5

Molecular Configuration of Synthetic and Biological Polymers

ALEXANDER RICH

Department of Biology, Massachusetts Institute of Technology, Cambridge 39, Massachusetts

IT is possible to make a rough division in biological systems between large molecules and small molecules. They may be regarded as having quite different roles. Small molecules are actively involved in metabolic cycles, and they provide most of the energy needed for the activity of a biological system. The large molecules often have quite a different role in that their configuration—the three-dimensional organization of the molecule—has a particularly important part in determining the function of the molecule. Indeed, on the molecular level, structure and function are closely related.

Almost all large molecules that are found in biological systems are polymers; that is, they are formed by the linear association of certain fundamental repeating units. This review discusses some of the general structural features of these polymeric molecules, starting with some synthetic polymers which illustrate certain features of structural organization, and then continuing on to a group of more biologically significant synthetic and naturally occurring polymers.

In recent years, there has been an increasing awareness of the widespread occurrence of helical polymer molecules, both from natural and synthetic sources. This is not surprising, since the helix is the most general configuration for a regular polymer. The position of adjacent residues along a helix can be specified by a translation distance and an angular rotation. The most familiar example of a helix is the twofold screw rotation in which adjoining residues are related by an angular rotation of 180°; however, this is a specialized configuration and is discussed in the following.

Most molecular structure determinations are the results of x-ray diffraction studies. For diffraction work on polymers, a very useful expression has been obtained by Cochran *et al.*,[1] who derived the Fourier transform for helical molecules. Their analysis utilizes the helical symmetry in the molecule and expresses the Fourier transform in terms of Bessel-function contributions to various layer lines. In this way, the position and intensity of the x-ray scattering by helical molecules can be computed relatively directly. Their analysis has been programed into computing machines in order to facilitate the testing of a variety of proposed helical molecular models.[2]

SYNTHETIC POLYMERS

Some understanding of the factors which determine the configuration of polymers can be gained from studies of the simpler synthetic polymers. It is illuminating to compare the configuration of polyethylene $-(CH_2)_n-$ with that of polytetrafluoroethylene $-(CF_2)_n-$ (known commercially as Teflon). The polyethylene molecule [Fig. 1(a)] consists of a series of CH_2 groups organized in a zigzag chain in which all of the carbon atoms lie in one plane. The distance between alternate CH_2 groups is 2.54 A. If, however, the hydrogen atoms are replaced by fluorine atoms, there is a change in configuration as is shown in Fig. 1(b).[3] Instead of lying in a planar zigzag, the carbon atoms in the backbone are now

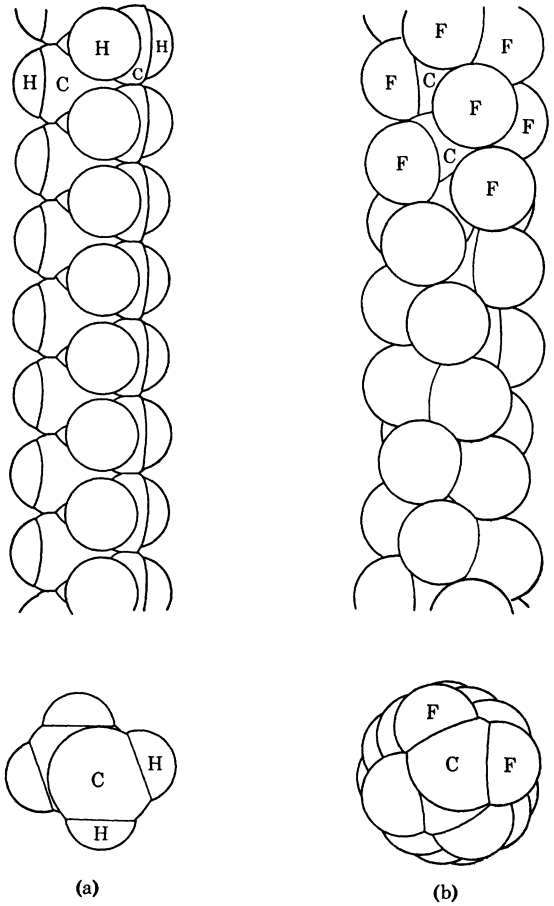


FIG. 1. (a) Planar zigzag chain formed by polyethylene (side and end views). The hydrogen atoms are not in van der Waals contact so that the carbon backbone remains fully extended. (b) Helical-chain configuration formed by the fluorocarbon molecule (side and end views). The bulkier fluorine atoms are in van der Waals contact and force the backbone chain of carbon atoms into a helical configuration [from C. L. Bunn and E. R. Howells, Nature **174**, 549 (1954)].

twisted. In the hydrocarbon chain, adjacent residues are related by a rotation of 180°; in the fluorocarbon chain, this angle has changed to 166°, and the molecule has a helical appearance. The fluorocarbon chain assumes this form as a result of the steric restraints imposed on it by the bulkier fluorine atoms. The hydrogen atoms with a van der Waals radius of 1.1 A are so small that they are not in contact in polyethylene. However, the fluorine atoms with a van der Waals radius of 1.35 A are too bulky to lie in a planar zigzag, and a helical twist is introduced into the carbon backbone in order to fit each fluorine atom into the groove between the two fluorine atoms on adjacent residues. This fairly small structural change brings about a considerable alteration in the properties of the two materials. Looking down the axis of the fluorocarbon molecule [Fig. 1(b)], one can see that the polymer is rounded in cross section. This rounded contour is responsible for the first-order phase transition which occurs near 20°C, and is undoubtedly owing to the rotation of these molecules in the crystalline state about their molecular axes. The hydrocarbon molecule, with its flattened cross section, does not have a phase change of this sort. This is an example of the effect of atomic size on molecular configuration and chemical properties.

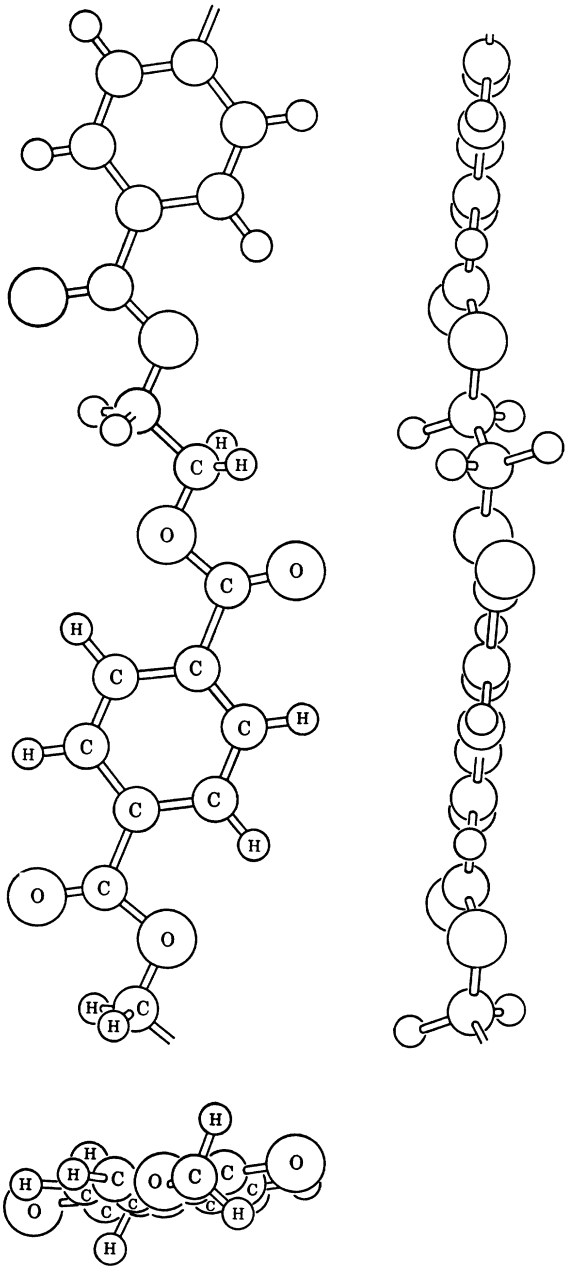


FIG. 2. Configuration of the polyethylene-terephthalate molecule. The unsaturated benzene groups lie nearly in the same plane so that the molecule has a ribbon-like configuration [from R. P. Daubeny, C. W. Bunn, and L. Brown, Proc. Roy. Soc. (London) **A226**, 531 (1954).

Another feature of interest is illustrated by the polymer molecule polyethylene terephthalate (known commercially as Dacron or Terylene). The repeating molecular unit contains a benzene ring as shown in Fig. 2. Succeeding residues lie almost in the same plane, hence the molecule has a ribbon-like configuration. Figure 3 illustrates how these molecules are packed in the crystal.[4] The flat, unsaturated rings lie largely on top of each other. In this orientation, they are stabilized considerably by the van der Waals interaction which is large for the highly polarizable π-electron system of the ring. This stabilization undoubtedly contributes to the high melting point (266°C) observed in the fiber. This type of interaction occurs in biological polymers and is most frequent in the nucleic acids where the large unsaturated ring systems usually pile on top of each other. These are discussed more thoroughly in the chapter devoted to the nucleic acids (p. 191).

When polymers contain electronegative atoms such as nitrogen and oxygen, there is often an opportunity for hydrogen bonds to be important in determining the configuration of the molecule. An example of this type of polymer is seen in polyhexamethylene adipamide, known as 66 nylon.[5] The structure of the α-crystal illustrated in Fig. 4 is most clearly understood by noting that the molecules form a sheet structure in which adjacent units on the "a" face of the unit cell are held together by the horizontal hydrogen bonds which are shown as dotted lines. This polymer is of particular interest, of course, because its residues are joined by peptide bonds ($-$NH$-$CO$-$) as are the amino-acid residues in a polypeptide or in a protein.

Hydrogen bonds represent fairly weak or secondary forces between atoms. Nonetheless, they are often of considerable importance in determining the configuration of polymeric systems, because the total number of bonds involved is of the same order of magnitude as the number of monomer units in the polymer molecule. This is usually a large number so that the total stabiliz-

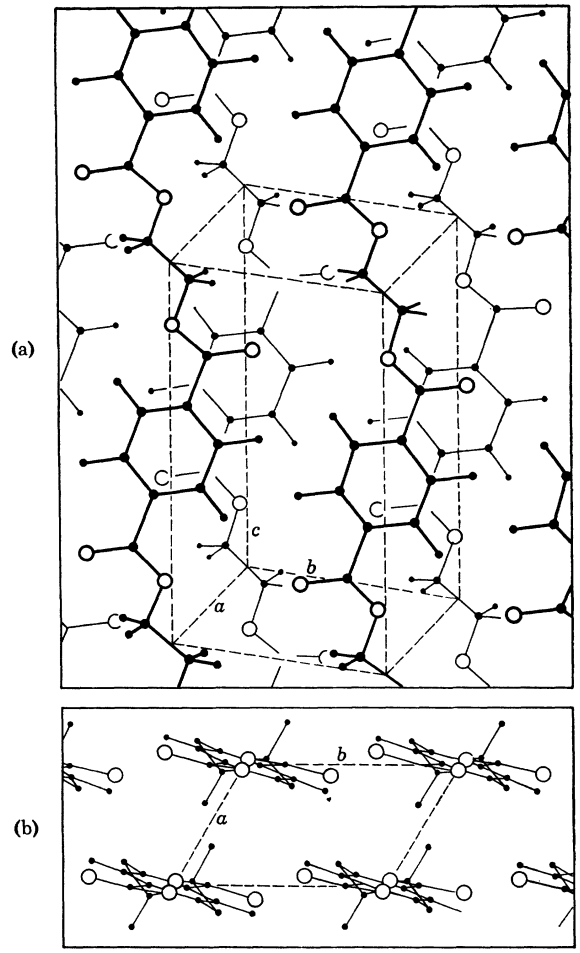


FIG. 3. The arrangement of polyethylene-terephthalate molecules in the crystal. (a) View perpendicular to the fiber axis. (b) View down the fiber axis. The larger dots represent carbon; smaller dots, hydrogen; open circles, oxygen [from R. P. Daubeny, C. W. Bunn, and L. Brown, Proc. Roy. Soc. (London) **A226**, 531 (1954)].

ing energy per covalently-linked molecule may become considerable.

BIOLOGICAL POLYMERS

In all of the polymer structures described in the foregoing, there is one regular repeating unit and the configuration of the molecule is the resultant of systematic and repeating steric effects, van der Waals stabilization, electrostatic forces, and hydrogen bonding. The biological polymers, however, have an additional element of subtlety. In the proteins, the peptide backbone is the repeating unit, while the sugar phosphate groups form the backbone in the nucleic acids. In addition to these repeating units, however, there is also a variety of side groups attached periodically in some definite sequence to the polymer backbone. Hence, in addition to the systematic, repeating interactions described in the foregoing, the configuration of the molecule must be dependent also upon the nature of these side groups and the sequence in which they occur.

In discussing the molecular configuration of biological polymers, one usually describes the results of structural investigations carried out in the solid state. This description is useful, since the solid-state structure often is retained in solution or in the living cell.

The biological polymers may be listed as follows:

(a) The nucleic-acid polymers. These are discussed at length in another chapter (Rich, p. 191). However, it should be mentioned that, in this polymer, each repeating residue has a negative charge located on the phosphate group. Thus, the configuration of a nucleic-acid molecule represents a compromise in which the internally repelling electrostatic forces are balanced by the cohesive forces in the molecule.

(b) The carbohydrate polymers. Only a few such as cellulose and chitin have a regular configuration with the polymer chains organized systematically. The ab-

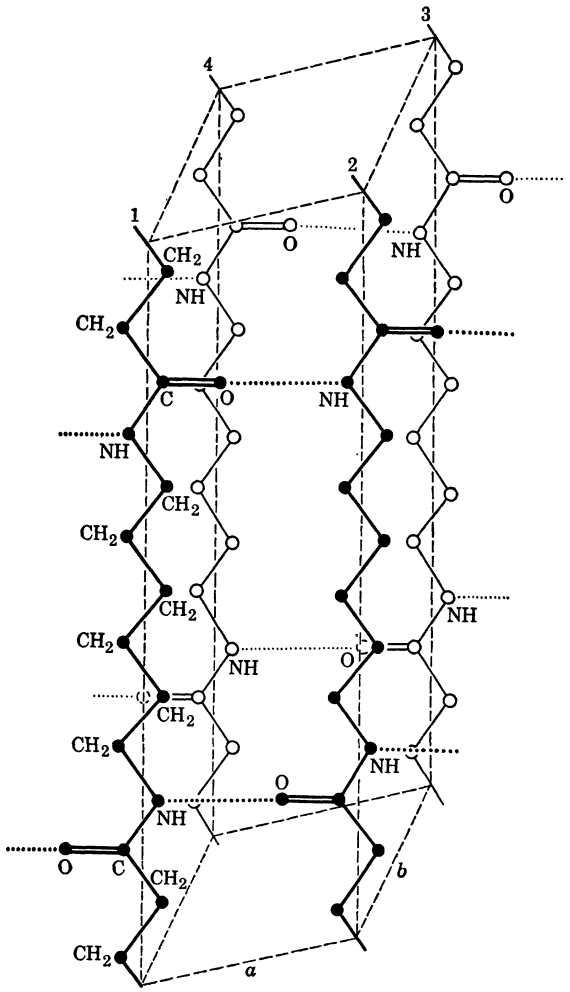


FIG. 4. The molecular structure of the α-crystal of 66 nylon. The molecules are hydrogen-bonded into sheets in the plane of the page [from C. W. Bunn and E. V. Garner, Proc. Roy. Soc. (London) **A189**, 39 (1947)].

sence of a highly ordered configuration in most carbohydrate polymers is probably an indication that their biological function does not require any elaborate three-dimensional organization. Instead, it may depend on the stereochemistry of the sugar-to-sugar linkage. For example, the amorphous particles of glycogen stored in the cell are apparently just a convenient depot from which glucose residues can be removed by enzymatic action when they are needed. These structures are not discussed here.

(c) Proteins. Perhaps the most varied biological polymer is the polyamino-acid chain from which proteins are built. The repeating chemical unit is a single peptide bond with one α-carbon atom between each bond in contrast to the six carbon atoms between peptide bonds in 66 nylon. In the proteins, however, about twenty different types of side chains can be attached to the single α-carbon atom. The nature of these side chains has a profound effect in determining molecular configuration (see following).

(d) Lipids. There are no lipid polymers, even though there are several lipid molecules with extremely long side chains which are almost polymeric in nature.

It is interesting to note that there are apparently no mixed polymers in biological systems. Thus, one never finds a polymer chain involving both nucleotides and amino acids or sugar residues. If such mixed polymers existed, they might be unable to interact systematically through their backbones, and hence would lack an important element which gives rise to configurational stability.

POLYAMINO ACIDS

Present knowledge of the structure of the simpler polypeptides and proteins has its foundation in the extensive work of Pauling and Corey and their collaborators who worked out the crystal structures of a number of amino acids and small polypeptides. From their work, several important conclusions can be drawn.

(1) The dimensions of bond lengths and bond angles of the peptide group were obtained to an accuracy of ± 0.02 A and $\pm 2°$.[6-8] They found a similarity in the dimensions of the peptide group regardless of the particular amino acids involved in the linkage.[9] These are shown in Fig. 5. The carbon-oxygen distance in the carbonyl group was found to be 1.24 A, even though the sum of the double-bond covalent radii is 1.21 A. In a similar way, the carbon-nitrogen bond distance in the amide group is 1.32 A, which is less than the sum of the single-bond radii (1.47 A). This points to approximately 40% double-bond character in the C—N bond, and 60% double-bond character in the carboxyl group owing to resonance between the two.

(2) As a direct consequence of the double-bond character in the C—N bond, the amide group is expected to have a planar configuration. Experimental results amply confirm this expectation. It has been

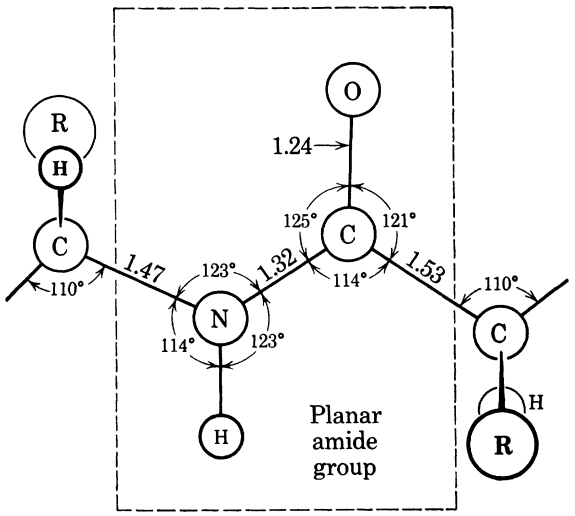


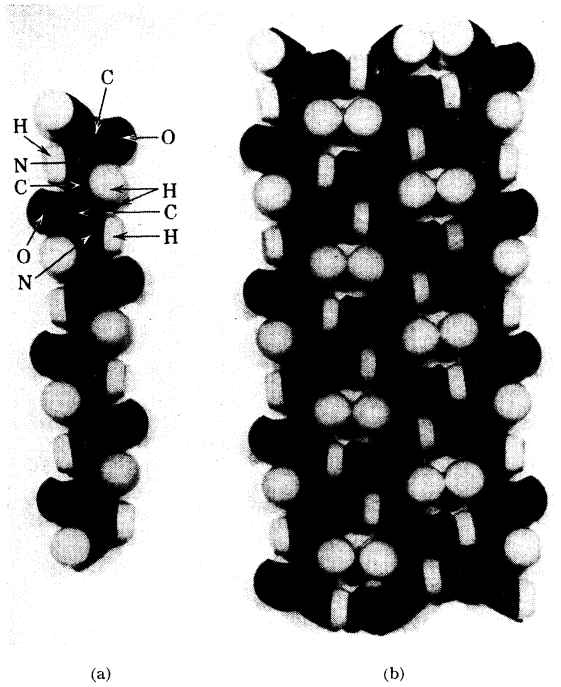
FIG. 5. Fundamental dimensions of polypeptide chains as derived from x-ray analyses of amino acids and simple peptides [from R. B. Corey and L. Pauling, Rend. ist. lombardo sci. **89**, 10 (1955)].

estimated that the distortion energy is about 1 kcal/mole for a deviation of 10° from planarity and about 3.5 kcal/mole for a 20° distortion.[6,7] In all of the peptide structures determined, the peptide group is found planar to within a few degrees.

(3) All of the small peptides now examined show a *trans*-configuration for the amide group. This is shown in Fig. 5 where the α-carbon atoms are attached to the planar amide group on opposite sides of the C—N bond. This conclusion from crystal-structure work is also supported by dipole moment studies.[10] Using infrared absorption, Badger and Rubaclava[11] have shown that the *trans*-configuration is more stable than the *cis*- by more than 2 kcal/mole.

(4) Another feature of the peptide group is the fact that it has a hydrogen atom attached to the nitrogen and is, therefore, capable of forming a hydrogen bond. All of the crystal structures analyzed were characterized by forming a maximum number of N—H···O hydrogen bonds. Most of the hydrogen bonds are linear to within a few degrees and have a length of 2.79 ± 0.12 A.[6,7]

Thus, the polypeptide backbone consists of a series of fairly rigid planar groups which may or may not be able to rotate relative to each other, depending upon the nature of the side groups attached to the α-carbon atom. This rotational barrier which may be as large as 3 kcal/mole is an additional important factor in determining the ultimate configuration of the molecules. Pauling and Corey first pointed out that a relatively small number of configurations will meet all of the four criteria described above. Thus, one can expect a fairly small number of stable polypeptide configurations. The following paragraphs review several of the more important ones.



(a) (b)

FIG. 6. A planar sheet composed of fully-extended polyglycine chains. The atoms are represented as solid out to their van der Waals radii. The hydrogen atoms attached to nitrogen have a depression in them for use in illustrating hydrogen bonding with carbonyl groups (>C=O). (a) Single extended polypeptide chain. (b) Antiparallel arrangement of polyglycine chains to form a sheet. Note that the hydrogen atoms in adjoining chains are in van der Waals contact. This would prevent the formation of this sheet from polypeptides other than polyglycine.

Extended Polypeptide Configurations

The simplest configuration of a polypeptide chain is one in which the polypeptide chain is extended so that all of the planar peptide units tend to lie in a common plane. In this extended form, alternate amino-acid residues have carboxyl groups pointing in the same direction, and adjacent residues are related by a translation of 3.6 A and a rotation of 180°. The usual extended form of polyglycine has a configuration which is close to that described in the foregoing. When fully extended, the polyglycine chain should have a repeat distance of 7.2 A. However, the fiber-axis repeat distance, obtained experimentally, is somewhat less[12]; hence, the chains are almost but not fully extended. An extended configuration is shown in Fig. 6, a photograph of atom models that are designed to show all of the van der Waals radii for the various atoms.[6] In Fig. 6, a single extended chain of polyglycine is shown also. This structure is helical in the same sense that polyethylene [Fig. 1(a)] is helical, with a rotation angle of 180°. Chains of polyglycine can be hydrogen bonded to make a sheet structure, as shown in Fig. 6(b), where the chains alternate in direction. Such a configuration in which the successive planar peptide residues are almost in the same plane could exist in polyglycine only and not in any other polypeptide, because steric interference would result in the presence of a bulkier side chain. This can be seen most clearly by noting that the hydrogen atoms in the sheet structure are in contact. Thus, as illustrated in Fig. 6, it would be impossible to have a larger β-carbon atom attached in the position of the hydrogen atoms.

It is possible, however, to make a sheet structure utilizing hydrogen bonds in the plane of the sheet by rotating the peptide units until the β-carbon atoms are no longer in contact as they are in the polyglycine configuration of Fig. 6. This configuration is important in the structure of silk fibroin.

Structure of Silk Fibroin

There is a rotational barrier between adjacent planar peptide units in the polypeptide chain, and, as a consequence, only a limited number of configurations are stable. One of the earliest to be investigated critically by Pauling and Corey was a semi-extended form in which adjacent planar peptide groups were related to each other by a translation of approximately 3.5 A and a rotation of 180°. A drawing of the configuration, looking down the plane of the peptide units, is illustrated in Fig. 7(a). In this configuration, the β-carbon atoms emerge from alternate sides of the polypeptide backbone, and the backbone itself has a zigzag rather than a fully extended form. The most significant feature of this configuration is the orientation of the C=O and N−H groups which line in one plane and are perpendicular to Fig. 7(a). These groups are oriented so that all hydrogen bonds can be formed by aligning a series of these polypeptide chains side by side. This has been done in Fig. 7(b) to form what is called the "antiparallel chain pleated sheet" structure. Alternate chains run in opposite directions, and in this way the successive amino and carboxyl groups are hydrogen-bonded to chains on either side of the original chain. By working carefully with molecular models, it is possible to show that the identity period along the fiber axis is 7.00 A, whereas the lateral distance between equivalent chains (alternate chains) is 9.50 A.

There are several interesting features in the antiparallel chain pleated sheet structure. The β-carbon atoms protrude at right angles from the plane of the sheet on alternate sides for successive residues along one chain. Furthermore, the β-carbon atoms are in phase on either side of the sheet; that is, they all appear at the same points along the fiber axis. Thus, there are large grooves between adjacent rows of β-carbon atoms.

The common form of silk fibroin is obtained from the silkworm *Bombyx mori*, and it has been the subject of x-ray diffraction investigations for many years. Most recently, Marsh *et al.*[13] have carried out an extensive and careful investigation and have arrived at a structure for this form of silk fibroin based on the antiparallel chain pleated sheet structure. Although the diffraction pattern is not simple, most of the reflections can be

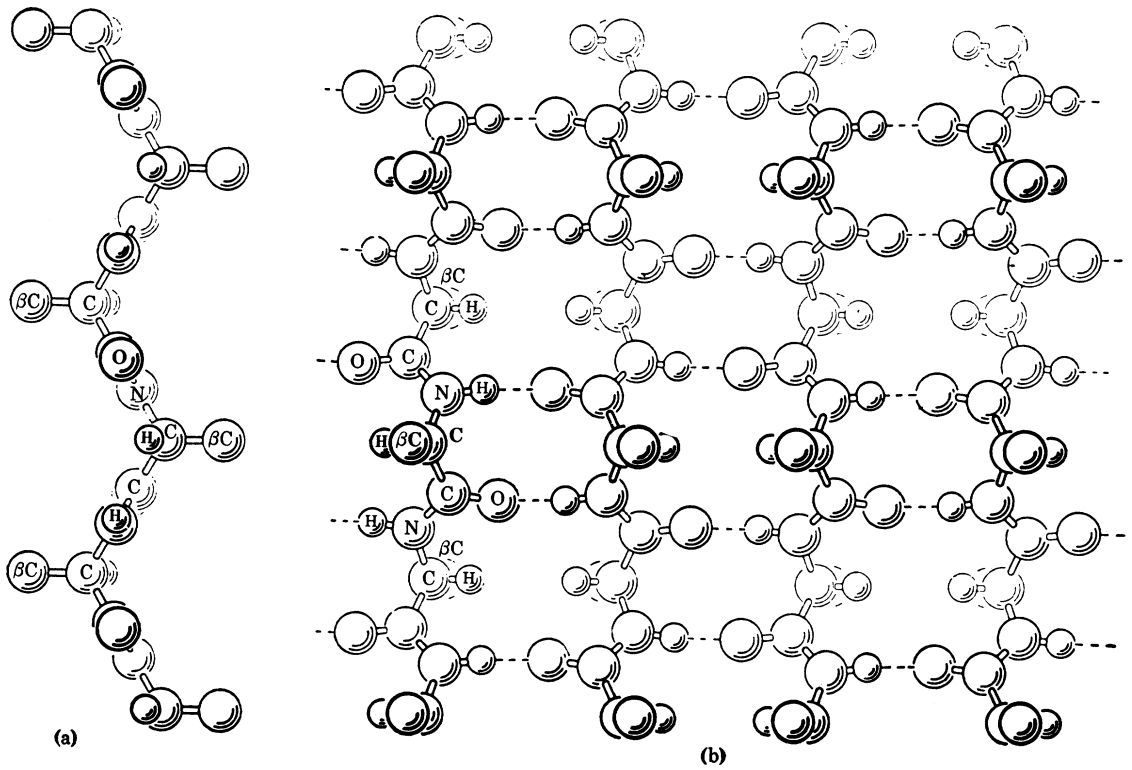


FIG. 7. (a) A drawing of a single polypeptide chain in the configuration found in the antiparallel chain pleated sheet. The chain is viewed in the direction of the hydrogen bonds and, therefore, along the plane of the sheet. Note the zigzag arrangement produced by alternate planar peptide residues. (b) A drawing of the antiparallel chain pleated sheet. In a given polypeptide chain, alternate β-carbon atoms are found on the same side of the pleated sheet [from R. E. Marsh, R. B. Corey, and L. Pauling, Biochim. et Biophys. Acta **16**, 1 (1955)].

accounted for by a simple unit cell which is orthogonal with dimensions 9.20 and 9.40 A at right angles to the fiber and a fiber-axis repeat of 6.97 A. The similarity between two of these dimensions observed experimentally and the pleated-sheet dimensions (7.00 A along the polypeptide chain and 9.50 A between equivalent alternate chains) suggested to these investigators that the antiparallel chain pleated sheet structure might serve as the structural basis of silk fibroin. Experiments carried out with rolled and flattened samples of silk fibroin produce the double orientation usually associated with layered structures. In this way, they were able to show that the 9.20-A reflection represented the repeat distance between molecular layers in silk fibroin. To deduce the structure, they had to arrange the layers of the sheet structure such that an identity repeat of 9.2 A was obtained.

One of the most remarkable features of silk fibroin is its unusual chemical composition. The simplest of the amino acids, glycine, constitutes almost one-half (44%) of the total number of amino acids present, while the next simplest amino acids (alanine and serine) comprise almost 40% of the residues. This suggested to the investigators that a simplified model of the structure might contain glycine residues with the alanine and serine residues in between at alternate sites along the polypeptide chain. In this way, all of the side groups protruding from one side of the sheet are from glycine residues and, therefore, consist simply of hydrogen atoms. The other alternate sites are occupied by other side chains—mainly alanine which has one methyl group or serine which has CH_2OH as a side chain. When the

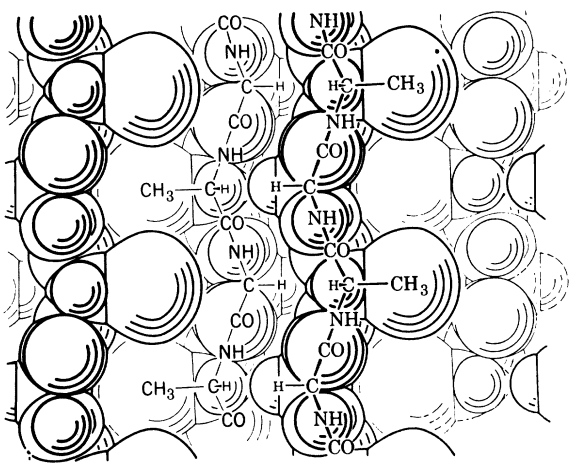


FIG. 8. A packing drawing of the structure of silk fibroin (*Bombyx mori*) viewed parallel to the pleated sheet. The fiber axis is vertical in this view [from R. E. Marsh, R. B. Corey, and L. Pauling, Biochim. et Biophys. Acta **16**, 1 (1955)].

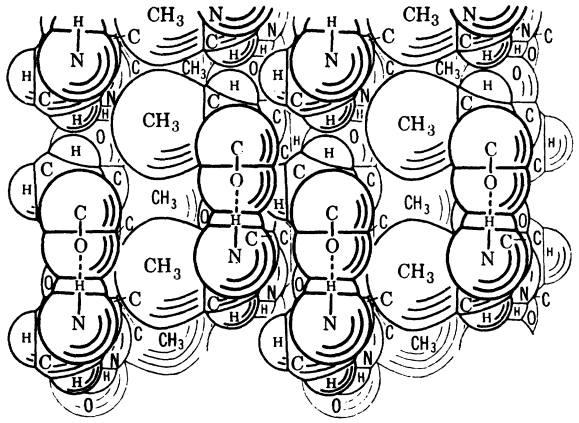


FIG. 9. A packing drawing of the structure of silk fibroin (*Bombyx mori*) viewed along the fiber axis. The hydrogen-bonding sheets are vertical in this view [from R. E. Marsh, R. B. Corey, and L. Pauling, Biochim. et Biophys. Acta **16**, 1 (1955)].

side chains are distributed in this manner, it is possible to pack pairs of sheets together so that the bulky side chains are located between the two sheets. In this way, the β-carbon atoms of one sheet protrude into the space between the β-carbon atoms from the other sheet. The distance from one sheet to the next is 5.7 A when they are packed in this way. However, when these double-sheet structures are stacked together, the glycine side chains are in contact and the distance from sheet to sheet over the glycine contacts is 3.5 A. Hence, the identity repeat for such a head-to-head, tail-to-tail stacking of sheets is $5.7+3.5=9.2$ A. This is, of course, exactly what is observed in the simple unit cell derived from the x-ray diffraction study of silk fibroin.

A packing drawing of the structure of silk fibroin is shown in Figs. 8 and 9. In Fig. 8, the view is perpendicular to the fiber axis and parallel to the pleated sheets. This view shows how the bulky methyl side chains are packed together with adjacent sheets filling alternate positions along the fiber axis, and how the small hydrogen side chains of the glycine residues fit together. Figure 9 is a view down the fiber axis. In this view, the hydrogen-bonding between the chains can be seen in the vertical direction, and the alternate packing arrangement of side chains is also clearly shown.

The reflections calculated from a simplified structural model of this type agree very well with the intensities obtained from the x-ray diffraction study of silk fibroin. This model does not account for all of the amino acids found in silk, including many with bulky side chains. They act to perturb this simplified model and produce some of the minor complications present in the diffraction pattern. Most of the properties of silk, however, can be understood from the simplified unit of structure.

Coiled Helical Configurations

The extended form of polyglycine shown in Fig. 6 is called polyglycine I and was the first form studied crystallographically by x-ray methods. In 1955, Bamford and his colleagues[14] described a second form of polyglycine which they called polyglycine II. The struc-

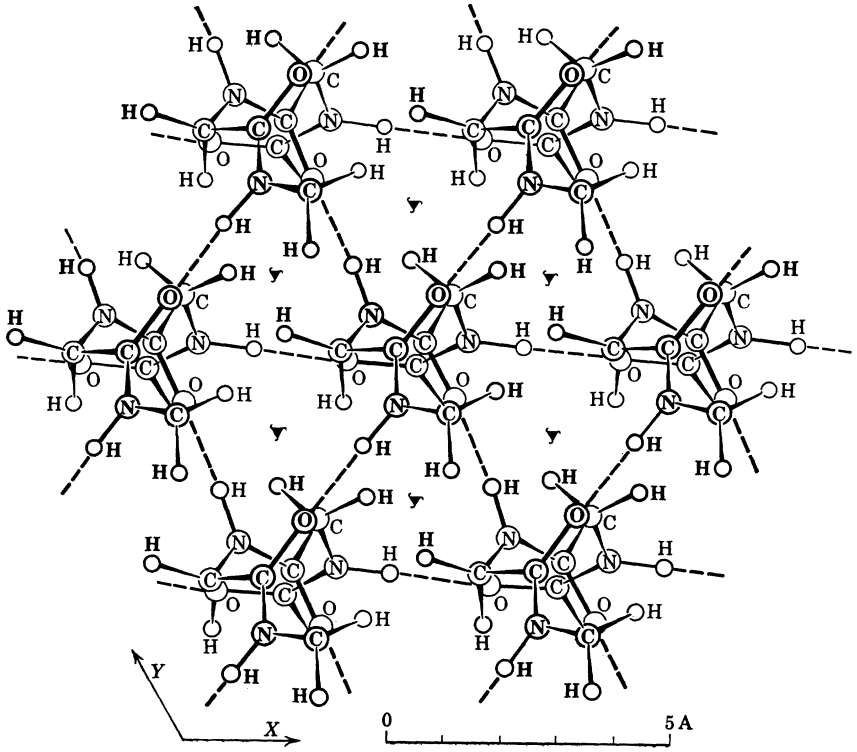


FIG. 10. The crystal structure of polyglycine II as viewed along the axis of the helical chains. The dashed lines represent hydrogen bonds [from F. H. C. Crick and A. Rich, Nature **176**, 780 (1955)].

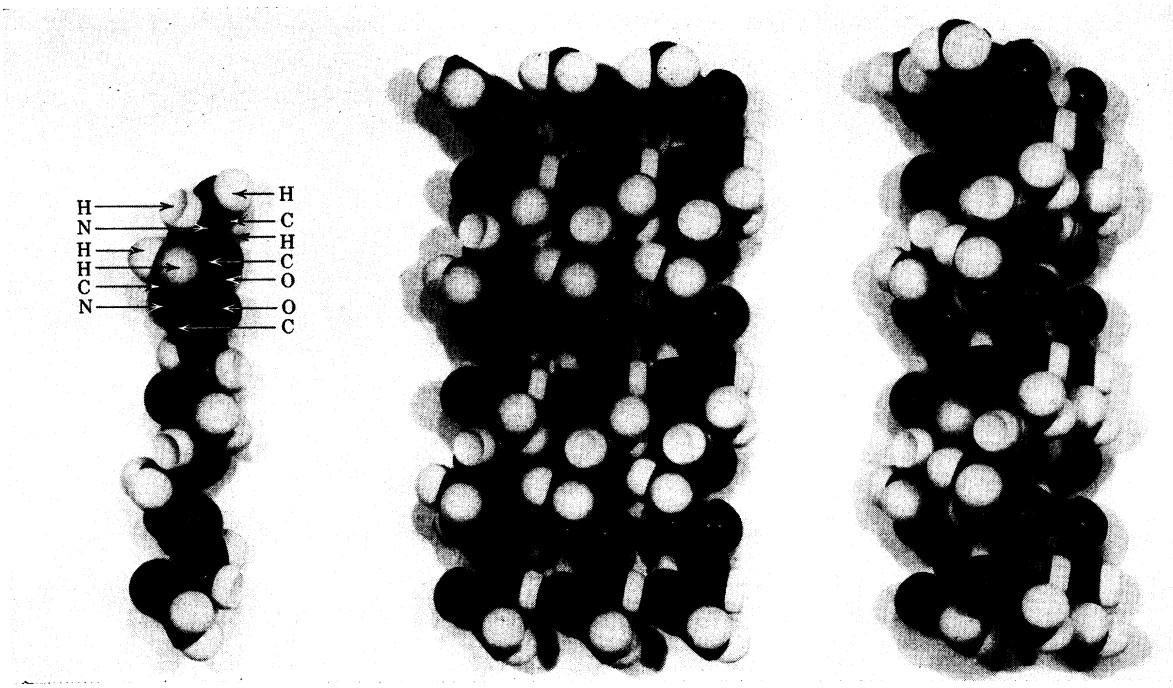


FIG. 11. (a) Molecular model of a single polyglycine chain arranged in the helical form found in polyglycine II. There are three peptide residues per turn in this form. (b) Three chains hydrogen bonded together to make a planar part of the polyglycine-II lattice. (c) Three chains packed in triangular form as found in the polyglycine-II lattice. The hydrogen bonds connecting these three chains are partly obscured.

ture of polyglycine II has been worked out by Crick and Rich,[15] and it is a particularly simple one. In it, all of the polypeptide chains are parallel, and each one is organized into a threefold screw axis so that adjacent residues are related by a translation of 3.1 A and rotation of 120°. The chains are packed in an hexagonal array, each chain being hydrogen bonded to six neighbors. The hydrogen bonds lie roughly perpendicular to the screw axis, and run in several directions and not solely in one plane as in the polyglycine sheet structure described in the foregoing. Figure 10 is a view of this structure down the fiber axis. In it, seven polypeptide chains are hexagonally packed with dashed lines indicating the linear hydrogen bonds which hold the adjoining chains together. The hydrogen-bond distance is 2.76 A, and the spacing between the polypeptide chains is 4.8 A. Since the translation along the axis is 3.1 A per residue, the axial repeat is 9.3 A because of the threefold screw symmetry. In this configuration, the polypeptide chains can fulfill all of the criteria developed by Pauling for stable configurations of polypeptide chains. It should be noted, however, that the polyglycine-II lattice could be formed only by a polymer of glycine. If bulkier side chains were present, the molecules would not be able to hydrogen bond, since the chains are already densely packed with only hydrogen atoms attached to the α-carbon atoms. A side view of the polyglycine-II chain structure is seen in Fig. 11, using molecular models which show the atomic van der Waals distances. On the left [Fig. 11(a)] is a single polyglycine chain arranged in a threefold screw axis. In the center [Fig. 11(b)], three of these chains are hydrogen-bonded to make one plane through the polyglycine-II lattice. It can be seen that only one-third of the hydrogen bonds is utilized in forming a sheet. The remaining two-thirds of the hydrogen bonds are used in holding the sheets together. Thus, in order to form all of the hydrogen bonds, a three-dimensional network must be built up in polyglycine II in contrast to the two-dimensional sheet structure seen in polyglycine I. In Fig. 11(c), the same three polyglycine chains are present as are in Fig. 11(b), but they are arranged so that they are no longer planar. One polyglycine chain is lying on top of the other two so that the three are hydrogen bonded. In this form, they represent a group of three polyglycine chains obtained by selecting a triangle of three close-packed chains (see Fig. 10). It should be pointed out that there are two ways of selecting a group of these three polyglycine chains. This can be seen most readily by observing the directions of the hydrogen bonds (N—H···O) which hold together the group of three. In one group, they are clockwise and, in the other, they are counterclockwise. This difference in the two groups is a consequence of the space-group symmetry, since the chains are arranged about a threefold screw axis and not about a sixfold screw axis.

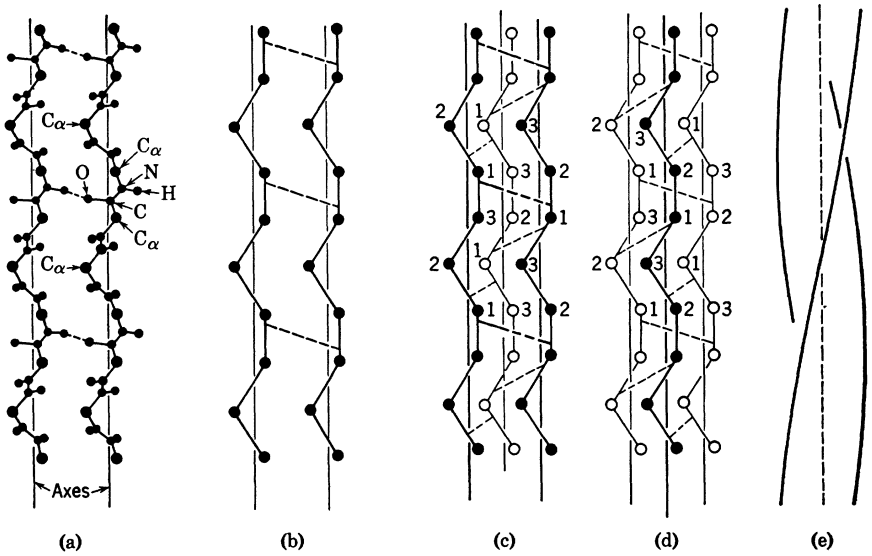


FIG. 12. A schematic drawing which illustrates how the collagen molecule may be generated from the polyglycine-II lattice. (a) Two molecular chains from polyglycine II. (b) The same chains showing only the α-carbon atoms (solid circles) and hydrogen bonds (dashed lines). (c) A third chain is added behind the two chains of (b). This generates collagen I. (d) The third chain is added in front of the two chains found in (b). This generates collagen II. (e) In the collagen molecule itself, the three molecular axes are coiled as shown in the diagram. The numbers on (c) and (d) correspond to the three kinds of sites found on the collagen molecule. Site number 1 is usually occupied by glycine [from A. Rich and F. H. C. Crick in *Recent Advances in Gelatine and Glue Research* (Pergamon Press, London, 1957), p. 1].

Structure of Collagen

Just as the structure of the antiparallel pleated sheet serves as a model for deriving the structure of silk fibroin, the polyglycine-II lattice can be used to derive the structure of collagen. Collagen, found in the form of elongate fibrils, is the major structural element of skin and connective tissue in the animal kingdom. One of its most unusual features is the chemical composition. One-third of its residues is glycine, and it has, in addition, a large number of pyrrolidine rings. Proline plus hydroxyproline make up about 24% of the residues. Because of the high glycine content, it is perhaps not surprising that there is a relationship between the polyglycine-II lattice and the collagen molecule. This relationship can be seen most clearly in Fig. 12, which shows a schematic development starting from the polyglycine-II lattice and ending in the collagen molecule.[16] In Fig. 12(a), two parallel polyglycine chains are sketched with the hydrogen bonds joining them (dotted lines). In Fig. 12(b), only the α-carbon atoms of the glycine residues are shown, with short solid lines schematically indicating the peptide units connecting them. It is seen that these polypeptide chains coil around the two axes and are held together by the hydrogen bonds (dashed lines). The collagen molecule may be derived from three polyglycine chains such as is shown in Figs. 12(c) and 12(d). In Fig. 12(c), a third polypeptide chain is added behind the two shown in Fig. 12(b), whereas in Fig. 12(d), the third chain is added in front of the original two. In this way, two groups of three polypeptide chains are created which are prototypes of two models, collagen I and collagen II. These correspond to the two different ways of collecting groups of three molecules as shown in the end view of the polyglycine-II lattice (Fig. 10). In the collagen molecule, however, the axes around which the polypeptide chains coil are no longer straight as in Figs. 10(a)–10(d), but are coiled slightly as shown in Fig. 12(e). Thus, the collagen molecule is regarded as a coiled-coil structure.

In silk fibroin, the spacings between the sheets in the molecule are uneven; the closer spacing (3.5 A) arises because there are only the side chains of glycine between the sheets. Similarly, in the collagen molecule, the three polypeptide chains can come close together, because the glycine residues are located near the center of the group of three chains, whereas the other two-thirds of the residues are projecting outward from the center of the molecule. Thus, two-thirds of the residues can have any side chains on them. And, what is more significant, they can accommodate pyrrolidine rings without producing any steric difficulties.

It is generally agreed that the collagen-II model is the most likely form for the collagen molecule,[17–19] and in it pyrrolidine rings can be found on two of the externally situated sites, while the third internal site must have a glycine residue. This explains why all collagens are found with 33% glycine in their composition. Figure 13 shows a form of the collagen-II molecule where two-thirds of the sites are filled with pyrrolidine rings such that a repeating sequence is used in each chain. In it, the sequence glycine-proline-hydroxyproline is used, since this is a common constituent of collagen

digests. This model of collagen was made using the polypeptide chains of Fig. 11(c), except that the axes of the polyglycine chains are now coiled rather than straight. As seen in Fig. 13, the collagen-molecule model is elongated, and has a distinct helical groove running down it which arises from the absence of a bulky side chain on the glycine site. The helical ridge on the surface of the molecule arises because the bulky side chains, in this case pyrrolidine rings, all lie closely packed. In the collagen molecule, however, there is a variety of amino acids found in that ridge, including many which have polar side chains. These are undoubtedly important in stabilizing collagen by hydrogen bonding, as well as in forming electrostatic bonds from one molecule to its nearest neighbor.

In the polyglycine-II lattice, the fundamental screw operation used in generating the lattice is a translation of 3.1 A and a rotation of 120°. Because collagen is a coiled-coil molecule formed from this lattice, the fundamental screw operation in generating the collagen helix is a translation of 2.86 A and a rotation of 108°. Of course, the asymmetric unit in the collagen molecule now consists of a group of three amino acids in contrast to the single amino acid in the polyglycine chain. In collagen, the coiling of the three polypeptide chains about one another is produced by the steric interaction of the β-carbon atoms which are present in the collagen molecule but which are absent in the polyglycine-II lattice.

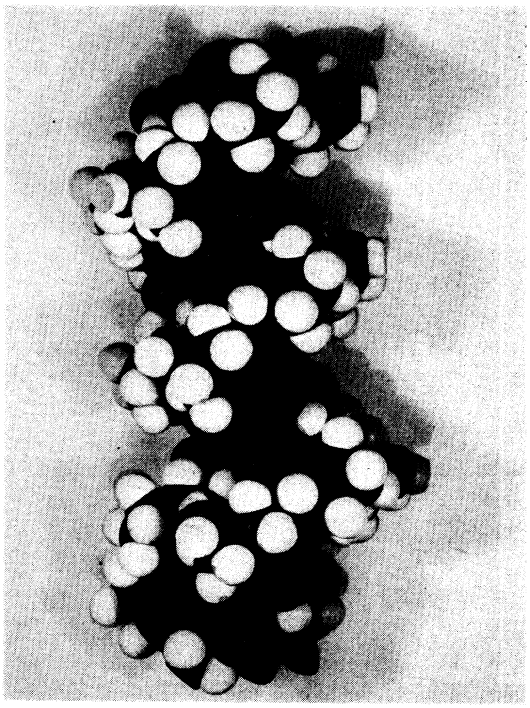


FIG. 13. A model of collagen II. The three polypeptide chains of Fig. 11(c) have been modified by the addition of proline and hydroxyproline residues to sites 2 and 3. This causes the three chains to coil about as shown in Fig. 12(e).

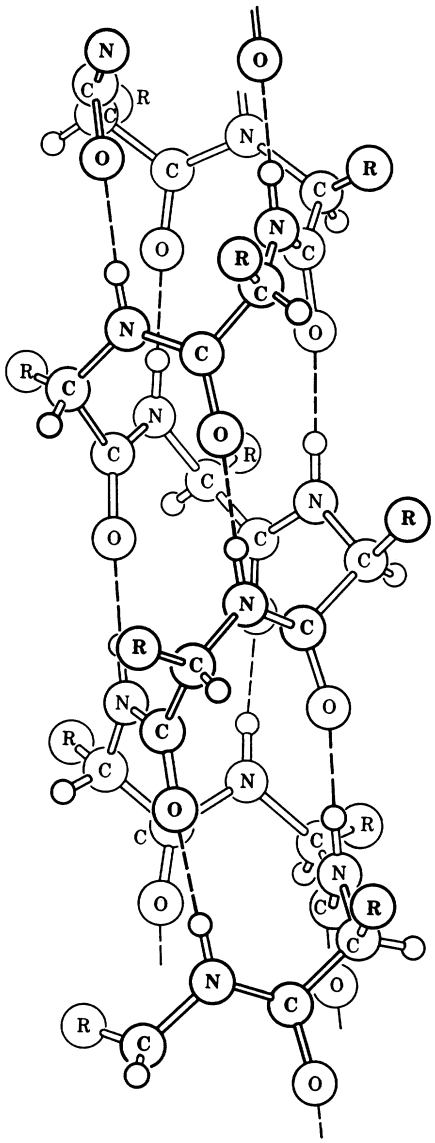


FIG. 14. A drawing of the α-helix [from R. B. Corey and L. Pauling, Rend. ist. lombardo sci. **89**, 10 (1955).

The α-Helix

An important feature of the structure of silk fibroin is the fact that alternate residues are glycine units along the polypeptide chain. This permits the sheets to pack very closely. Similarly, in the collagen molecule, an important feature is the fact that every third residue is glycine, thereby allowing the three chains to come close together so that they can hydrogen bond. In a sense, both of these structures are very specialized and are a consequence of a restricted amino-acid composition. The best example of a polypeptide configuration which can be formed with minimal or no compositional restrictions is the α-helix.[20] In the α-helix (Fig. 14), the peptide units are oriented so that their planes are tangent to a cylinder

around the axis of the helix. In this way, hydrogen bonds can be formed which are parallel to the axis, and thus serve to hold together the various turns of the helix. The fundamental operations for generating the α-helix are a rotation of 100° around the axis and a translation of 1.5 A along the axis. Each amide group is connected by a hydrogen bond to the third amide group from it along the peptide chain. There are 3.6 residues per turn of the helix, and the total rise of the helix per turn is 5.4 A. The helix formed in this way is packed very firmly, and there is no open space in the center.

Of all of the configurations discussed hereto, the α-helix is by far the most important. It is a completely general structure in that it will accommodate any amino-acid side chain. When proline residues are incorporated into the structure, the helix axis changes direction, but this incorporation probably does not destroy the organization of the helix. This structure has been found in a variety of synthetic polypeptides,[21,22] and there is evidence for its existence in many globular proteins as well as hair, muscle, wool, and myosin.[23,24] The α-helix should produce a reflection at a spacing of 1.5 A, which is the translation distance along the fiber axis for each amino-acid residue. Perutz[23] has found this reflection in a variety of proteins, and this has provided strong inferential evidence for the α-helix.

It has been suggested that the α-keratin proteins such as hair contain helices twisted around one another to form coiled-coil structures.[25,26] Thus, these molecules may form by hexagonal packing into cable-like structures involving several strands which coil around each other. Structures of this sort give reasonable agreement with the observed x-ray diffraction photograph obtained from hair.

The importance of the α-helix lies in its extreme generality. Knowledge of the detailed configuration of globular proteins, as discussed by Kendrew (p. 94), is just beginning to be advanced. It seems quite likely that the α-helix will form an important feature of many of these molecules. Significantly, the α-helix buries the systematically repeating part of the polypeptide chain, using it to form a supporting fabric of great stability. In so doing, it exposes to the external environment all of the side chains with their great diversity. Thus, the α-helix enables a protein structure to maximize its variability and, thereby, produce the considerable chemical versatility of the proteins.

SUMMARY

On considering the molecular structure of biological polymers, the proteins and the nucleic acids, one can see that they are more complex than are synthetic polymers. Nonetheless, there are many common structural features in all of these materials. The helix is seen as the most general form for organizing successive residues along a polymer chain. The helix is found in many forms, ranging from fully extended to tightly coiled molecules. In all of these structures, space is filled efficiently, thereby maximizing the amount of van der Waals stabilization. Most synthetic polymers achieve a limited degree of structural complexity, as for example, coiling into a helical form. Materials of biological origin, by contrast, often go beyond this in complexity. They may form coiled-coils, as in collagen or in the α-keratin structure. On even a higher level, these polymers may fold in a very complicated fashion to produce globular molecules in which the interactions of the different side chains play a critical part in maintaining configurational stability.

BIBLIOGRAPHY

[1] W. Cochran, F. H. C. Crick, and V. Vand, Acta Cryst. **5**, 581 (1952).

[2] D. R. Davies and A. Rich, Acta Cryst. (to be published).

[3] C. W. Bunn and E. R. Howells, Nature **174**, 549 (1954).

[4] R. P. Daubeny, C. W. Bunn, and C. J. Brown, Proc. Roy. Soc. (London) **A226**, 531 (1954).

[5] C. W. Bunn and E. V. Garner, Proc. Roy. Soc. (London) **A189**, 39 (1947).

[6] R. B. Corey and L. Pauling, Proc. Roy. Soc. (London) **B141**, 10 (1953).

[7] R. B. Corey and L. Pauling, Rev. Sci. Instr. **24**, 621 (1953).

[8] L. Pauling and R. B. Corey, Fortschr. Chem. org. Naturstoffe **8**, 310 (1954).

[9] R. B. Corey and L. Pauling, Rend. ist. lombardo sci. **89**, 10 (1955).

[10] S. Mizushima, T. Simanouti, S. Nagakura, K. Kuratani, M. Tsuboi, H. Baba, and O. Fujioka, J. Am. Chem. Soc. **72**, 3490 (1950).

[11] R. M. Badger and H. Rubalcava, Proc. Natl. Acad. Sci. U. S. **40**, 12 (1954).

[12] W. T. Astbury, Nature **163**, 722 (1949).

[13] R. E. Marsh, R. B. Corey, and L. Pauling, Biochim. et Biophys. Acta **16**, 1 (1955).

[14] C. H. Bamford, L. Brown, E. M. Cant, A. Elliott, W. E. Hanby, and B. R. Malcolm, Nature **176**, 396 (1955).

[15] F. H. C. Crick and A. Rich, Nature **176**, 780 (1955).

[16] F. H. C. Crick and A. Rich in *Recent Advances in Gelatine and Glue Research* (Pergamon Press, London, 1957), p. 20.

[17] A. Rich and F. H. C. Crick, Nature **176**, 915 (1955).

[18] P. M. Cowan, S. McGavin, and A. C. T. North, Nature **176**, 1062 (1955).

[19] G. N. Ramachandran, Nature **177**, 710 (1956).

[20] L. Pauling, R. B. Corey, and H. R. Branson, Proc. Natl. Acad. Sci. U. S. **37**, 205 (1951).

[21] L. Pauling and R. B. Corey, Proc. Natl. Acad. Sci. U. S. **37**, 241 (1951).

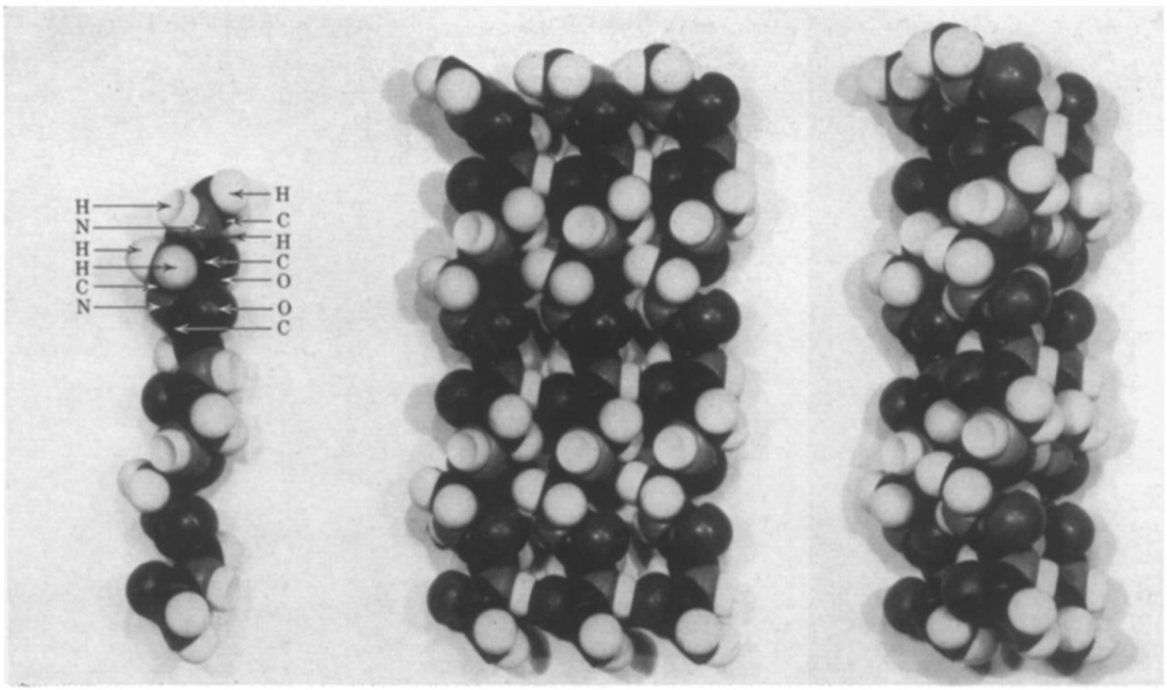
[22] C. H. Bamford, L. Brown, A. Elliott, W. E. Hanby, and I. F. Trotter, Proc. Roy. Soc. (London) **B141**, 49 (1953).

[23] M. F. Perutz, Nature **167**, 1053 (1951).

[24] W. T. Astbury, Proc. Roy. Soc. (London) **B141**, 1 (1953).

[25] F. H. C. Crick, Nature **170**, 882 (1952).

[26] L. Pauling and R. B. Corey, Nature **171**, 59 (1953).



(a) (b) (c)

FIG. 11. (a) Molecular model of a single polyglycine chain arranged in the helical form found in polyglycine II. There are three peptide residues per turn in this form. (b) Three chains hydrogen bonded together to make a planar part of the polyglycine-II lattice. (c) Three chains packed in triangular form as found in the polyglycine-II lattice. The hydrogen bonds connecting these three chains are partly obscured.

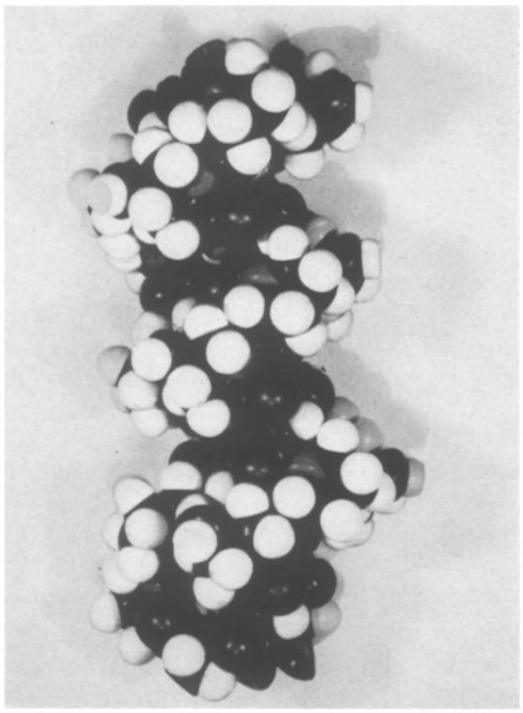
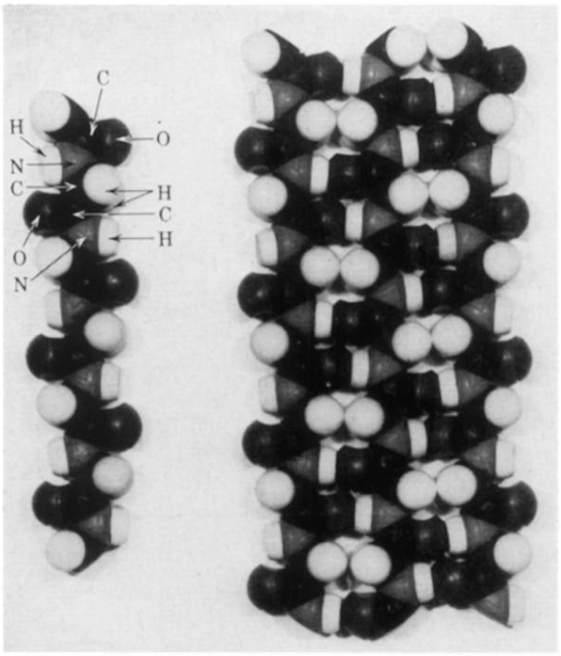


FIG. 13. A model of collagen II. The three polypeptide chains of Fig. 11(c) have been modified by the addition of proline and hydroxyproline residues to sites 2 and 3. This causes the three chains to coil about as shown in Fig. 12(e).



(a) (b)

FIG. 6. A planar sheet composed of fully-extended polyglycine chains. The atoms are represented as solid out to their van der Waals radii. The hydrogen atoms attached to nitrogen have a depression in them for use in illustrating hydrogen bonding with carbonyl groups (>C=O). (a) Single extended polypeptide chain. (b) Antiparallel arrangement of polyglycine chains to form a sheet. Note that the hydrogen atoms in adjoining chains are in van der Waals contact. This would prevent the formation of this sheet from polypeptides other than polyglycine.