THE STRUCTURE OF FIBROUS PROTEINS[1,2]

MAURICE L. HUGGINS

Kodak Research Laboratories, Rochester, New York

Received October 23, 1942

Following a presentation of some of the more important general principles which can reasonably be assumed to apply to protein structures, models have been proposed and compared with x-ray and analytical data for silk fibroin (*Bombyx mori*), β-keratin, α-keratin, and collagen.

The problem of the determination of the arrangements of atoms in proteins is an exceedingly difficult one. Even if we bring to bear the very powerful techniques now available, such as those involving the use of x-rays, the electron microscope, and the ultracentrifuge, there seems to be little hope of solving the problem in the near future by *deductive* methods alone. The most promising method of attack at this stage seems to require the making of tentative assumptions regarding the structures of protein molecules, followed by experimental testing of these assumptions. To avoid waste effort we should of course be guided in making these assumptions by such knowledge as is already available regarding related structure and by such limitations as are imposed by deduction from experimental data on the proteins themselves.

This is the point of view which will be taken in this paper. Starting with certain assumptions which appear reasonable on the basis of analogy or for which there is direct experimental evidence, we shall speculate regarding possible types of structure for proteins and then, where possible, test our speculative pictures by means of available applicable data.

For the present, we shall limit our discussion to *fibrous* proteins.

ASSUMPTIONS AND GENERAL PRINCIPLES

We shall assume, following Emil Fischer, that fibrous proteins are built up, at least primarily, of long polypeptide chains or very large rings. We shall assume these chains to be composed of the "residues" of amino acids, at least approximately in the relative proportions determined by accepted analytical methods. We shall assume, as is customary, that the orientation of the bonds around each asymmetric α-carbon atom is uniformly *levo*—that is, the same as in levolactic acid.

[1] Communication No. 889 from the Kodak Research Laboratories, Rochester, New York.

[2] Presented (except for minor changes) under the title "Hydrogen Bonds in Proteins" at the Symposium on "The Hydrogen Bond and Related Topics" at the Memphis Meeting of the American Chemical Society, April 21, 1942. An outline of the material herein was included in the author's recent review of "X-ray Studies of the Structure of Compounds of Biochemical Interest" in the 1942 Annual Review of Biochemistry (36). The structure models described in this paper have also been briefly discussed at previous meetings of the American Chemical Society (September, 1937; April, 1939; April, 1940) and at the Gibson Island Conference on X-ray and Electron Diffraction, July, 1940.

We shall make the reasonable assumption that the distances between closest atoms and the angles between adjacent bonds are in proteins approximately the same as in comparable small molecules of known structure (table 1). We shall also assume that the attraction between CO and NH groups which has been found to produce intermolecular NHO bridges in all the comparable small molecule crystals of known structure is also effective in producing similar (intermolecular and intramolecular) bridges in proteins. Because of the resulting

TABLE 1*

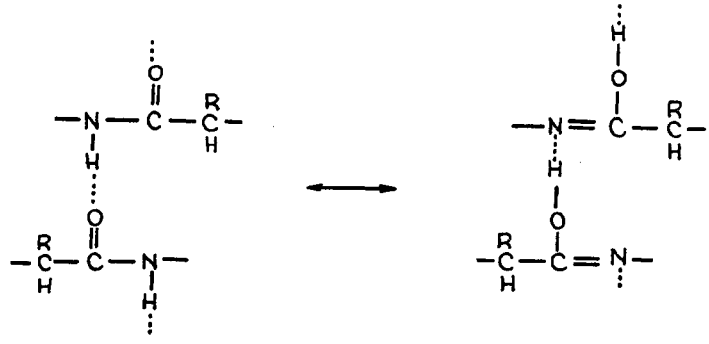
Interatomic distances and interbond angles

Experimental values for three simple compounds and estimated values for proteins

	DIKETO-PIPERAZINE (25)	GLYCINE (4)	ALANINE (41)	PROTEINS
	Å.	Å.	Å.	Å.
N—C	1.41	1.39	1.42	1.41
C—C′	1.47	1.52	1.54	1.52
C′—O	1.25	1.25, 1.27	1.23, 1.25	1.25
C′—N	1.33			1.33
N—H···O	2.85	2.76, 2.88	2.78, 2.84, 2.88	2.85
	degrees	*degrees*	*degrees*	*degrees*
∠NCC′	120	112	112, 113	112
∠CC′O	120	119	118	118
∠C′CR			110	110
∠OC′O		122	124	
∠CC′N	120			118
∠C′NC	120			118

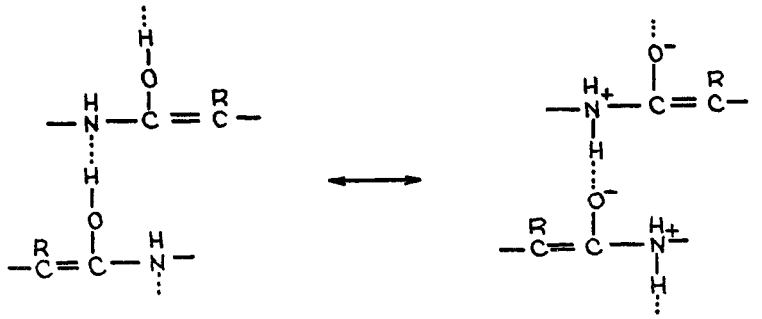
*C in this table refers to a carbon atom of a CHR or CH_2 group; C′ refers to a carbon atom of a carbonyl group.

increased stability, we expect these bridges to be linked together in long chains or in rings capable both of resonance of the following sort



and of synchronized oscillations of the hydrogens and the mobile electron systems (23). That NHO bridges actually are present in quantity in proteins has been confirmed by infrared spectrum studies (15, 23). These studies show that OHO bridges, if present at all, are present only in very small amounts.

Since most of the amino acid residues in proteins contain a hydrogen atom attached to the α-carbon atom, and since it might be expected that such a hydrogen would be capable of shifting to the neighboring oxygen, we should perhaps consider resonating structures of this sort:



For the present, however, we shall neglect such a possibility.

The hydrogen atoms of the CHR groups may also form bridges to carbonyl oxygen atoms, since there is good evidence for C—H···O bridges in comparable structures (24, 28, 30, 32, 34, 54, 60). This possibility will be discussed below in connection with the structure of collagen.

A protein which gives an x-ray diffraction pattern showing lines or spots instead of, or in addition to, the broad ill-defined bands characteristic of liquids and truly amorphous solids, must possess, over large regions at least, a regular crystalline type of structure. Polypeptide chains extending through the crystalline regions must each have a screw axis of symmetry, or else two or more chains must be grouped around screw axes or other symmetry elements. The unbalanced forces on opposite sides of a chain which has no screw axis—e.g., any of the earlier chain structures advocated for α-keratin by Astbury (13, 14, 29) or the one that he has most recently proposed (7, 10) for collagen—would tend to bend it continuously in the same direction. Fibrous proteins giving crystalline x-ray patterns can have layer structures composed of unsymmetrical chains only if the chains in each layer are alternately oriented in opposite directions, in such a way as to give a symmetry axis or symmetry plane or a set of symmetry centers between adjacent chains.

A principle which is logically reasonable and which has been amply verified by structure analyses of a great many substances is that like atoms or atomic groups tend to be surrounded in a like manner. Because of the variety of R groups present in any given protein, some differences between the environments of corresponding groups must be expected, but these differences should be minor ones. In general, a structural pattern for a protein in which like groups are all surrounded in a like manner, except for differences between the R groups, is more probable than one in which this is not the case. This is an argument against all of the structures advocated by Astbury for α-keratin (8, 11, 13, 14, 29) and against his latest collagen model (7, 10).

Another important structure principle is that of *close-packing*. In addition to the attractive forces connected with covalent-bond and hydrogen-bridge

formation and the Coulomb and polarization forces acting when one (or both) of two atoms or groups of atoms possesses a net electrical charge, two atoms or groups attract each other by what we may call a "van der Waals attraction" and repel each other by an "interpenetration repulsion," having a magnitude depending on the amount of interpenetration of the "electron clouds" of the two atoms or groups (figure 1). Both of these forces increase in magnitude as the distance decreases, but the interpenetration repulsion increases more rapidly. At distances greater than a certain minimum, characteristic of the kinds of atoms or groups, the net force is an attraction; at smaller distances, it is a repulsion.

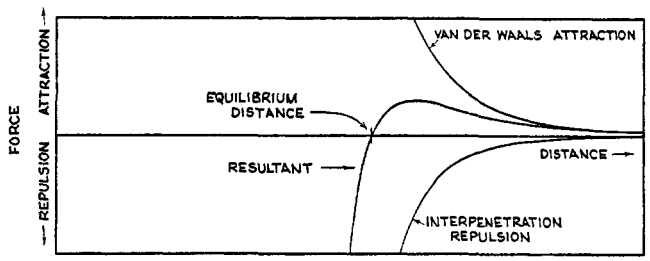


FIG. 1. Illustrating the variation with the distance between two atoms of their mutual van der Waals attraction and interpenetration repulsion forces and of the resultant force.

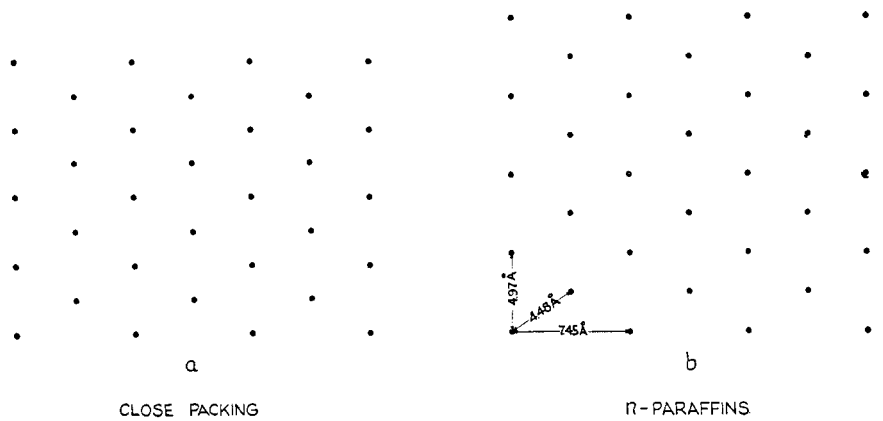


FIG. 2. (a) Close-packing of spheres having their centers in a plane; (b) distribution of chain axes in a normal paraffin crystal.

From these considerations it follows that, other things being equal, the most stable arrangement for an assemblage of molecules is one in which the component atoms and groups are packed together so that (*a*) the distances between neighbors are close to the equilibrium distance, (*b*) each atom or group has as many close neighbors as possible, and (*c*) there are no large unoccupied regions. In other words, each structure tends to be as "close-packed" as possible, consistent with the "sizes" of its component atoms or groups. Like spherical atoms or groups in a single plane thus assume a hexagonal arrangement (figure 2a) in which each has the maximum number (six) of close neighbors, all at the same distance. With non-spherical atoms or groups a similar arrangement, but dis-

torted, is produced. In a crystal of a paraffin or of one of its simple derivatives, the chain axes line up parallel to each other in the same sort of hexagonal array (but usually distorted). In normal long-chain paraffins (37, 49), for example, each chain axis has four others around it at a distance of 4.48 Å. and two at a distance of 4.97 Å. (figure 2b).

We shall now consider various types of hypothetical structures conforming to the foregoing assumptions and general principles and see which satisfy best the requirements of the x-ray data for certain fibrous proteins.

EXTENDED ZIGZAG CHAIN STRUCTURES

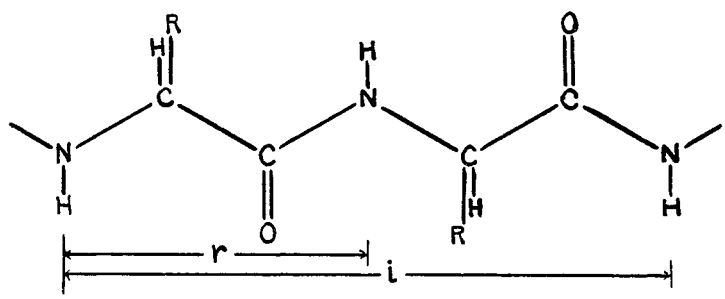
If a polypeptide chain, having bond distances and bond angles as given in the last column of table 1, is extended as much as possible, a zigzag structure results, in which all of the chain atoms (C, C′, N) and the carbonyl oxygen atoms lie in one plane; the H atoms of the CHR groups (and probably also those of the

TABLE 2

Approximate distances in certain structures

	PARALLEL TO FIBER AXIS		PERPENDICULAR TO FIBER AXIS	
	Identity distances if R's all equivalent	Average distance per residue	Between chains	Between layers
	Å.	Å.	Å.	Å.
Extended chain (calculated)	7.2	3.6	4.5	
Silk	7.0	3.5	4.5	9.2
β-Keratin	6.7	3.33	4.65	9.7
α-Keratin	10.3	1.7	9	9.8
Collagen	5.7	2.9	4.4	11

NH groups) and the nearest carbon atoms of the R radicals, on the other hand, do not lie in this plane. If all of the residues have a levo-configuration, the C—R bonds extend alternately above and below the plane of the zigzag.

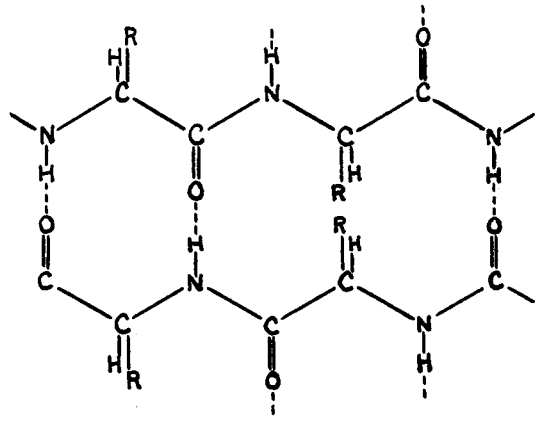


The distance r per residue in the direction of the chain axis is about 3.6 Å.; the identity distance i (neglecting differences in the R radicals) is twice this length.

X-ray data from silk fibroin (21, 31, 39, 40, 43, 44, 45, 51, 52, 56) and from β-keratin (12, 13) (stretched hair, horn, quill, fingernails, etc.) show apparent identity distances (table 2) in the direction of the fiber axis from 6.7 Å. to 7.0 Å.,

and it has been assumed (12, 47, 48), justifiably, that these substances are composed of polypeptide chains which are nearly fully extended.

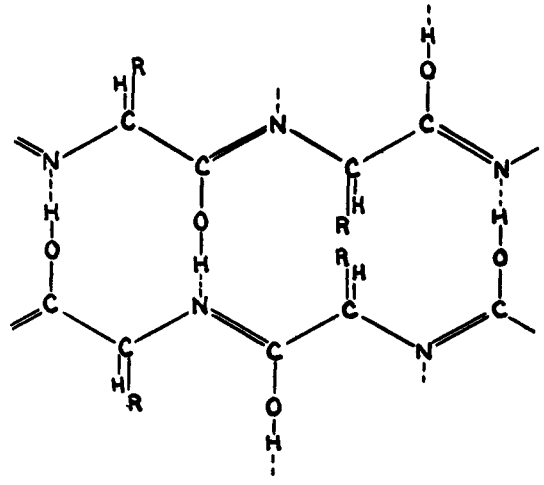
Reasoning from the principles outlined above, one would expect fully extended polypeptide chains in a protein to line up in sheets, with their chain axes parallel and with the hydrogen atom of each NH group linked to the oxygen atom of a CO group in the adjacent chain by means of an NHO hydrogen bridge, in this manner:



A

Although the x-ray data from silk and β-keratin cannot at present be said to prove that this hypothetical type of structure is correct, they seem to be in agreement with it, as will be shown below.

It may be noted that this structural arrangement permits the long-chain resonance and synchronized oscillations which, as has been pointed out above, make the hydrogen bonds especially stable. The other extreme resonance structure can be represented formally in the following way:



B

These formulas are somewhat idealized. For example, it is probable that the zigzag chains and the hydrogen bridges connecting them are not coplanar.

According to the Lewis theory of valence (42), corroborated by much experimental evidence on the structures of small molecules, the NH bonds in structure A should not be in the plane containing the centers of the nitrogen atom and the two carbon atoms to which it is bonded. Also, the C—O and O—H bonds in structure B and the C=O and O···H bonds in structure A should not be colinear. Moreover, coplanar zigzag chains would require the R groups attached to adjacent chains to be too close together.

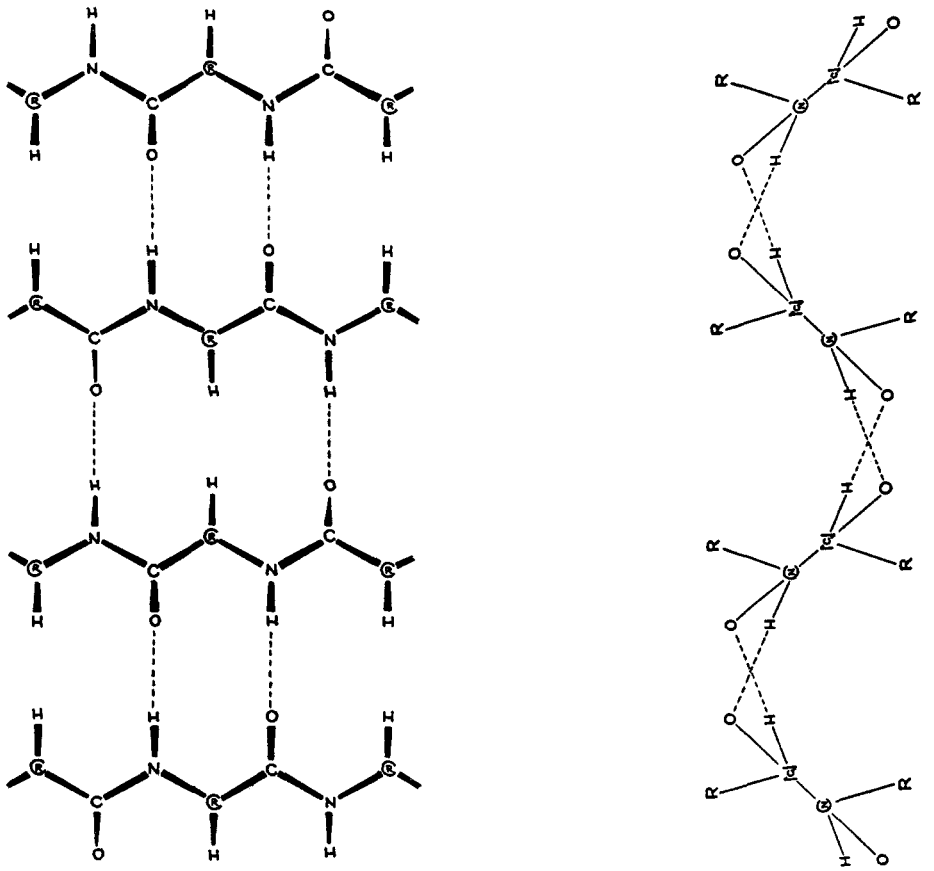


FIG. 3. Two views of a hypothetical structure pattern for a layer of extended polypeptide chains.

Figures 3 and 4 represent two structures (35) to which the foregoing objections do not apply. The latter seems less reasonable than the former, since, assuming a like (levo) configuration around each asymmetric carbon atom of a CHR group, the R groups of alternate chains are differently situated. The C—R bonds in half of the chains are approximately normal to the median plane of the layer. Those in the other half of the chains are so oriented as to place the first carbon atom of the R group only about 3 Å. from a carbon atom in the neighboring

chain. This is somewhat closer than the expected equilibrium distance (3.6–4.0 Å.). Only if all the "R groups" in half of the chains were hydrogen—i.e., if all of the residues in these chains were glycine residues—would the structure of figure 4 seem to be satisfactory.

The bond distribution shown in figure 5 should also be considered. Here the sequence —NH·CHR·CO— runs in the *same* direction for all the chains; in

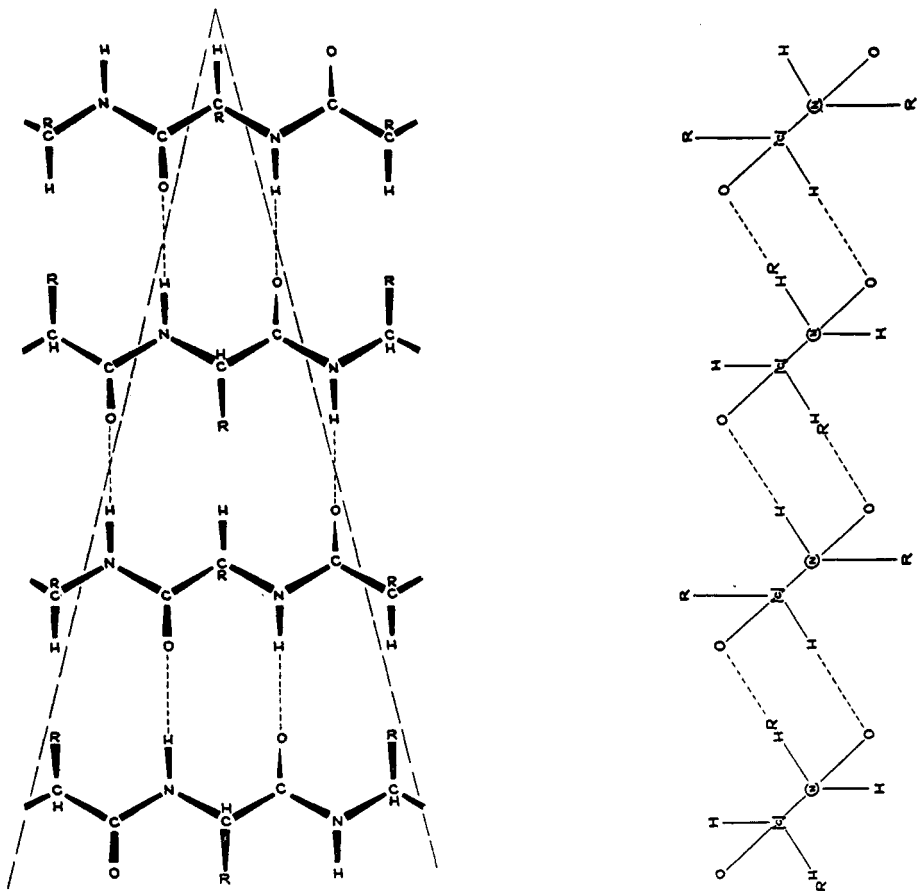


FIG. 4. Two views of another hypothetical pattern for a layer of extended polypeptide chains.

the structures of figures 3 and 4, *alternate* chains run in opposite directions.[3] The hydrogen bridges would tend to straighten out, causing the chains to coil up to some extent. This pattern therefore seems improbable for proteins in which the x-ray data show the chains to be practically fully extended.

For chains which are not fully extended, we should also consider structures

[3] In reference 35 it was incorrectly stated that the sequence is the same for all chains in the structure of figure 4.

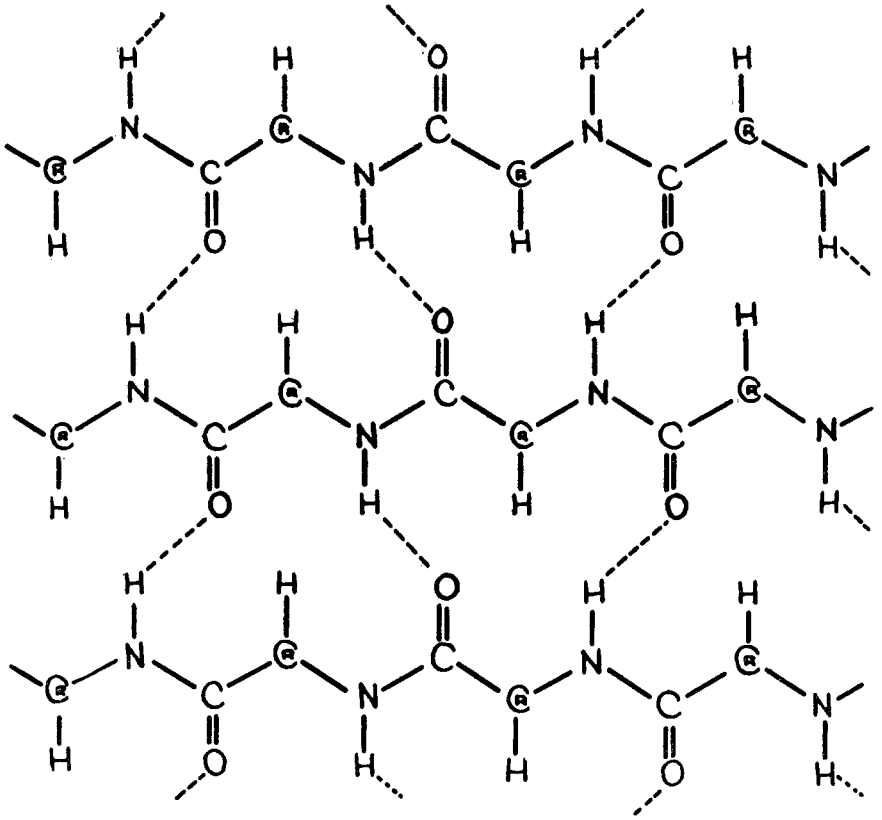


FIG. 5. Bond distribution in another hypothetical structure for a layer of extended polypeptide chains.

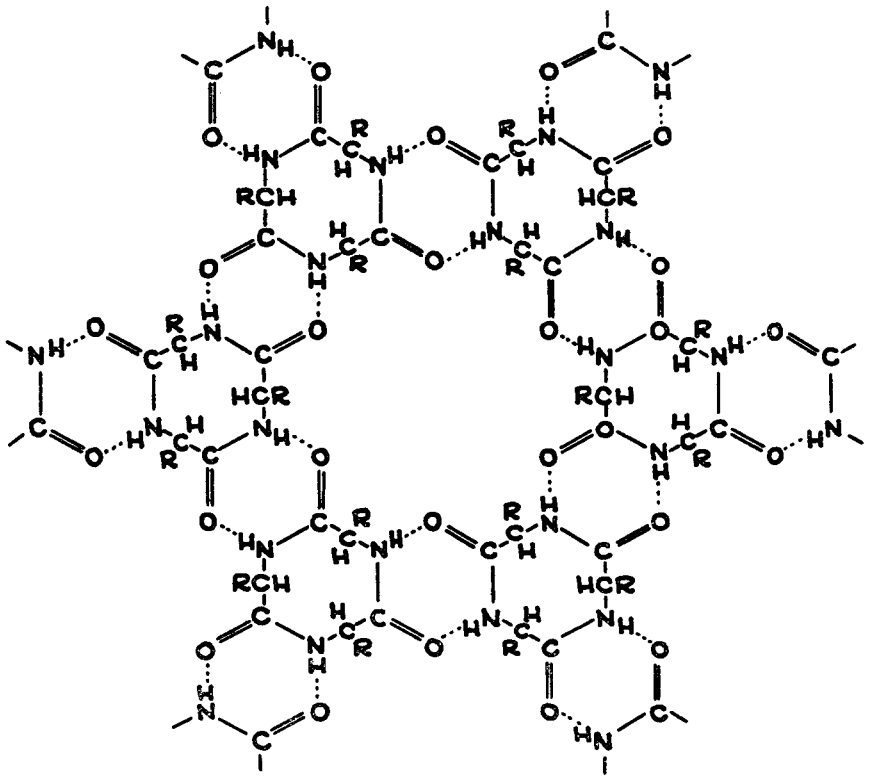


FIG. 6. Projection, on a plane normal to the axes of the spirals, of a hypothetical structure composed of spiral polypeptide chains, connected to one another through NHO bridges.

like those of figures 6 and 7, in which each chain spirals in such a way that it is connected by NHO bridges to *three* or more others, instead of just to *two*, as in the layer structures just described. Without entering into a detailed discussion of these structure patterns, we shall merely state here that for silk, α-keratin, β-keratin, and collagen they seem less likely, on the basis of the x-ray data, than the structures which will be described.

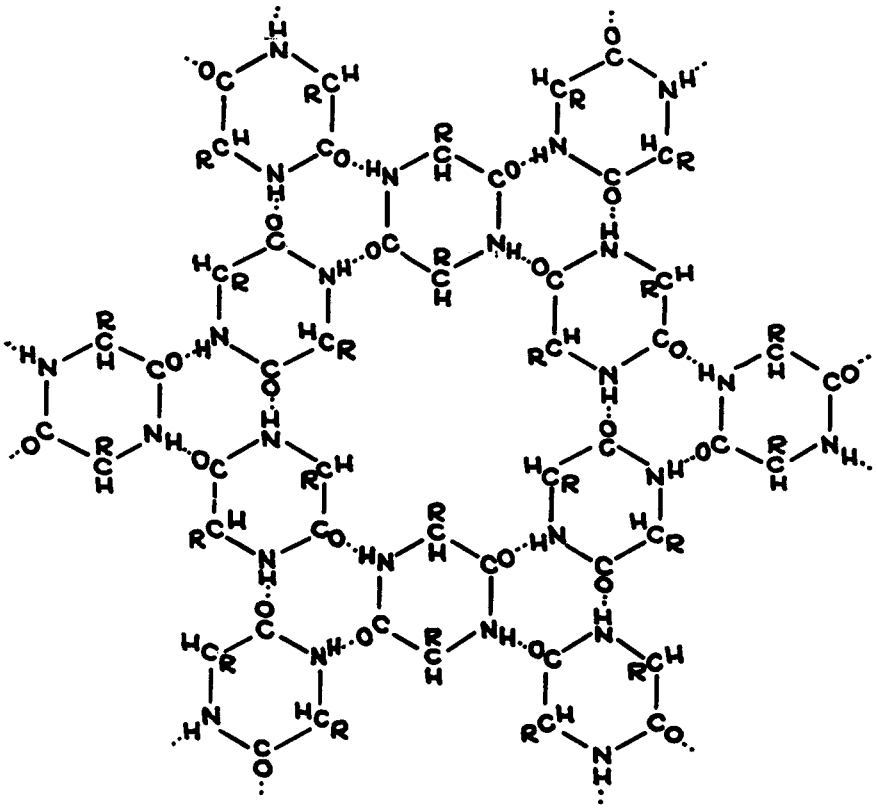


FIG. 7. A projection of another hypothetical structure composed of spiral polypeptide chains, linked together through NHO bridges.

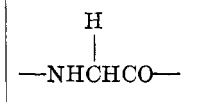
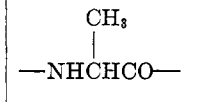
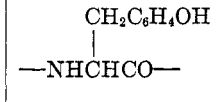
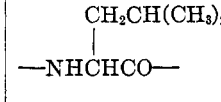
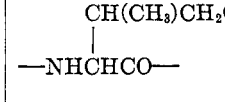
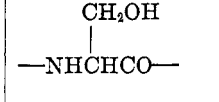
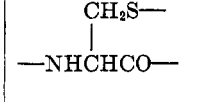
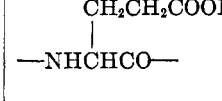
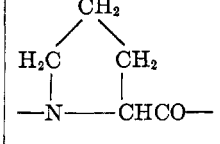
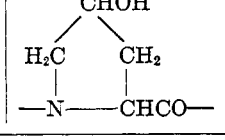
SILK FIBROIN

The x-ray studies of silk fibroin show that there are at least two kinds, having structures which are definitely different. The best and most complete x-ray data are from the type known as *Bombyx mori*; its structure will be that considered here.

Determinations of the composition of silk (2, 19, 46, 57) show (table 3) that approximately half of the residues are glycine (R = H) and about one-fourth are alanine (R = CH_3). Meyer and Mark (46, 47, 48) suggest that the substance consists of a crystalline portion, in which glycine and alanine (or alanine *plus* serine) residues are present in equal numbers, and an amorphous portion com-

TABLE 3

Approximate compositions of silk fibroin, keratin, and collagen

AMINO ACID		RESIDUE FRACTIONS		
		Silk fibroin	Wool keratin	Collagen
Glycine		0.49	~1/6	0.33
Alanine		0.25		0.09
Tyrosine		0.05		
Leucine		0.016	~1/9	0.05
Isoleucine				
Serine		0.014	~1/9	
Cystine ÷ 2			~1/8 (or 1/9)	
Glutamic acid			~1/9	
Proline				0.14
Hydroxyproline				0.10

posed of glycine residues and residues of all the other amino acids present. The argument for this is not very strong, however. Since about 15 per cent of the

residues are still unknown, it seems best, for the present, to consider silk fibroin as composed of equal numbers of glycine residues and of other residues, treating these others as if they were all alanine, but realizing that many of them are not.

Kratky and Kuriyama (39, 40) were able to account satisfactorily for the locations of all of their x-ray reflections from *Bombyx mori* on the basis of any one of the (pseudo) unit cells listed in table 4. A, B, C, D, E, and F in this table correspond to Kratky and Kuriyama's cells I, II, VIII, X, V, and VI, respectively. Their dimensions have been changed to those of the equivalent units having the interaxial angle nearer 90°. Their designations a, b, c, and γ have also been changed to c, a, b, and β, respectively. The probable error of each of the unit distances is perhaps 0.2 to 0.3 Å. (McNicholas (45) gives 7.04 Å. as the mean value of b, computed from twenty-two different diffraction photographs.) The length b is the (pseudo) identity distance in the direction of the fiber axis; a probably lies in the layer plane; $c \sin \beta$ is the perpendicular distance between layer planes, about 9.16 Å. in each case.[4]

TABLE 4

Dimensions of possible unit cells for silk fibroin (Bombyx mori), according to Kratky and Kuriyama

	a	b	c	β	NUMBER OF RESIDUES
	Å.	Å.	Å.		
A	4.72	6.95	9.25	86° 10′	2
B	4.95	6.95	9.15	89° 26′	2
C	4.60	6.95	9.26	81° 39′	2
D	17.71	6.95	9.75	70° 1′	8
E	8.92	6.95	9.40	76° 40′	4
F	8.98	6.95	9.37	77° 43′	4

A structure for silk fibroin consisting of layers of the type of figure 4, with alternate chains within each layer composed entirely of glycine residues, seems to be ruled out by the identification (1) of the dipeptides *d*-alanylglycine and glycyl-*l*-tyrosine.

Unit cells A, B, and C can be ruled out as improbable on several counts, the most important being that they would require all chains to be oriented in the same way. Assuming an extended zigzag chain, with glycine and other residues alternating, all of the C—R bonds would extend on the same side of the plane of the zigzag in all of the chains. There would be no symmetry elements in the structure whatever. All the chains and all of the layers would tend to bend continuously in the same direction. With a structure of this sort, silk would give x-ray diffraction photographs characteristic of amorphous materials.

[4] Meyer and Mark (48) state that Herzog and Kratky, in a private communication, give as the simplest possibility a unit cell having the following dimensions: 8.80 Å., 7.00 Å., 9.68 Å., 75° 50′. Since Meyer and Mark's book was apparently published early in 1930—the preface is dated May, 1930—whereas the Kratky and Kuriyama paper was not received for publication until October 25, 1930, we may assume that these (Herzog and Kratky) dimensions were superseded by those given by Kratky and Kuriyama.

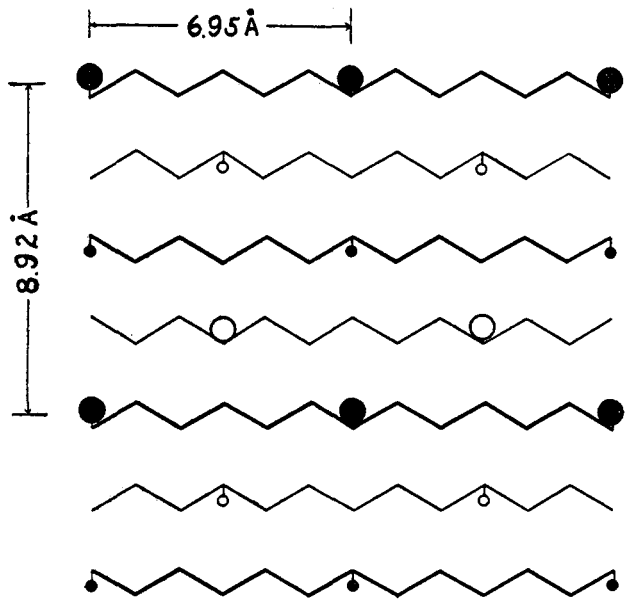


FIG. 8. Structure E for silk fibroin, assuming the layer structure of figure 3. The relative positions of the zigzag chains in two adjacent layers (heavy lines, lower layer; light lines, upper layer) are shown, with the positions of those R groups (large circles) and H atoms (small circles) of the CH_2 groups in the glycine residues which lie between the median planes of the two layers.

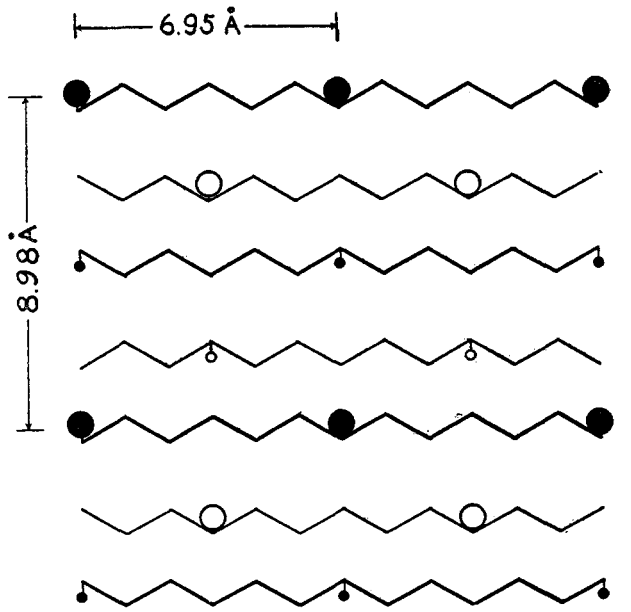


FIG. 9. Structure F for silk fibroin, assuming the layer structure of figure 3. This shows the relative positions of the chains in two adjacent layers, with the disposition of R groups and H atoms between these layers.

Another argument against the A, B, and C units is that, with glycine and other residues alternating in each chain, 010 reflections should be observed, whereas none have been reported.

Still another argument against the A, B, and C units is that the a distance in each is only about half the minimum possible value for a structure of the type of figure 3, and the b distance is close enough to that calculated for fully extended chains to make a bond distribution like that depicted in figure 5 improbable. Moreover, the chains in adjacent layers would be nearly over and under each other, giving an unreasonable packing of R groups and H atoms between layers.

The unit cell D gives a similar unlikely distribution between the layers.

Units E and F seem much more reasonable Assuming a structure of the type shown in figure 3, with the C—H and C—R bonds tilted as indicated in the projection on the right-hand side of that figure (the angle of tilt being somewhat arbitrary), the interlayer distribution of H atoms and R groups (mainly CH_3) is as shown in figures 8 and 9. Both of these appear reasonable. A decision between them does not seem possible at the present time.

β-KERATIN

As Astbury has pointed out (6, 13), x-ray photographs of β-keratin are similar, as regards their major features, to those of silk fibroin.

With the meager x-ray data obtainable, a good direct determination of the dimensions and symmetry of even a pseudo unit cell (i.e., the true unit if all the R groups were alike) is not possible. Astbury states, however, that his data can be accounted for on the basis of an orthogonal cell of the following dimensions:

$$a = 9.3 \text{ Å.}; b = 6.7\text{–}6.8 \text{ Å.}; c = 9.8 \text{ Å.}$$

Tentatively, we may accept these. Assuming layers of the figure 3 type, as in silk, b is the length of two residues in the direction of the chain axis, a is twice the distance between adjacent chain axes in each layer, and c is the average distance between layers. The larger value of c and the smaller value of b than in silk may reasonably be attributed to the larger average size of the R groups.

This unit cell places the chains in adjacent layers directly over or under each other, rather than shifted as shown in the structures deduced for silk (figures 8 and 9), but the x-ray data from β-keratin can hardly be said to prove this point. If Astbury is correct about this, the chain-over-chain arrangement may be a result of cystine (and perhaps other) bridges between chains in different layers.

From known interatomic distances (33, 50) and the assumption of tetrahedral bond angles ($\sim 109.5°$), one can compute a maximum distance of 8.0 Å. between two chain axes joined by a cystine radical. This is perhaps not in too poor agreement with the observed average interlayer distance of 9.8 Å., since the variety of R groups doubtless produces large distortions from the idealized structure models represented in the figures and since the bond angles may be somewhat larger than the tetrahedral angle.

Another alternative which appears more reasonable is that the cystine radicals

bridge between adjacent chains *in the same layer.* The calculated distance for this ($\sim$ 4.7 Å.) is in good agreement with the experimental distance (4.65 Å.) between chain axes.

Interchain bridges need not be limited to hydrogen bridges and cystine bridges. There may also be amide or ester linkages, for example, such as could be formed by condensation between serine and glutamic acid. From the analytical figures (3, 5, 8, 11, 22, 58), approximately one-ninth of the residues are serine, one-ninth are glutamic acid, and one-ninth are "cystine ÷ 2." (Each cystine molecule obtained in the analysis comes from two "cystine ÷ 2" residues.) It seems reasonable to guess that there is a cystine or serine–glutamate bridge connecting each chain to its immediate neighbors at every third residue

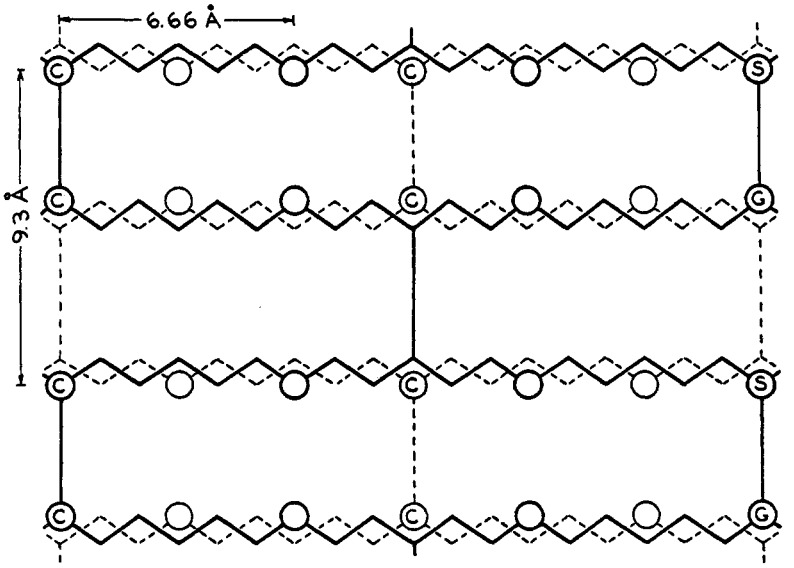


FIG. 10. Possible structure of β-keratin, showing the disposition of the zigzag chains in two adjacent layers, with the packing of R groups (circles) between them. A suggested distribution of cystine (C–C) and serine–glutamate (S–G) bridges is also indicated.

(figure 10). This is perhaps too speculative at this stage of the game, however; these details are not required by the present considerations.

α-KERATIN

When hair is stretched in steam, changing from α-keratin to β-keratin, there is an extension of the polypeptide chains of about 100 per cent, according to Astbury (13). If this is correct, the average length per residue in the direction of the fiber axis is about 1.7 Å. in the α-form. There is a strong x-ray reflection (or pair of reflections), due to planes approximately normal to this axis, spaced about 5.1 Å. apart,—about three times this average extension per residue. It seems necessary to assume, with Astbury, some sort of coiling of the chains.

Figures 11, 12, and 13 show three ways[5] in which a polypeptide chain can be coiled, consistent with the following assumptions:[6] (*1*) bond distances and angles are the expected ones; (*2*) atoms not directly bonded together are not too close together; (*3*) like atoms (or groups) are surrounded equivalently; (*4*) adjacent turns are connected by NHO hydrogen bridges. The structures of figures 11 and 12 have twofold screw axes of symmetry; there are two amino acid residues per coil; the identity distance, equal here to the distance between corresponding points in adjacent coils, is roughly 5 Å. The 5.1 Å. reflection can be accounted for on the basis of either of these structures, *provided one introduces the additional assumption that alternate R groups are much more potent x-ray scatterers*

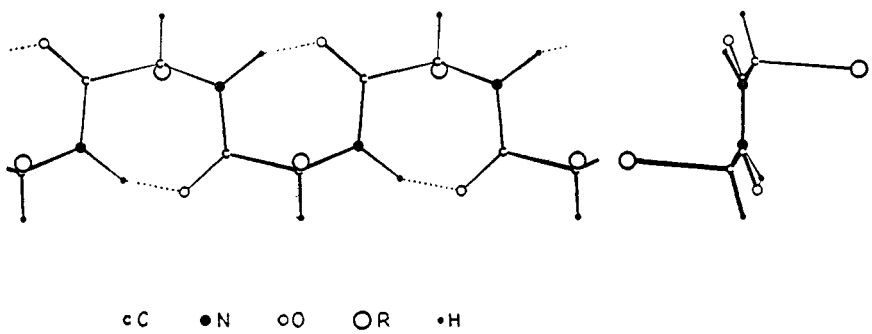


FIG. 11. Two views of a hypothetical coiled structure for a polypeptide chain, with intrachain NHO hydrogen bridges.

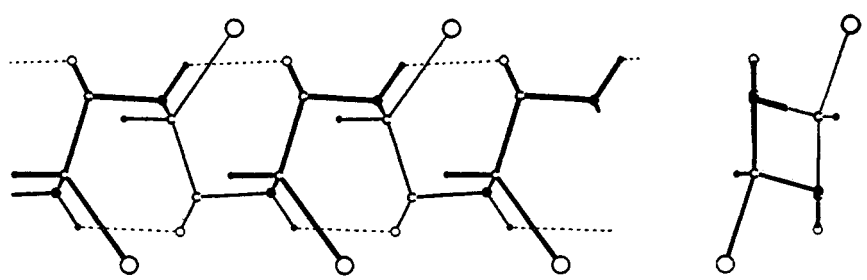


FIG. 12. Two views of a hypothetical spiral structure for a polypeptide chain, with intrachain NHO bridges.

than the intermediate ones. The change from α- to β-keratin, in either case, involves an extension of only about one-third, however, rather than the 100 per cent extension required by Astbury.

It might be possible, by distorting the bond angles considerably, to compress

[5] These three structures were discussed, with slides and models, in a paper presented by the writer before the Division of Biological Chemistry at the Rochester Meeting of the American Chemical Society, September, 1937. The structure of figure 13 is apparently identical with one described in a recent paper by H. S. Taylor (55). The author is glad to express here his indebtedness to Professor Taylor for calling his attention to this paper.

[6] None of the models discussed for α-keratin by Astbury (7, 10, 13, 14, 29) is in agreement with these assumptions.

the model pictured in figure 12 until the distance per *three* residues in the direction of the spiral axis became 5.1 Å. This would give the 100 per cent α-to-β extension. To account for the 5.1 Å. x-ray reflection, one would then need to assume every *third* R group to be an especially strong x-ray scatterer.

In the structure depicted in figure 13 there are about three residues per turn of the spiral. The distance per residue (measured parallel to the spiral axis) is about 1.7 Å. This gives the 100 per cent extension in the α-to-β change, and also the 5.1 Å. x-ray spacing—provided every third R group is assumed to scatter x-rays much more strongly than the others.

It may be noted that there is nothing about this structure which requires exactly three residues per turn of the spiral. In fact, it would seem, from the models that have been made, that the bond distance and angle requirements are best satisfied by a slightly smaller number of residues per turn.

Still another hypothetical α-keratin structure will now be described. This may best be done by showing the relationship to the structure of figure 11. Figure 14a shows the ribbon-like structure of figure 11, slightly idealized for

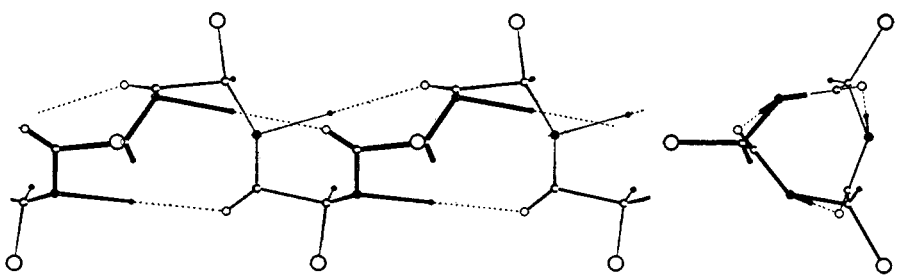


FIG. 13. Two views of another hypothetical spiral structure for a polypeptide chain, with intrachain NHO bridges.

simplicity; figure 14b is an edge-on view of this arrangement, showing the simple zigzag bending of the ribbon; figure 14c represents an alternative manner of bending, also possible without distortion of bond angles and distances from their preferred values. This arrangement gives the 100 per cent α-to-β extension and also accounts for the observed 5.1 Å. x-ray spacing—*without any additional assumptions regarding the scattering powers of the R groups.* At the present time, this type of structure seems to the writer more reasonable than any other of which he is aware.

It is worth pointing out that, if we assume β-keratin to have the structure of figures 3 and 10 and α-keratin to have that of figure 14 (a and c), the transition from β to α (or *vice versa*) involves no breaking of bonds (except the O···H bonds of the hydrogen bridges), either in cystine or other cross links or elsewhere, and no radical changes in bond distances or angles at any stage of the process. This statement is true whether the cross links connect chains within the same layer or in different layers.

The transition from β- to α-keratin can be visualized in the following way: The β-keratin structure can be readily warped in the manner indicated in fig-

ure 15, with no breaking of covalent bonds (either within the chains or in cross links) or hydrogen bridges and with but little change in energy. By a relatively

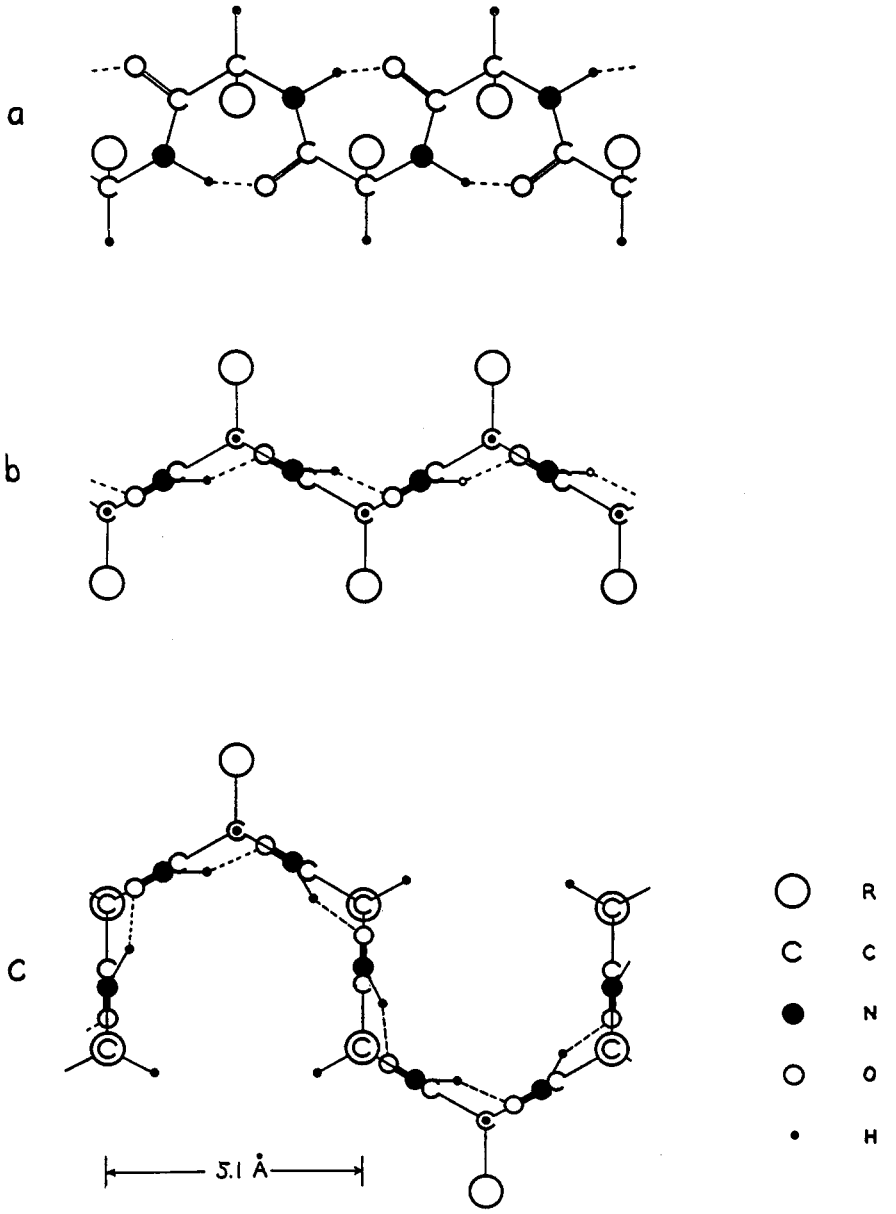


FIG. 14. (a) An idealized representation of the structure of figure 11; (b) an edge-on view of this structure, showing the zigzag folding; (c) showing another manner of folding the ribbon-like structure represented in 14a.

simple shift of position of each bridging hydrogen, the NHO bridges *between* chains can now be replaced by similar bridges *within* the chains. This gives the

distribution of atoms and bonds within each chain shown in figure 14a. The atoms shown in this figure, however, cannot be coplanar; the ribbon-like structure of the chain must be bent, at the carbon atoms of the CHR groups. Considering each chain by itself, we might expect this bending to take place in such a way as to give the structure of figure 11 and figure 14b. Since the x-ray and extensibility data seem to favor the figure 14c structure, we may infer that

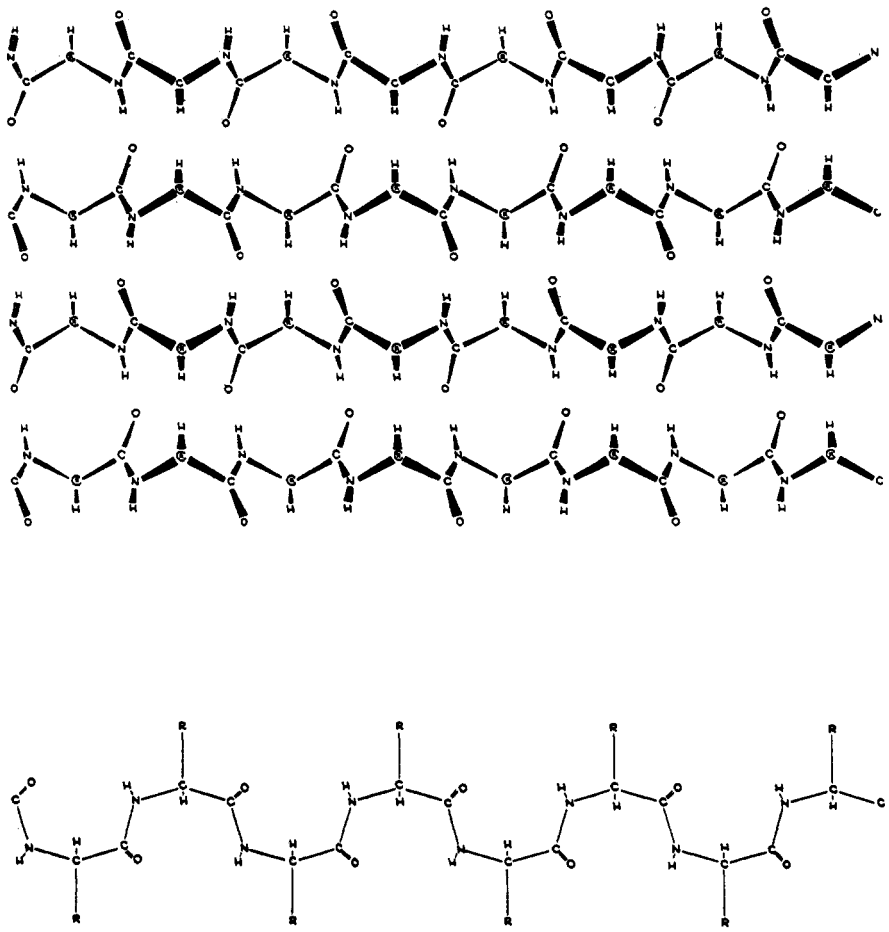


FIG. 15. Illustrating a possible intermediate stage between α-keratin and β-keratin.

the latter type of folding is more stable than the former—perhaps owing to a better distribution of R groups in the complete structure.

In addition to the strong 5.1 Å. "meridian" x-ray reflection already mentioned, α-keratin also gives two strong (though not well-defined) "equator" reflections, from sets of planes parallel to the fiber axis. One has a spacing of 27 $\pm$ 2 Å. The other is described by Astbury and Street (12) in the following words: "The most prominent interference on the equator is the disproportionately large spot formed round (001). There is little doubt that this is not a single reflection.

It is spread over about 3 Å. . . . The region of maximum density corresponds to 9.8 Å. . . . " This information, plus the small amount of additional information available, is insufficient to enable one to deduce a single unique arrangement of coiled chains—even assuming them to be of a particular type, say that represented in figures 14a and 14c. In one of several structures which look reasonable, the chains are tied together in layers through cystine and serine–glutamate bridges in the manner indicated in figure 16, these layers being stacked together with an average interlayer distance of about 9 Å. The strong 27 Å. spacing suggests that either the spacing between these layers is not uniform (every third spacing being relatively small, perhaps), or else the compositions of successive layers are not the same (every third layer having an excess of R groups which

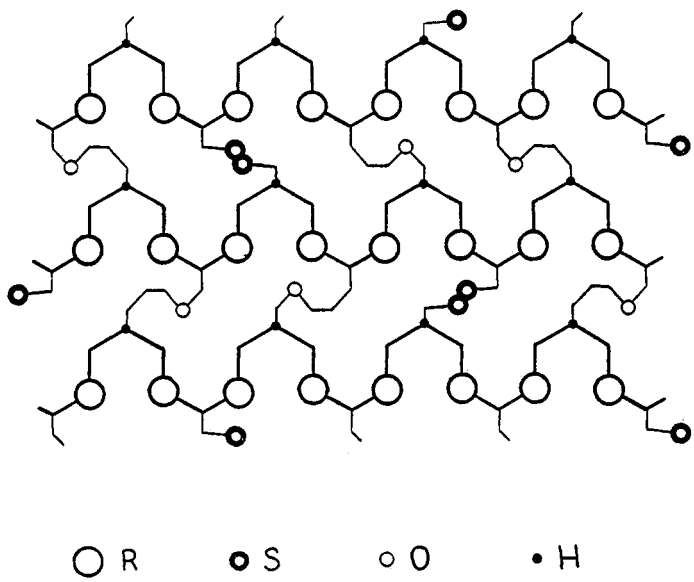


FIG. 16. Hypothetical distribution of spiral chains of the type of figure 14 (a and c) in a layer of the α-keratin structure, assuming cystine and serine–glutamate bridges connecting the chains.

are especially strong x-ray scatterers, for example), or both. A strong third order (9 Å.) spacing would be expected; it may well be present, being included in the strong, very diffuse spot described above.

In addition to the few strong diffuse reflections discussed above, certain substances (e.g., sea gull's quill and porcupine quill) classed by Astbury as α-keratin show also a considerable number of other reflections, some corresponding to quite large (up to 100 Å.) interplanar distances. These reflections may be attributed to the distribution of R groups in the structure; the true unit of structure must be much larger than the pseudo unit of the idealized pattern just described. Astbury (9) interprets MacArthur's (9) "meridian" reflections from porcupine quill as indicating a probable identity distance in the direction of the fiber axis of 658 Å. Pending the publication of more complete data, however, further discussion of the possible significance of these results seems unwarranted.

COLLAGEN

Figure 17 shows a way in which polypeptide chains can be coiled spirally so as to give NHO bridges *between* different chains, holding them together in layers. (The structures represented by figures 6 and 7 are not layer structures.) The R groups extending above or below each layer are very uniformly distributed, without crowding. The layers can be piled together in such a way (figure 18) as to give approximately a "close-packed" arrangement of R groups between each pair.

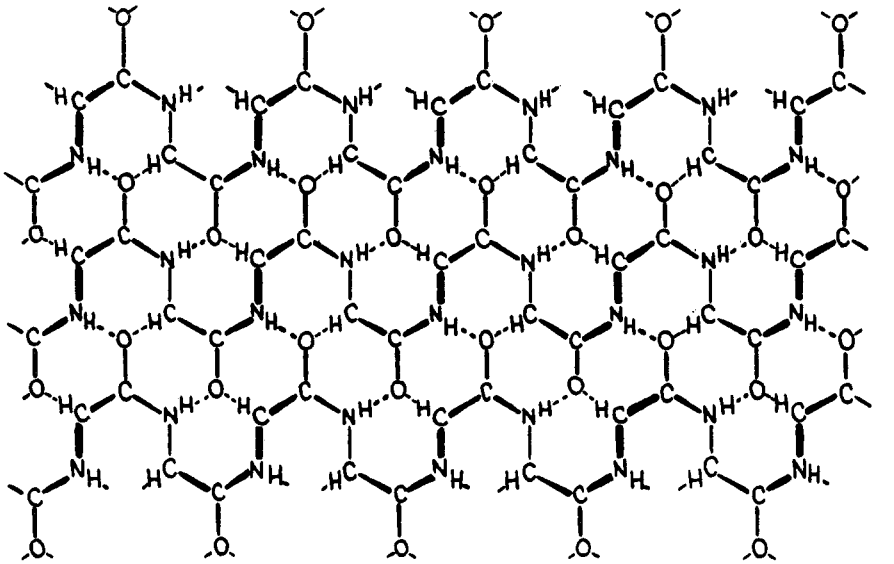


FIG. 17. Hypothetical structure of a layer of spiral polypeptide chains in collagen. An R group (or H atom, in case the residue in question is glycine) is assumed to be directly over or directly under the C of each CH.

The pseudo unit of the structure represented in these figures (on the assumption that all R groups are equivalent) has the following dimensions:

$$a \sim 4.5 \text{ Å.}; b \sim 5.8 \text{ Å.}; c \sim 22 \text{ Å.}; \beta \sim 90°$$

(The c distance, assumed to be twice the average distance between the median planes of adjacent layers, varies greatly with the water content.)

This structure seems a reasonable one for collagen. It accounts simply for the observed (7, 10, 26, 59) strong 2.9 Å. x-ray reflection from planes nearly normal to the fiber axis and the strong equator reflections corresponding to distances of about 4.4, 5.4, and 11 Å. The indices of these four reflections, on the basis of the pseudo unit, are 022, 101, 004, and 002, respectively. In spite of this, however, it must be emphasized that this structure must be considered only as a guess. It is certainly neither deducible from the x-ray observations nor proven correct by them.

Because of the differences in the R groups, the true unit must be much larger than the pseudo unit just described. The observed large-spacing reflections

from planes parallel to the fiber axis may be considered as evidence for this. The long-spacing reflections from planes normal to the fiber axis, such as the many orders of a 640 Å. spacing reported by Bear (16), are doubtless the result of the banded structure (of the same periodicity) which collagen fibers possess, according to electron-microscope studies of Schmitt, Hall, and Jakus (53). Dense regions, which might possibly have a structure of the type of figure 17, alternate with relatively transparent regions having a much more open structure.

The analytical data (7, 10, 17, 18, 20, 27, 38) on collagen suggest that one-third of the residues in collagen are glycine, one-sixth are proline, one-ninth are hydroxyproline, one-ninth are alanine, etc. (see table 3). The large fraction of residues which are proline or hydroxyproline, in which the carbon atom of the

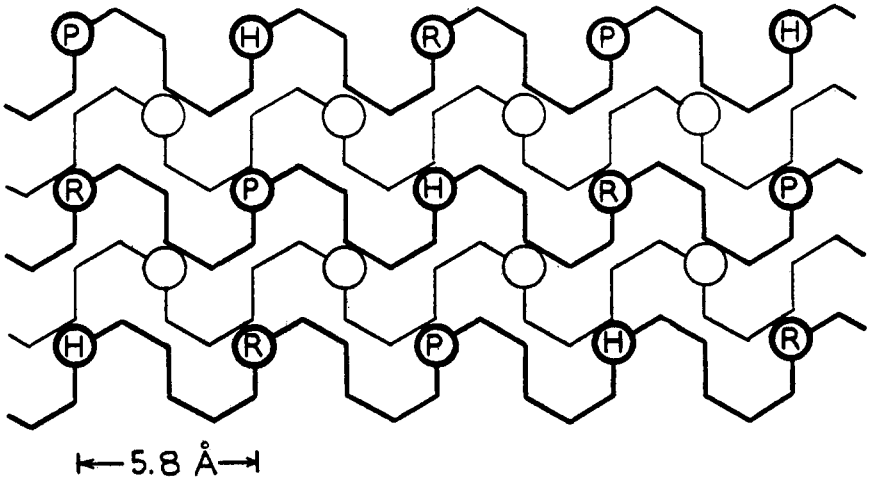
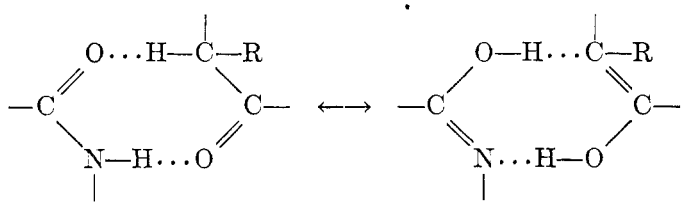


FIG. 18. Possible manner of packing adjacent layers of the type of figure 17. Heavy circles containing the letters P, H, and R denote R groups (or H atoms) pointing *up* from the lower layer; light circles, without letters, denote R groups (or H) pointing *down* from the upper layer. The letters P, H, and R indicate a suggested distribution of proline (and oxyproline), glycine, and other residues, respectively, on the assumption that these three classes are present in equal numbers.

CHR and the adjacent nitrogen atom in the polypeptide chain are both part of a five-membered ring, is probably important. Assuming, as usual, a levo-configuration around this carbon atom, the orientation of the ring relative to the neighboring bonds in the chain is fixed. Residues of this sort cannot fit into any of the structures discussed above for silk, α-keratin, or β-keratin, but they do fit readily into a structure of the type of figure 17. The ring extends either above or below the layer, without any crowding. This may explain why collagen assumes this type of structure. It seems equally reasonable to assume, however, that the collagen structure is, in general, the stable one, the α-keratin structure being adopted only when there are sufficient cystine or other bridges between adjacent chains.

In figure 17, C—H···O hydrogen bridges, connecting the CHR groups

with CO groups, have been indicated. Although CHO bridges are not common, they would be expected in a structure of this type, being stabilized by resonance and by synchronized oscillations of neighboring bridges. One out of several modes of resonance contributing to the stability of the structure can be represented by the formulation



Wherever one of the R groups is proline or oxyproline, the adjacent nitrogen atom in the chain (which is also a part of the proline ring) has no hydrogen atom directly attached to it; therefore, it cannot form a hydrogen bridge to the nearby carbonyl oxygen. In view of the neighboring CHO bridges, however, the absence of the NHO bridge would not be expected to cause a rupture or great distortion of the structure pattern.

CONCLUSION

No claim can be made that the structures proposed here for silk fibroin, keratin, and collagen are proven, that they are correct in all their details, or that they are complete. The writer does claim, however, that they are in considerably better agreement with the available experimental data than are the other structures which have been proposed. Further analytical and x-ray data, he feels sure, will either verify these structures or will show that others quite closely related to them are correct; they should also lead to further refinements, especially as regards the distribution of R groups.

In this paper no reference has been made to the structures of globular proteins. There is every reason to believe, however, that the same principles apply as in the case of the fibrous proteins.

REFERENCES

(1) ABDERHALDEN, E.: Z. physiol. Chem. **62,** 315 (1909); **63,** 401 (1909); **65,** 417 (1910); **72,** 1 (1911).
(2) ABDERHALDEN, E.: Z. physiol. Chem. **120,** 207 (1922).
(3) ABDERHALDEN, E., AND WELLS, H. G.: Z. physiol. Chem. **46,** 31 (1905).
(4) ALBRECHT, G., AND COREY, R. B.: J. Am. Chem. Soc. **61,** 1087 (1939).
(5) ARGIRIS, A.: Z. physiol. Chem. **54,** 86 (1907).
(6) ASTBURY, W. T.: Trans. Faraday Soc. **29,** 193 (1933).
(7) ASTBURY, W. T.: J. Intern. Soc. Leather Trades' Chem. **24,** 69 (1940).
(8) ASTBURY, W. T.: Chemistry & Industry **60,** 491 (1941).
(9) ASTBURY, W. T.: J. Chem. Soc. **1942,** 337.
(10) ASTBURY, W. T., AND BELL, F. O.: Nature **145,** 421 (1940).
(11) ASTBURY, W. T., AND BELL, F. O.: Nature **147,** 696 (1941).
(12) ASTBURY, W. T., AND STREET, A.: Trans. Roy. Soc. (London) **A230,** 75 (1931).
(13) ASTBURY, W. T., AND WOODS, H. J.: Nature **126,** 913 (1930); Trans. Roy. Soc. (London) **A232,** 333 (1933).

(14) ASTBURY, W. T., AND WRINCH, D. M.: Nature **139,** 798 (1937).
(15) BATH, J. D., AND ELLIS, J. W.: J. Phys. Chem. **45,** 204 (1941).
(16) BEAR, R. S.: J. Am. Chem. Soc. **64,** 727 (1942).
(17) BERGMANN, M.: J. Biol. Chem. **110,** 471 (1935).
(18) BERGMANN, M., AND NIEMANN, C.: J. Biol. Chem. **115,** 77 (1936).
(19) BERGMANN, M., AND NIEMANN, C.: J. Biol. Chem. **122,** 577 (1938).
(20) BERGMANN, M., AND STEIN, W. H.: J. Biol. Chem. **128,** 217 (1939).
(21) BRILL, R.: Ann. **434,** 204 (1923).
(22) BUCHTALA, H.: Z. physiol. Chem. **52,** 474 (1907).
(23) BUSWELL, A. M., KREBS, K. F., AND RODEBUSH, W. H.: J. Phys. Chem. **44,** 1126 (1940).
(24) BUSWELL, A. M., RODEBUSH, W. H., AND ROY, M. F.: J. Am. Chem. Soc. **60,** 2528 (1938).
(25) COREY, R. B.: J. Am. Chem. Soc. **60,** 1598 (1938).
(26) COREY, R. B., AND WYCKOFF, R. W. G.: J. Biol. Chem. **114,** 407 (1936).
(27) DAKIN, H. D.: J. Biol. Chem. **44,** 499 (1920).
(28) EARP, D. P., AND GLASSTONE, S.: J. Chem. Soc. **1935,** 1709.
(29) FRANK, F. C.: See Astbury, W. T., J. Textile Inst. **27,** P282 (1936).
(30) GLASSTONE, S.: Trans. Faraday Soc. **33,** 200 (1937).
(31) GOLDSCHMIDT, S., AND STRAUSS, K.: Ann. **480,** 266 (1930).
(32) GORDY, W.: J. Am. Chem. Soc. **60,** 605 (1938); Nature **142,** 831 (1938); J. Chem. Phys. **7,** 163 (1939).
(33) HUGGINS, M. L.: Phys. Rev. **28,** 1086 (1926).
(34) HUGGINS, M. L.: J. Org. Chem. **1,** 407 (1936).
(35) HUGGINS, M. L.: J. Chem. Phys. **8,** 598 (1940).
(36) HUGGINS, M. L.: Ann. Rev. Biochem. **11,** 27 (1942).
(37) KOHLHAAS, R., AND SOREMBA, K.-H.: Z. Krist. **A100,** 47 (1939).
(38) KOSSEL, A., AND GROSS, R. E.: Z. physiol. Chem. **135,** 167 (1924).
(39) KRATKY, O.: Z. physik. Chem. **B5,** 297 (1929).
(40) KRATKY, O., AND KURIYAMA, S.: Z. physik. Chem. **B11,** 363 (1931).
(41) LEVY, H. A., AND COREY, R. B.: J. Am. Chem. Soc. **63,** 2095 (1941).
(42) LEWIS, G. N.: *Valence and the Structure of Atoms and Molecules.* The Chemical Catalog Company, Inc., New York (1923).
(43) MATANO, T.: J. Soc. Chem. Ind., Japan **42,** Suppl. binding 30 (1939).
(44) MATSUNAGA, Y.: Naturwissenschaften **24,** 446 (1936); Mem. Coll. Sci. Kyoto Imp. Univ. **A20,** 157 (1937).
(45) MCNICHOLAS, H. J.: Textile Research **11,** 39 (1940).
(46) MEYER, K. H., FULD, M., AND KLEMM, O.: Helv. Chim. Acta **23,** 1441 (1940).
(47) MEYER, K. H., AND MARK, H.: Ber. **61,** 1932 (1928).
(48) MEYER, K. H., AND MARK, H.: *Der Aufbau der hochpolymeren organischen Naturstoffe.* Akademische Verlagsgesellschaft, Leipzig (1930).
(49) MÜLLER, A.: Proc. Roy. Soc. (London) **A120,** 437 (1928).
(50) PAULING, L., AND HUGGINS, M. L.: Z. Krist. **A87,** 205 (1934).
(51) PHILIPP, H.: Umschau **41,** 624 (1937).
(52) SAKURADA, I., AND MATSUSHITA, Y.: J. Soc. Chem. Ind., Japan **40,** Suppl. binding 58 (1937).
(53) SCHMITT, F. O., HALL, C. E., AND JAKUS, M. A.: J. Cellular Comp. Physiol. **20,** 11 (1942).
(54) STANFORD, S. C., AND GORDY, W.: J. Am. Chem. Soc. **63,** 1094 (1941).
(55) TAYLOR, H. S.: Proc. Am. Phil. Soc. **85,** 1 (1941).
(56) TROGUS, C., AND HESS, K.: Biochem. Z. **260,** 376 (1933).
(57) VICKERY, H. B., AND BLOCK, R. J.: J. Biol. Chem. **93,** 105 (1931).
(58) VICKERY, H. B., AND LEAVENWORTH, C. S.: J. Biol. Chem. **83,** 523 (1929).
(59) WYCKOFF, R. W. G., AND COREY, R. B.: Proc. Soc. Exptl. Biol. Med. **34,** 285 (1936).
(60) ZELLHOEFFER, G. F., COPLEY, M. J., AND MARVEL, C. S.: J. Am. Chem. Soc. **60,** 1337 (1938) and subsequent papers by Copley and coworkers in the same journal.