety of functional units that have very precise three-dimensional structures. Below we take a look at some of the proteins that are involved in mechanobiology of vertebrate tissues.

2.2.4.1 Keratins

Keratins are found in the superficial layer of skin as well as in the intermediate filaments that support the cell cytoskeleton. Intermediate filaments (IF) are proteins 80 to 120 Å in diameter found in cytoskeletal fibers that reinforce cells. They have molecular weights ranging from 40,000 to 210,000 and are composed of a central domain of 310 to 350 amino acids. There are six types of intermediate filaments; types I and II are composed of keratins that are the largest group of IFs. There are about 30 different protein chains that make up 20 epithelial keratins and 10 hair keratins. Epidermal keratinocytes synthesize two major pairs of keratin polypeptides: K5/K14 of the basal layer and K1/K10 of the cells in the suprabasal layer. Epithelial keratins are expressed in pairs with type I (K between 10 and 20) having acidic groups and type II (K 1 to 9) having neutral and basic groups. It is believed that keratins are expressed as pairs with one acidic and one neutral/basic chain in each molecule.

The basic structural feature of each double stranded (type I/type II hybrid) molecule is the presence of four interrupted α helical sequences termed 1A, 1B, 2A, and 2B which are interrupted by three nonhelical sequences termed L1, L1–2, and L2 (see Figure 2.25). In addition, there is

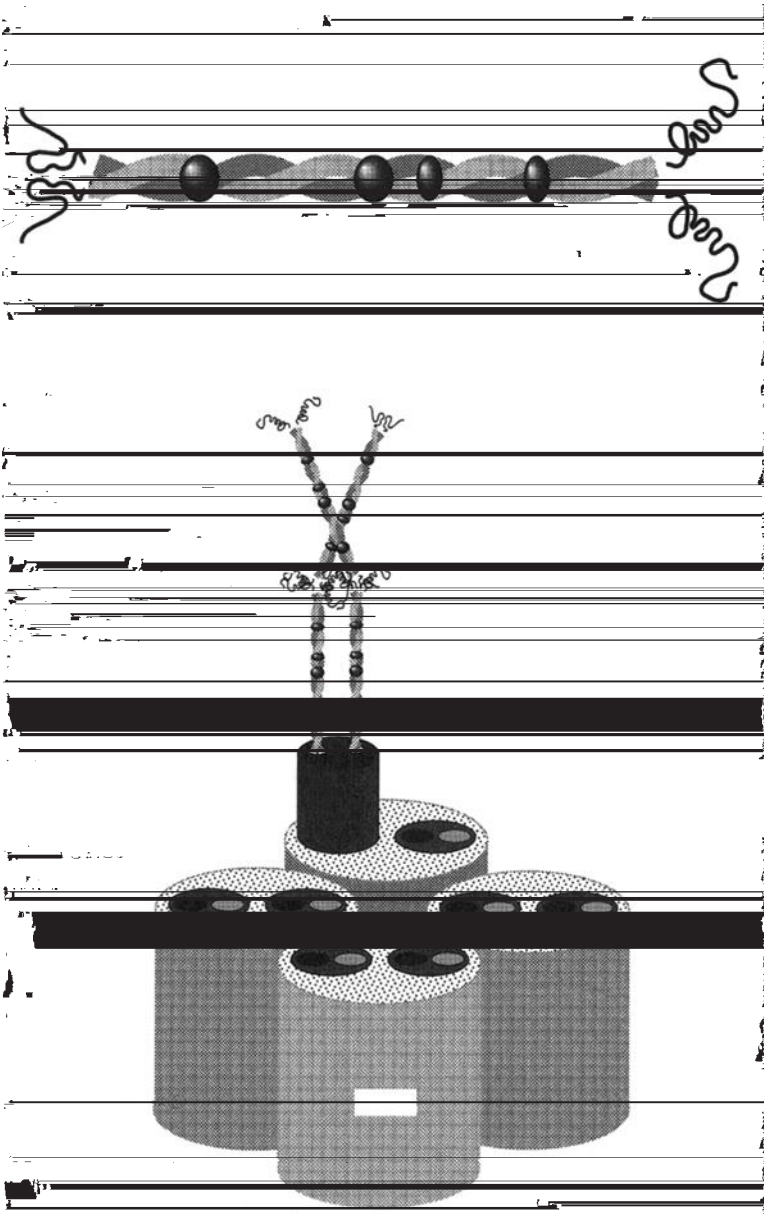


FIGURE 2.25. Structure of keratin protofibrils. The diagram illustrates the structure of keratin in intermediate filaments containing α helical sequences. Two coils are wound around each other and then packed into protofilaments. Eight protofilaments are packed into a filament.

a head and tail domain added to the interrupted α helical domains. The head and tail regions are different for type I and type II keratins. Filament assembly requires at least one type I and one type II keratin that associate parallel and in register to form a coiled-coil. Two coiled-coil dimers then form a ropelike structure in an antiparallel unstaggered or nearly half-staggered array termed a tetramer or protofilament (diameter 2 to 3 nm). Two tetramers form a protofibril (4 to 5 nm) and four protofibrils form filaments 8 to 10 nm in diameter (Figure 2.25).

2.2.4.2 Actin and Myosin

Actin is a protein that is found in cells in the form of filaments and in conjunction with tropomyosin and troponin, forms thin filaments in muscle. Actin exists in two states, a monomeric globular state (overall 3-D sphere-like structure), termed G-actin, and an assembled state, termed filamentous or F-actin. G-actin in the presence of actin binding proteins (ABP) and CA^{+2} and Mg^{+2} assembles into actin filaments in the cell cytoplasm. Actin filaments act as structural supports within the cell cytoskeleton. The 3-D structure of G-actin has been determined from X-ray diffraction studies on complexes containing Ca-ATP-G-actin-DNase I containing small amounts of Mg^{+2} ions (DNase I is added to inhibit F-actin formation). The structure of G-actin at a 2.8 Å resolution consists of small and large domains each divided into two subdomains. ATP is bound in a cleft between the two domains. The actin molecule is composed of helical domains connected by domains containing β extended structures. Actin molecules polymerize into filaments that are coiled-coils described by a left-handed helix with a pitch of 5.9 nm and a right-handed helix with a pitch of 72 nm (Figure 2.26).

Myosin is an enzyme that catalyzes hydrolysis of ATP and converts the energy released into movement through muscle contraction. During muscle contraction, an array of thick filaments containing myosin actively slides by an array of thin filaments containing actin. Myosin has a molecular weight of about 500,000 and consists of six polypeptide chains: two heavy chains (molecular weight 400,000) and two sets of light chains, with each light chain having a molecular weight of 20,000. The molecule consists of a long tail connected to two globular heads as illustrated in Figure 2.27. Each globular head is composed of about 850 amino acids residues contributed by one

FIGURE 2.26. Structure of F-actin. The diagram illustrates assembly of G-actin into double helical segment of F-actin.

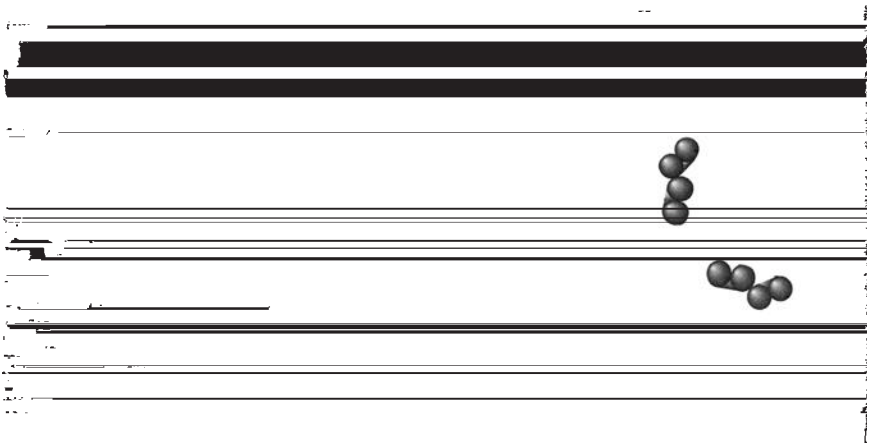


FIGURE 2.27. Structure of myosin. The diagram shows the structure of myosin, which is composed of a helical segment with a flexible bend that is attached to two head units.

of the two heavy chains and two of the light chains. The remaining portions of the heavy chains form an extended coiled-coil that is about 1500 Å in length. The globular head contains the ATP binding sites and the actin-binding region; and the long rodlike portion of myosin forms the backbone of the thick filament. The myosin head has a length of over 165 Å and is about 65 Å wide and 40 Å deep at its thickest end. The secondary structure is dominated by many long α helices and the myosin head is characterized by several prominent clefts and grooves that allow myosin to bind both ATP and actin.

2.2.5 Cell Attachment Factors

The fibronectins are one class of high molecular weight multifunctional glycoproteins that are present in soluble form in plasma (0.3 g/l) and other bodily fluids, and in fibrillar form in extracellular matrix. They bind to cell surfaces and other macromolecules including collagen and gelatin, the unfolded form of collagen, as well as fibrinogen and DNA. Fibronectin mediates cell adhesion, embryonic cell migration and wound healing. It is composed of two chains, α and β with molecular weights of about 260,000 that are covalently linked via two disulfide bonds near the carboxy termini. This macromolecule may contain some β structure and contains a heparin-binding domain, a cell-binding domain, and a collagen/gelatin-binding domain. The molecule can adopt an extended conformation at high ionic strength or a compact one at physiologic ionic strength suggesting that the molecule is flexible (Figure 2.28). Other types of cell adhesion molecules

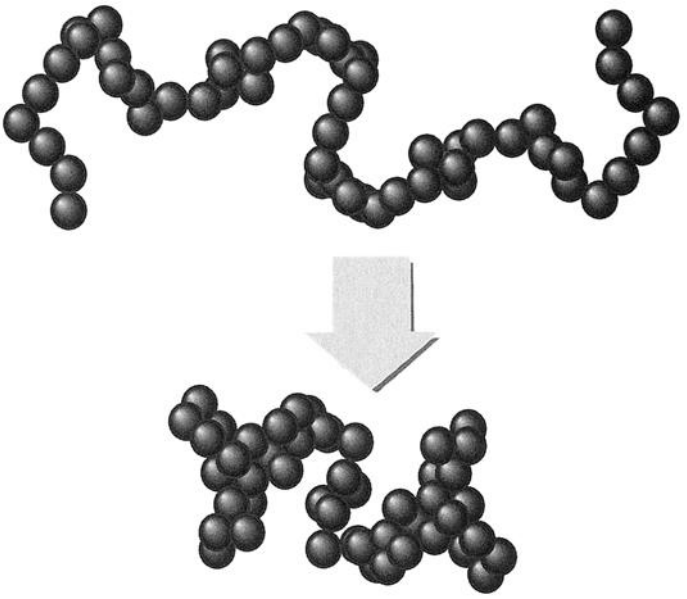


FIGURE 2.28. Structural changes in fibronectin. Structure of fibronectin at high (top) and low (bottom) ionic strength. Note collapse of molecule at low ionic strength.

include laminin, chondronectin, osteonectin, and a variety of other glycoproteins that moderate adhesion between cells and their ECMs.

Laminins are a family of extracellular matrix proteins that are found in basement membrane and have binding sites for cell surface integrins and other extracellular matrix components. They consist of α, β, and γ chains with molecular weights between 140,000 and 400,000. These chains associate through a large triple helical coiled-coil domain near the C-terminal end of each chain (see Figure 2.29). Eight different laminin chains have been identified, α1, α2, α3, β1, β2, β3, γ1, and γ2. The most extensively characterized of the seven forms of laminin is laminin-1 (α1β1γ1), which assembles in the presence of calcium to form higher-ordered structures in basement membranes with type IV collagen (Figure 2.29).

2.2.6 Integrins

Integrins are a family of membrane glycoproteins consisting of α and β subunits (see Figure 1.1). The binding site appears to contain sequences from both subunits, and their cytoplasmic domains form connections with the cytoskeleton. In this manner integrins form a connection between the cytoskeleton and extracellular matrix. In accomplishing this there are 11 α subunits and 6 β subunits forming at least 16 integrins including glyco-

protein IIb/IIa expressed by megakaryocytes (cells in bone marrow from which platelets are derived) and platelets, LFA-1, Mac-1, and p150/95 are expressed by leukocytes. Many integrins bind to components of the extracellular matrix including collagen, fibrinogen, fibrin, laminin, and other proteins. Specific examples include the interaction of endothelial cells with core protein of basement membrane perlecan; the expression of αvβ3 integrin by blood vessels in wound granulation tissue during angiogenesis (formation of new blood vessels); adhesion of platelets and other cell types to collagen involving integrin α2β1; adhesion of natural killer cells (T cells) to fibronectin via VLA-4 and VLA-5; and adhesion of neutrophils to collagen via α2 integrins. Other integrins bind to cell membrane proteins, mediating cell–cell adhesion.

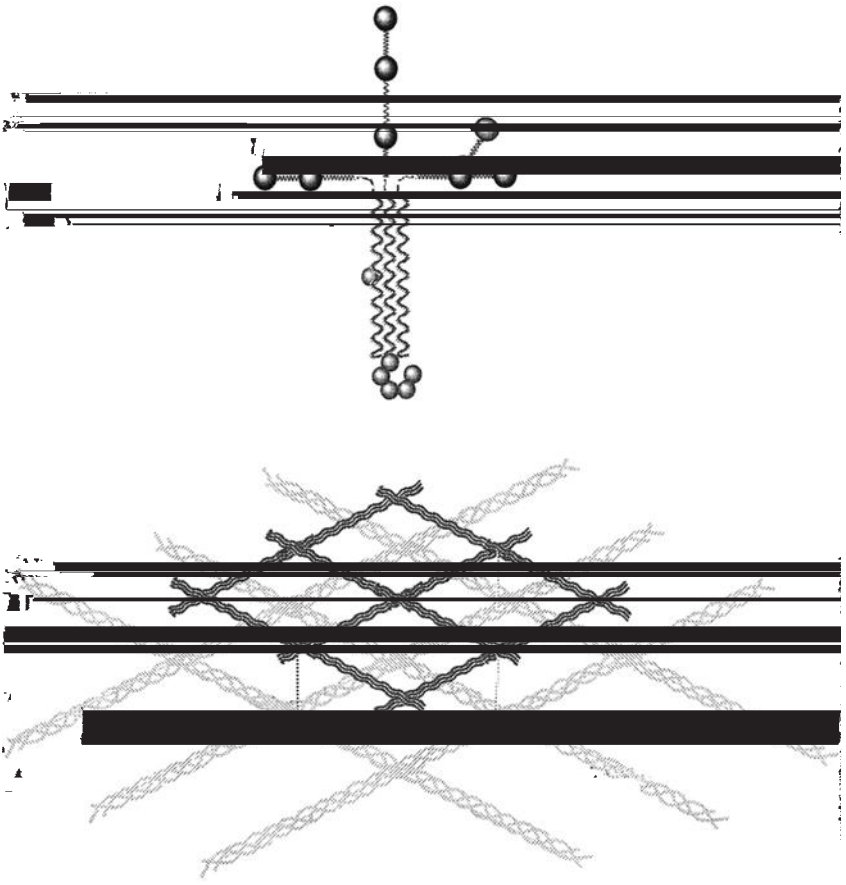


FIGURE 2.29. Network structure of laminin in basement membranes. Basement membranes contain a laminin network that is covalently cross-linked to a type IV collagen network.

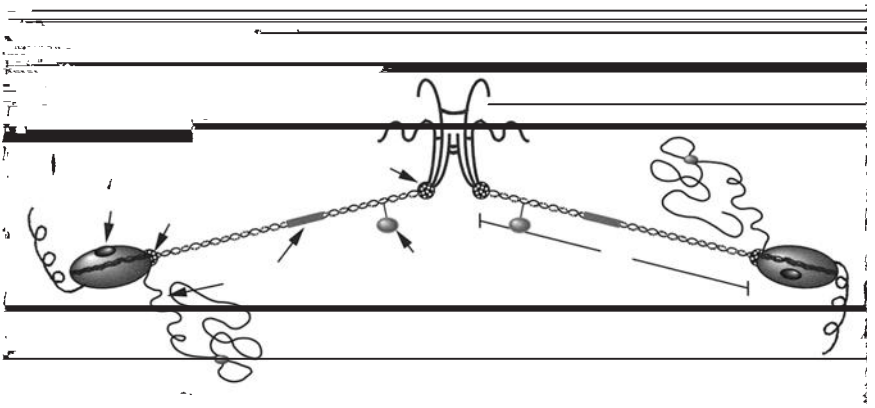


FIGURE 2.30. Domain structure of fibrinogen. The fibrinogen molecule is composed of globular end regions separated from a central domain by helical threads. The central domain has cross-links that hold all the polypeptide chains together. Fibrinopeptides A and B (FPA and FPB) are found in the central domain.

2.2.7 Fibrinogen

Fibrinogen is a plasma protein, formed in the liver that is the basis for the formation of a blood clot. The blood concentration of this protein is about 3% and in the presence of other blood proteins pieces of the fibrinogen molecule, fibrinopeptides A and B, are cleaved and fibrinogen is polymerized to form a fibrin clot (Figure 2.30).

2.2.8 Tubulin

Tubulin is a self-assembling protein that forms microtubules (MTs) within the cell. MTs are involved in a variety of cell functions including vesicle movement, chromosome segregation, and cell motility. MTs are assemblies of heterodimeric proteins, α/β-tubulins. The protein consists of two subunits, a modified α-tubulin and a modified β-tubulin. The two monomers have about 40% of the same amino acid sequences and their structures are similar except for a few differences in the loops. Each monomer contains a pair of central β sheets surrounded by α helices. The tubulin monomers self-assemble into hollow structures, termed microtubules.

2.3 Polysaccharide Structure

Although sugar polymers are less abundant on a mass basis throughout vertebrate tissues they still are important components of the cell surface and extracellular materials. Unlike proteins that are made up of four basic struc-

tural forms, polysaccharides are found in a variety of conformational forms. These forms are a consequence of the flexibility of polysaccharides.

2.3.1 Stereochemistry of Sugars

The stereochemistry of polysaccharides is found using procedures similar to those used with polypeptides. The only real difference is that the repeat unit is a sugar and not a peptide unit. Polysaccharides are found as free molecules such as glycogen, starch, hyaluronan, and as side chains on molecules such as proteoglycans. There are three possible conformations of the basic glucose repeat unit: ^{4}C1 or C1 chair conformation, ^{1}C4 chair conformation, or the boat conformation (see Figure 2.3). The most observed conformation of glucose and its derivatives is the C1 chair conformation, also known as the ^{4}C1 conformation. Macromolecular repeat units of glucose can be linked through oxygen atoms that are either up, α linkage, or down, β linkage. D and L sugar isomers occur based on the position of the OH groups on the sugar ring.

2.3.2 Stereochemistry of Polysaccharides

Polysaccharides are large molecules formed when sugar rings are polymerized. In this book we are interested in hyaluronan and the sugar side chains of proteoglycans because these macromolecules make up the interfibrillar matrix that surrounds collagen and elastic fibers and cells. These molecules are highly flexible and allow tissues such as cartilage to compress during joint loading and also assist in collagen fibril rearrangement during loading by allowing interfibrillar slippage.

Hyaluronan (HA) is a component of every tissue or tissue fluid in higher animals. The highest concentrations are found in cartilage, vitreous humor, and umbilical cord and even blood contains some HA. At physiologic pH, the molecule has been shown to adopt a helical structure and therefore the molecule can be modeled as a series of helical segments that are connected with flexible segments. However, based on results of solution studies, the molecular structure of HA appears to be more complicated.

Stereochemical calculations have been conducted on polysaccharides similar to those done on polypeptides to determine the allowable conformations for different repeat units. Figure 2.31 is a diagram showing a repeat unit containing β-D-glucuronic acid and β-D-N-acetyl glucosamine (hyaluronan). Both units are derivatives of glucose and are connected via oxygen linkages where the position of the linkage is either oriented equatorially (i.e., the bonds of the side chains are not perpendicular to the chain backbone), which is termed a β linkage, or they can be perpendicular to the backbone, which is termed an α linkage. If the linkage involves the carbon at the first and third position we refer to it as either an α or β 1–3 linkage

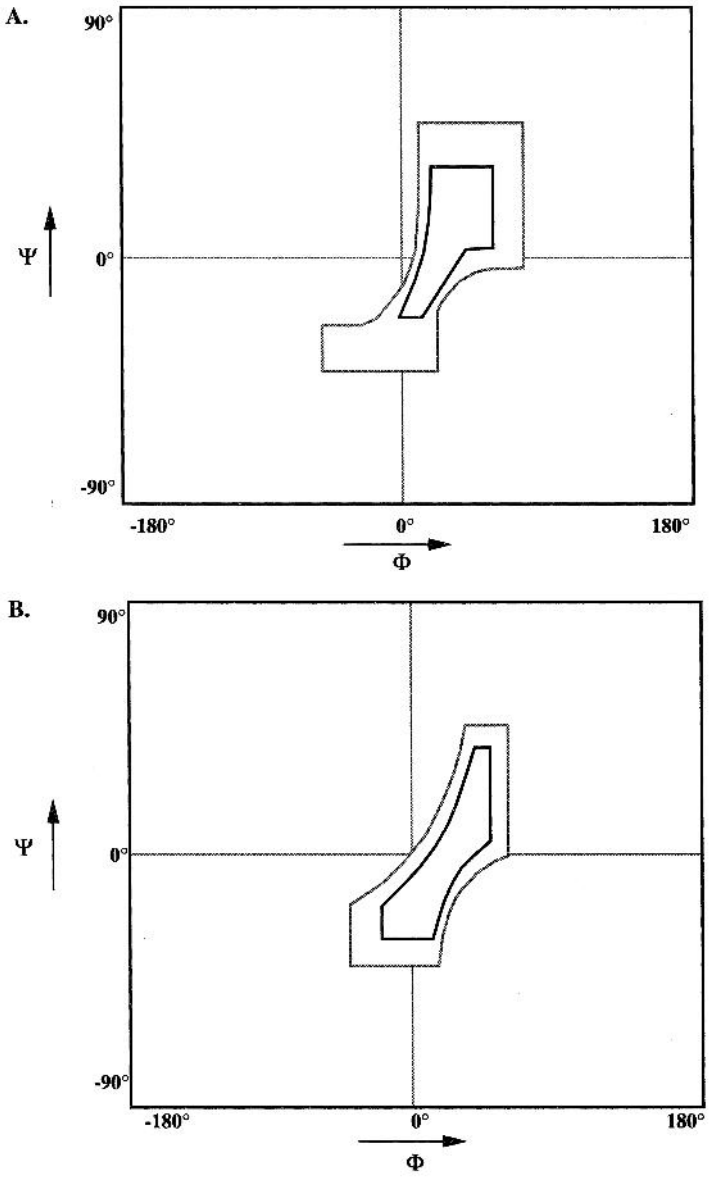


FIGURE 2.31. Conformational plots for β-(1–3) and β-(1–4) linkages in hyaluronan. Plots of fully allowed (inner solid lines) and partially allowed (outer solid lines) conformations of ϕ and ψ for (**A**) D-glucuronic acid, which is β-(1–3)-linked to *N*-acetyl glucosamine, and (**B**) *N*-acetyl glucosamine, which is β-(1–4)-linked to D-glucuronic acid. Note that the allowed conformations center around 0°,0° and show that hyaluronan has flexibility.

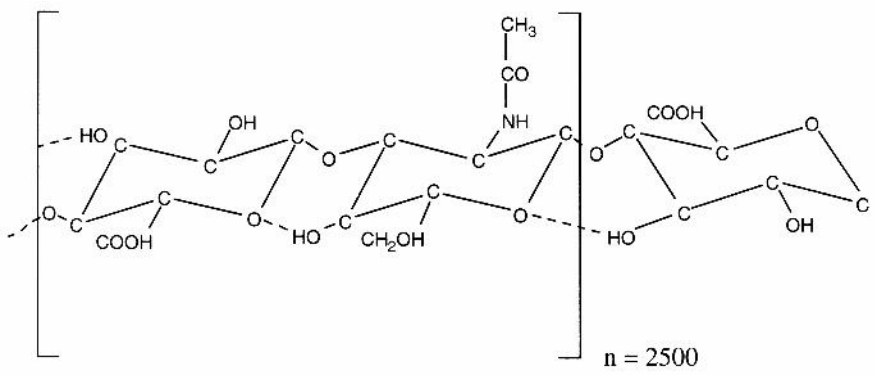


FIGURE 2.32. Repeat disaccharide of hyaluronan. Hyaluronan is composed of β-(1–3) linkage of glucuronic acid to D-*N*-acetyl glucosamine that is linked β-(1–4) to D-glucuronic acid. Both β-D-glucuronic acid and β-D-*N*-acetyl glucosamine are in the C_1 chair conformation. The 0°,0° conformation shown occurs when the atoms attached to the carbons connected to the oxygen linking the sugar units are along the axis of the unit (axially) and eclipse each other when viewed along the chain.

and if it involves the first and fourth carbon we refer to the linkage as either α or β 1–4.

The repeat disaccharide shown in Figure 2.32 is composed of D-glucuronic acid β (1–3) linked to D-N-acetyl glucosamine linked β (1–4) to D-glucuronic acid. Both sugar units are shown in the C1 chair conformations. The 0°,0° conformation that is illustrated occurs when the atoms attached to the 1 and 3 or 1 and 4 positions (i.e., the axial-oriented side chains) eclipse each other when viewed from a plane perpendicular to the bonds. The stereochemical plot that is produced when the dihedral angles are rotated through 360 degrees is shown in Figure 2.31. The stereochemical plot for HA shows a limited number of conformations, that is, about 4% of the theoretical total compared to the stereochemical plots for proteins. It is interesting to note that as we discuss further below, hyaluronan behaves to a first approximation at high shear as a flexible molecule as opposed to collagen, which is more rigid and attains its flexibility in another manner. Hyaluronan is flexible because even though the conformational plot has what appears to be a limited number of allowable conformations that center around 0,0, flexibility is associated with the continuous range of allowable conformations around the 0,0 position. What this tells us is that polymers of glucose and glucose derivatives such as cellulose have inherent chain flexibility; however, introduction of a hydrogen bond can limit this flexibility as has been postulated for HA. It turns out these polymers are also thixotropic; that is, at low shear they are rigid and at high shear they are flexible. This property of thixotropy of high molecular weight polysaccharides is a reflection of their stereochemistry and the ability to form hydrogen bonds along the chain backbone that are broken at high strain rates.

2.3.2.1 Supramolecular Structure

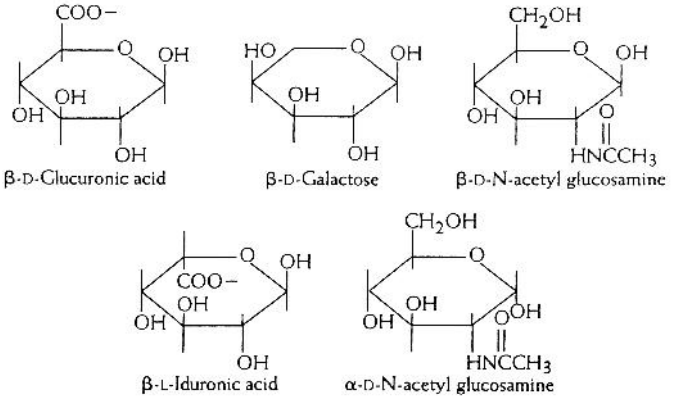
A number of helical structures have been proposed for HA in the solid state based on results of X-ray diffraction studies. Although these reports are more than isolated studies, the majority of solution data suggests that HA exists in solution and probably in the interfibrillar matrix as flexible coils that form entangled domains at high molecular weight. Therefore HA and other polysaccharides do not appear to form higher-ordered structures but exist as a series of flexible domains that are characterized by viscosities that are low enough to promote fluid flow during shearing of collagen and elastic fibers. In cartilage the high charge density of negatively charged GAG side chains of proteoglycans (see below) prevents tissue compression and limits flow outflow during locomotion. A diagram of the extended structure of HA is found in Figure 2.32.

2.3.3 Structure of Glycosaminoglycans

Most of the polysaccharides of interest in this text are termed glycosaminoglycans, polymers that contain an amino sugar in the repeat unit. Glycosaminoglycans that are abundant in mammalian tissues include hyaluronan, chondroitin sulfate, dermatan sulfate, keratan sulfate, and heparin–heparan sulfate (see Table 2.4). Most of these glycosaminoglycans,

Table 2.4. Glycosaminoglycans found in mammalian tissues

Glycosaminoglycan	Repeat disaccharide
Hyaluronan	β-D-Glucuronic acid + β-D-N-acetyl glucosamine
Chondroitin sulfate	β-D-Glucuronic acid + β-D-N-acetyl galactosamine
Dermatan sulfate	β-D-Glucuronic acid or β-L-Iduronic acid + α-D-N-acetyl galactosamine
Heparin, heparan sulfate	β-D-Glucuronic acid or β-L-Iduronic acid + α-D-N-acetyl glucosamine
Keratan sulfate	β-D-Galactose + β-D-glucosamine



with the exception of hyaluronan, are composed of short polysaccharide chains and have limited secondary structure. However, extensive studies on hyaluronan have indicated that it likely has some limited secondary structure. The rate of oxidation (reaction with oxygen) of hyaluronan and chondroitin sulfate indicate that the C(2)–C(3) glycol group in the glucuronic acid moiety is very slowly oxidized. This could be explained by steric hindrance of the glycol group (OH) in hyaluronic acid and chondroitin sulfate (which could hydrogen bond to N-acetyl glucosamine) but not in dermatan sulfate (which contains N-acetyl galactosamine and cannot hydrogen bond to the same glycol groups). The proposed hydrogen bond scheme is shown in Figure 2.32 and is consistent with a highly extended ribbonlike helix with an axial rise per disaccharide of 9.8 Å. The stiffening of hydrogen bond arrays over short segments of the molecule explains the large expanses (space-filling role) that are occupied by HA molecules. A conformational plot for the repeat units found in HA is shown in Figure 2.33 showing the inherent flexibility of the β(1–3) and β(1–4) linkages.

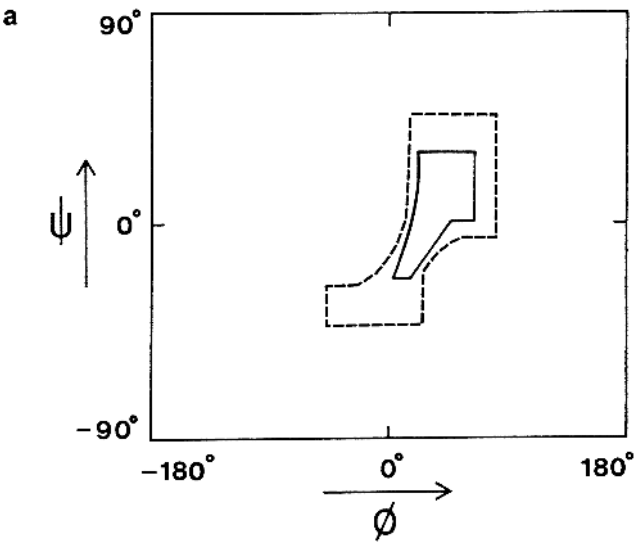


FIGURE 2.33. Conformational plot of β (1→3) and β (1→4) linkages in hyaluronic acid: plots of fully allowed (—) and partially allowed (---) conformations of ϕ and ψ for (a) D-glucuronic acid which is linked β (1→4) to N-acetyl glucosamine and (b) N-acetyl glucosamine which is linked β (1→3) to D-glucuronic acid. Allowed conformations center around (0°, 0°) and indicate that stereochemically the backbone of hyaluronan has some flexibility.

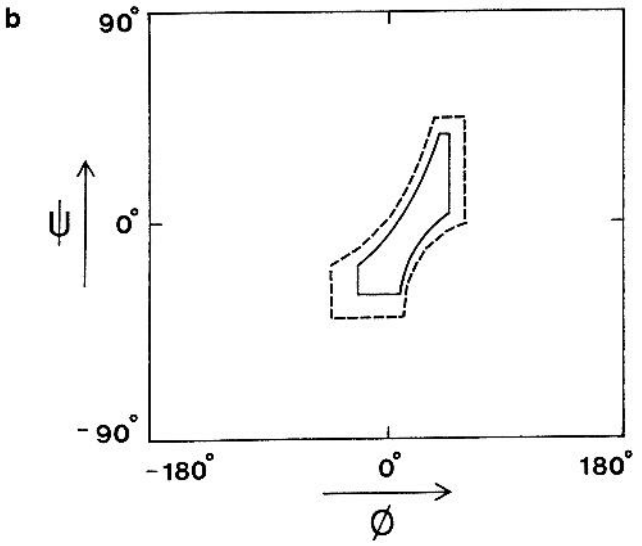


FIGURE 2.33. *Continued*

2.4 Glycoprotein and Proteoglycan Structure

There are other macromolecules found in tissues that are combinations of polypeptides and polysaccharides. These include glycoproteins that are composed of polypeptides to which sugar chains are attached and proteoglycans that are composed of a protein core onto which polysaccharides are grafted. Fibronectin and collagen are actually glycoproteins because both contain sugar rings and sugar polymers that are attached to these molecules.

Proteoglycans are a diverse family of glycosylated proteins, which contain sulfated polysaccharides as a principal constituent. Diverse structures are found for these molecules that depend on the type of glycosaminoglycan side chains, length and net charge. Aggrecan found in large amounts in cartilage (50 mg/g of tissue) is highly glycosylated with 200 chains containing chondroitin sulfate and keratan sulfate. The central core protein is about 220,000 in molecular weight with the overall aggrecan weight reaching 2 to 3 million (Figure 2.34). It interacts with hyaluronan via link protein. In contrast decorin and biglycan have relatively small core proteins (about 40,000) that have a leucine rich repeat and have one (decorin) and two (biglycan) chondroitin/dermatan sulfate side chains (Figure 2.34). Decorin binds to collagen fibrils whereas biglycan does not (Figure 2.35). A unique heparan sulfate proteoglycan, perlecan, is the major PG found in basement membranes. It has a large protein core containing about 3500 amino acids, consisting of multiple domains. Perlecan is able to self-associate or interact with several other basement membrane macromolecules including laminin and type IV collagen. In the kidney, heparan sulfate PG contributes a negative charge to the basement membrane and is thought to exclude serum

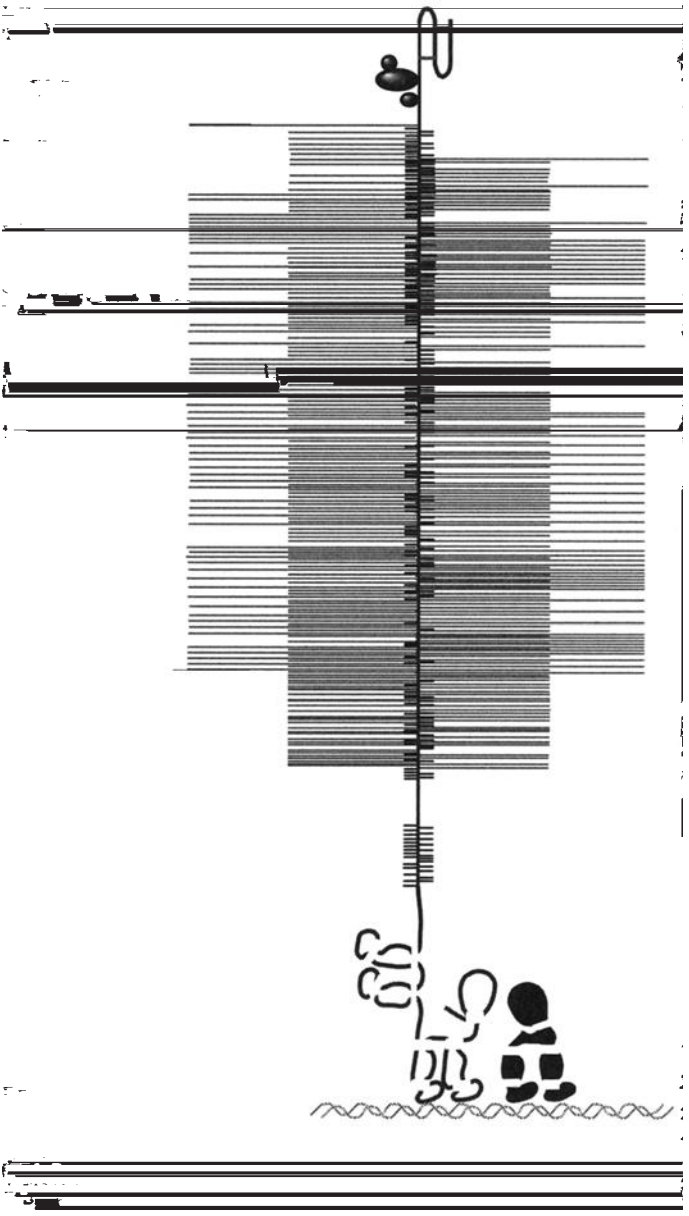


FIGURE 2.34. Interaction between aggrecan and hyaluronan. The diagram illustrates the interaction between large aggregating proteoglycan (aggrecan), link protein, and hyaluronan.

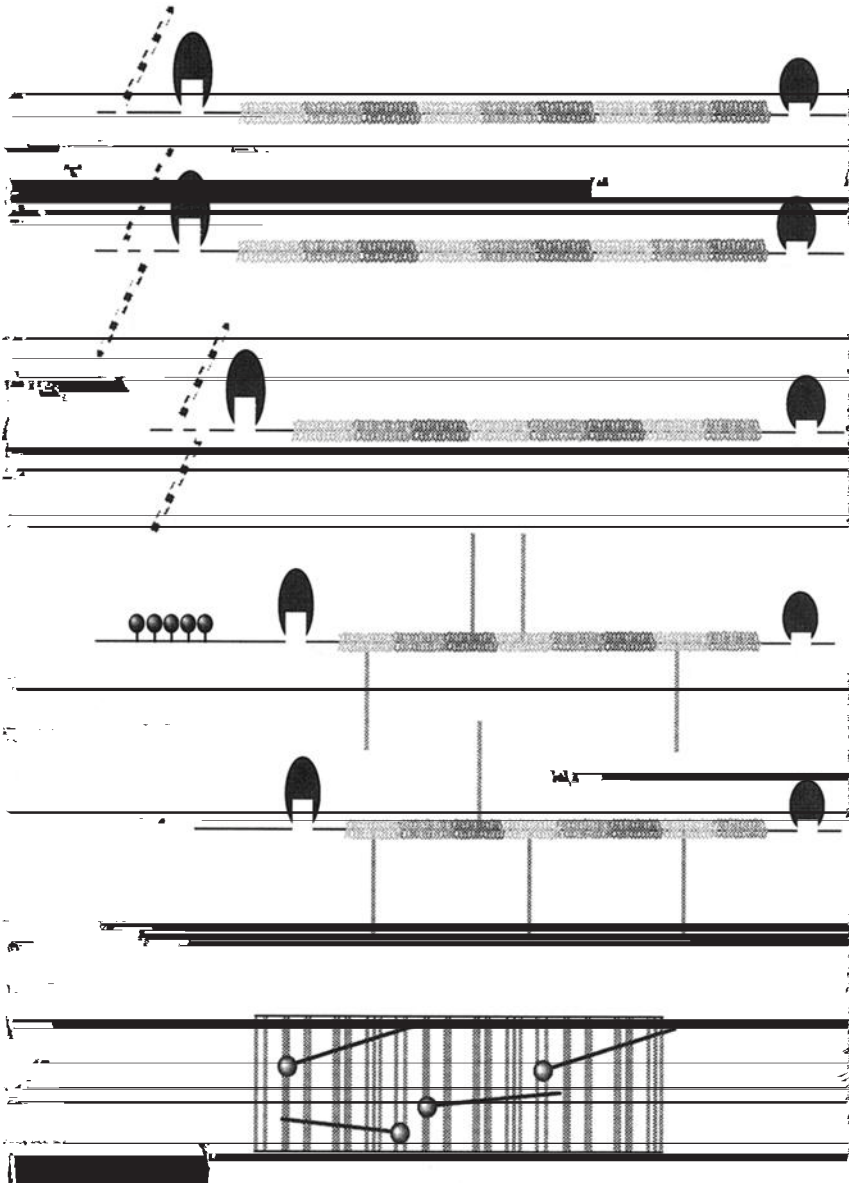


FIGURE 2.35. Comparison between proteoglycan structures. The diagram illustrates structures of decorin (one glycosaminoglycan side chain), biglycan (two glycosaminoglycan side chains), proteoglycan-Lb, fibromodulin, and lumican. Also shown is the specific binding of decorin to the d and e bands on collagen fibrils.

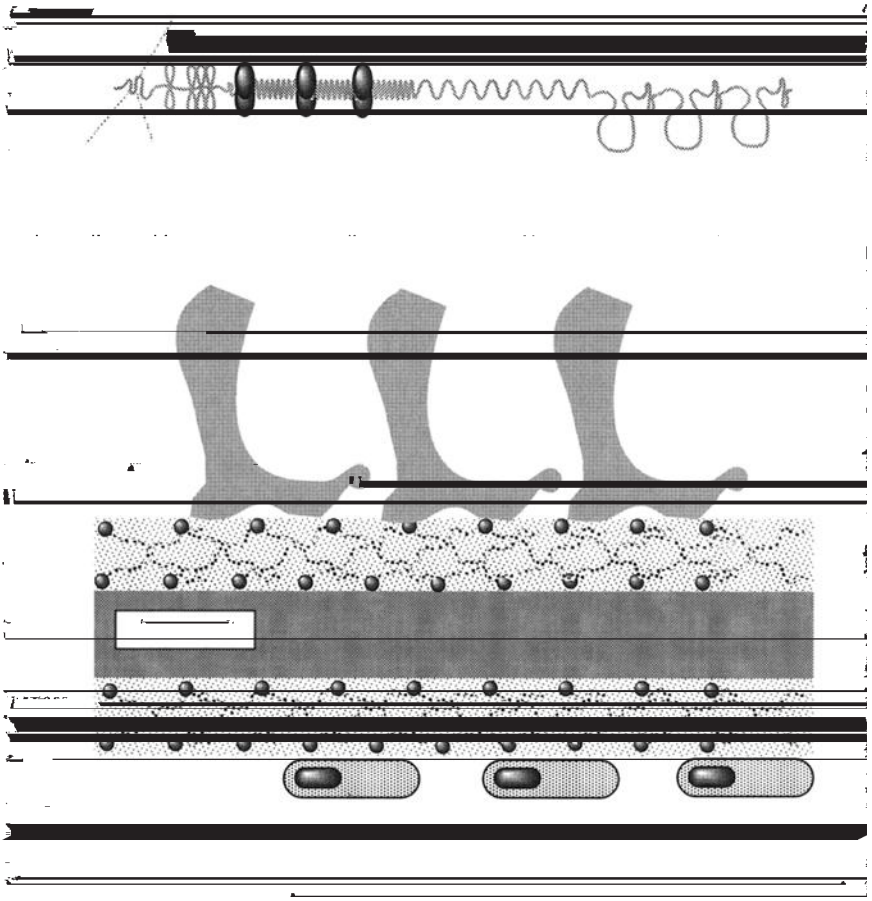


FIGURE 2.36. Model of perlecan in basement membranes. The illustration shows the structure of perlecan, a basement membrane proteoglycan, and localization of proteoglycans in the basement membrane.

proteins from being filtered out of the blood. Although the structure of perlecan is still not completely understood a model has been developed (Figure 2.36).

Virtually every mammalian cell has heparan sulfate proteoglycans as a plasma membrane component. They are inserted into the cell membrane either through a transmembrane domain in their core protein (syndecans), or via a modified glycosaminoglycan region that is linked to the cell membrane (glypican). Heparan sulfate PGs interact with numerous molecules such a growth factors, cytokines, extracellular matrix proteins, enzymes, and protease inhibitors. They are believed to act to transduce signals that emanate from the interplay between components in the extracellular

matrix. Syndecan is associated with epithelial cell differentiation after migration and with the intracellular actin cytoskeleton; expression of syndecan by cells appears to inhibit epithelial cell invasion into collagen.

Syndecan consists of a core protein that is inserted through the cell membrane and contains both heparan sulfate and chondroitin sulfate chains. Glypican is covalently linked to the head group of membrane phospholipids in the plasma membrane and contains only heparan sulfate side chains.

2.5 Stereochemistry of Lipids

Lipids are the major components of cell membrane and are also found in blood. The stereochemistry of lipids is very similar to that of poly(ethylene) chains and proteins with small side chains. A freely rotating polymer chain (lipids without double bonds in the backbone) has a stereochemical plot that is very similar to Figure 2.11. Therefore most hydrocarbon chains are quite flexible and adopt a number of conformations. In order to pack hydrocarbon chains efficiently into the cell membrane, the hydrocarbon component is in a planar zigzag. The reason that the hydrocarbon component is only 14 or so carbon atoms long in cell membranes is that longer chains would crystallize as a result of the van der Waals bonds that form between hydrogen atoms between the chains. That is exactly what happens with poly(ethylene) when it is polymerized into chains with molecular weights exceeding 100. This is easily illustrated by examining the physical form of hexane which is a low molecular weight poly(ethylene). At room temperature hexane is a liquid. As the chain length of the hydrocarbon is increased the material is first a wax (i.e., it will flow under pressure and heat such as paraffin) and then at higher molecular weights the material becomes a solid, although if we heat it enough the solid will melt. Therefore, when the chains get long enough the secondary forces between them prevent the molecules from being flexible and allow the chains to crystallize. The flexibility of the chains can be assessed by noting the viscosity of gasoline is very low compared to that of a 2% solution of carboxy methylcellulose (CMC), a food additive used to make food products thick.

The small nature of the hydrogen atoms attached to the backbone and the similarity in size (the chain is composed of a carbon backbone with hydrogen side chains) allow these polymer chains to pack efficiently. They pack efficiently because of the stereochemistry of the side chains. If we look at a drawing of a hydrocarbon chain in a zigzag conformation, the hydrogens as viewed from the side of the chain are staggered and alternately fall between the other hydrogens (see Figure 1.7) or they can be positioned on top of the previous ones (eclipse each other). As we discussed above when we referred to Equation (2.2), the Gibbs free energy should be minimized in the most favorable conformation. The free energy is minimized when the hydrogens are staggered with respect to the backbone of the hydrocarbon chain.

2.6 Stereochemistry of Nucleic Acids

DNA and RNA are the blueprints for making proteins and other polymers involved in cell and tissue metabolism. Examples of the repeat units in DNA and RNA are given in Figure 2.4. The stereochemistry is similar to that of polysaccharides in that a sugar unit is linked through oxygen to another sugar unit. The difference is that the oxygen linkage has several other atoms attached to it between the sugars giving additional flexibility to the chain backbone. Therefore the stereochemical plot for nucleic acid polymers will be similar to that of polysaccharides except there will be additional allowed conformations. This flexibility is necessary for DNA so that it can fold into a double helix as we discuss below.

2.6.1 Primary and Secondary Structure of DNA and RNA

The structure of DNA was originally thought to contain equal amounts of the four purines and pyridines. By the late 1940s it was found that the ratios of adenine and thymine were always very close to unity; the same was true for guanine and cytosine. This implied that for some reason, every molecule of DNA contained equal amounts of adenine and thymine and also equal amounts of guanine and cytosine. Using this chemical information together with X-ray diffraction patterns of DNA, a model was proposed for the structure of DNA in the early 1950s. It was proposed that a molecule of DNA consists of two helical polynucleotides wound around a common axis to form a right-handed "double helix". In direct contrast to the arrangements in helical polypeptides (where the amino acid side chains are directed to the outside of the helix), the purine and pyrimidine bases of each polynucleotide chain are directed towards the center of the double helix in a manner such that they faced each other. Based on stereochemical considerations, it was further suggested that the only possible way that the nitrogen bases could be arranged within the center of the double helix that was consistent with the predicted dimensions was that in which the purine always faced the pyrimidine. Based on consideration of the possible hydrogen bond patterns between purines and pyrimidines it was concluded that adenine must be matched with thymine and guanine with cytosine.

The parameters of the double helix that is formed by DNA include a diameter of 20 Å, a rise per nucleic acid residue of 3.4 Å with ten residues per complete turn. The two chains that make up the molecule are antiparallel; the chains grow by adding repeat units to the 3′ group on ribose or the 5′ group on ribose and therefore one chain is joined 3′ to 5′ by phosphodiester bonds and in the other chain the riboses are joined by 5′ to 3′. The two polynucleotides are twisted around each other in such a way as to

produce two helical grooves in the surface of the molecule. The structure of RNAs are different than DNA and do not form a double helix but fold to form different structures such as the large and small subunit of the ribosome.

2.7 Relationship Between Higher-Order Structures and Mechanical Properties

We have learned about the types of macromolecular structures found in tissues. These include helices, extended structures, and random coils. Ultimately, the properties of biological polymers are dictated not only by the macromolecular structure, but also by the levels of structural hierarchy that are found in tissues. For instance, mechanical loading very easily deforms elastin. In contrast, structures containing keratin are less easily deformed not only because they are made up of α helices, but also the α helices are packed into higher-ordered structures. It is the higher-ordered structures that in the case of keratin and even collagen dictate whether a tissue is soft and pliable or hard and rigid. Therefore although helical molecules are in general more difficult to deform (require more force per unit area) than random coils, the presence of higher-order structure gives biological systems the ability to tailor structures and therefore physical properties. The relationship among macromolecular structure, energy storage, and mechanotransduction is explored further in later chapters. The purpose of this chapter is to underscore that force transfer is best accomplished using a polymer that is in an extended conformation. This allows stress to be transferred with a minimum of deformation.

2.8 Summary

Biological tissues contain a variety of different types of macromolecules that exist in α helices, single and double helices, extended chain structures, β pleated sheets, random coils, and collagen coiled-coils that form supramolecular structures that maintain cell and tissue shape, resist mechanical forces, act to transmit loads, generate contractile forces, provide a mechanical link between extracellular matrix and the cell cytoskeleton, and act in recognition of foreign cells and macromolecules. The key element needed to understand how these molecules form is: (1) the chemistry of the repeat unit, (2) the nature of the backbone flexibility, (3) the types of hydrogen bonds that form, (4) the nature of the secondary forces other than H-bonds that form between the chains, (5) the manner in which individual chains fold, (6) the way in which folded chains assemble with other chains, and (6) the manner that assembly is limited. Unfortunately, we are only beginning

to understand the intricacies of how chain folding and assembly lead to biological form and function. However, the little we know makes this field so very exciting because nature has created some very clear structure–function relationships. At the very least we now know that structural materials that are aimed at reinforcing tissues have many levels of organization that pack molecules into a regular array. For instance, materials that provide tissue integrity (i.e., keratin and collagen) contain linear regions of highly ordered structure and crosslinks within the molecule to prevent extensive molecular slippage. Macromolecules that transmit force to cells such as actin in the cell cytoskeleton, and myosin in skeletal muscle contain more globular regions that are connected by turns and bends.

Suggested Reading

Laurent TC, Fraser RE. Hyaluronan, *FASEB J.* 1992;6:2397.

Pauling L, Corey RB. The polypeptide conformation in hemoglobin and other globular proteins, *Proc Natl Acad Sci.* 1951;37:282.

Pauling L, Corey RB. Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets, *Proc Natl Acad Sci.* 1951;37:729.

Ramachandran GN, Kartha G. Structure of collagen, *Nature.* 1955;176:593.

Ruoslahti E. Integrins, *J Clin Invest.* 1991;87:1.

Schultz GE, Shrimer RH. Patterns of folding and association of polypeptide chains. In: *Principles of Protein Structure*, New York: Springer-Verlag; 1979: Chapter 5, 66–107.