On the Structures of Fibrous Proteins. I. A New Polypeptide Structure of α-Keratin

By Yoshiya KANDA

(Received October 31, 1949)

I. Introduction.—A question of the first importance for the structural-chemical study of proteins is to determine by the results obtained from physical and chemical experiments how the polypeptide chains are constructed from their constituent molecules or molecular residues, and how they are assembled. For fibrous proteins several model structures have been proposed by different authors, but all of them appear to be still unsatisfactory. Astbury's model for α-keratin,[1] for example, involves some unreasonable points.[2] Simanouti and Mizushima[3] offered a bent form model for α-keratin and an extended form for β-keratin, concluding that these models could reasonably explain many problems concerning the structures of proteins to an appreciable extent. However, the tetrahedral configuration they assumed to represent the C atom of the carbonyl radical does not accord with experimental facts. It is known from the results of electron diffraction experiments on formaldehyde[4] and acetaldehyde[5] that the three bonds of the C atom in a carbonyl radical lie in one plane, making angles of about 120° mutually. Mizushima and Simanouti[6] reported also that the identity distance of α-keratin chain calculated from their bent form model, using Pauling's covalent radii, agrees very well with the observed 5.15 Å.[1] If we calculate this distance, taking 109°28′ for the bond angles on each of the α-carbon and the N atoms, and 120° for those on the C atom in the carbonyl radical, it becomes 5.12 Å. However, the N···O distance of the N–H···O hydrogen bond comes out to be 1.98 Å. This is much shorter than the ordinarily accepted value for this bond of about 2.85 Å. Assuming the tetrahedral angles also for the C atom in the carbonyl radical, as indicated in Simanouti and Mizushima's figure, the identity distance becomes 4.91 Å., and then the overall N–H···O distance shortens to a much smaller value 1.82 Å. This would mean the existence of a steric hindrance between the N and the O atoms.

Recently Mizushima, Simanouti *et al.*[7] studied the structure of N-methyl acetamide, which they regarded as a constructive unit of polypeptide chains, by ultraviolet absorption, Raman effect, infrared absorption and dipole moment measurements. They found that this molecule has no *enol*-form, and it involves a resonance between two structures, and the two methyl radicals were concluded to be in *trans* position in a coplanar configuration. These findings all seem to favor their bent and extended models for keratins, but whether the relative orientations of atoms found in the simple N-methyl acetamide molecule would directly represent the circumstances obtaining in a highly polymerized substance like keratin chains is a question we must be very cautious about. Their conclusions appear to need more experimental verifications in this very point, although their results about the N-methyl acetamide molecule itself appears very good. We would rather consider that the -C-N-C-C- chain structure in the N-methyl acetamide

(1) W. T. Astbury and A. Street, *Phil. Trans. Roy. Soc.*, **A230**, 75 (1931). W. T. Astbury and H. T. Wood, *Phil. Trans. Roy. Soc.*, **A232**, 333 (1933).

(2) K. H. Meyer and H. Mark, "Hochpolymere Chemie, II" (1940).

(3) T. Simanouti and S. Mizushima, This Bul-Bulletin, **21**, 1 (1948).

(4) D. P. Stevenson, J. E. Lu Valle and V. Schomaker, *J. Am. Chem. Soc.*, **61**, 2508 (1939).

(5) D. P. Stevenson, H. D. Burnham and V. Schomaker, *J. Am. Chem. Soc.*, **61**, 2922 (1939).

(6) S. Mizushima and T. Simanouti, "Structures of Proteins." (in Japanese), p. 77. Gakuzyutu Tosho Syuppansha, Tokyo (1949).

(7) S. Mizushima and T. Simanouti, *Kagaku* (in Japanese), **19**, 252 (1949).

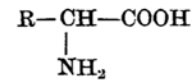
molecule does not necessarily be kept exactly in fibrous proteins, where long polypeptide chains are tangled and sidely connected at various positions in a very complicated way.

We now turn to examine Huggins' model,[8] proposed and discussed in detail early in 1943, for the polypeptide structures of silk-fibroin, α- and β-keratins and collagen. In his structures Huggins assumed bond distances and angles based on Corey's data[9] of X-ray analyses for diketopiperazine,[10] glycine[11] and alanine.[12] However, in his model for α-keratin [Fig. 11 on p. 210, Fig. 14 (a), (b) and (c) on p. 212 of his paper], the calculated N···O distance of the N–H···O bond is 2.17 Å. (for the case of the identity distance 5.15 Å.), disagreeing with 2.85 Å. given in his Table 1 for this distance. If we stretch the chain to obtain a model for β-keratin until all bonds in the main chain come in one plane, the calculated identity distance becomes 7.48 Å., which is very much larger than the observed value 6.64 Å. In glycine and alanine molecules there are positive and negative charges and the N—C_α and C_α—C bond distances are apt to be influenced by these charges. In the diketopiperazine molecule several resonating structures are proposed. However, there is no charge in the polypeptide main chain, and the occurrence of resonance phenomena has not been confirmed. Therefore we cannot agree with Huggins in assuming the same bond distances and angles for fibrous proteins as occurring in simple molecules of glycine, alanine and diketopiperazine. Moreover, if we look into the stereochemical significance of C_α atom in his models [Fig. 11 and 12 on p. 210, Fig. 13 on p. 211, Fig. 14 (a), (b) and (c) on p. 212], we may easily verify that these models are constructed solely from D-acids. As a matter of fact, most of his models can as well be constructed from L- as from D-constituents. In this respect it may be mentioned that Simanouti and Mizushima's bent form model is also derivable even if we interchange the positions of R and H.

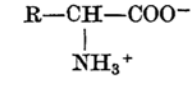
No explanation seems to have ever been offered as to why D-type amino acids do not exist in natural proteins. So long as we assume protein structures in which the positions of R and H attached to C_α are interchangeable, we should expect both D- and L-amino acid products in the hydrolyses of natural proteins. That this is not the case is shown by every evidence, and we are forced to consider that the true structures for proteins are those in which only one single type (L) of amino acids are contained as constituents.

From the above arguments it is clear that both Huggins' and Simanouti and Mizushima's models for keratins are not entirely satisfactory. We are going to propose, in this and in the succeeding paper, new models for α-, β-, and feather-keratins and silk-fibroins which appear more reasonable than those of the above authors, and in this paper we shall limit ourselves to the treatment of the structure of α-keratin.

II. Amino Acids.—We shall describe here briefly about amino acids constituting polypeptide chains of proteins from the structural-chemical standpoint. Amino acids obtained by hydrolysis of natural proteins are all α-amino acids and their general formula is given by

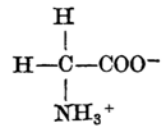


but in accordance with the X-ray diffraction results of crystals of glycine[11] and alanine,[12] we had better write the formula as

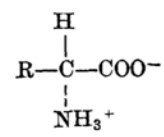


Amino acids are classified into three kinds.

(1) When R=H, this is optically inactive glycine. Two H atoms are attached to the C_α atom and three to the N atom.

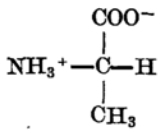


(2) When R=CH_3, C_2H_5, etc. (about 27 radicals), the acids are all optically active. Here we naturally expect the existence of D- and L-type antipodes, whereas α-amino acids obtained by hydrolysis of natural proteins are all of the L-type. By the studies of Freudenberg,[13] Kuhn,[14] etc. the absolute configuration of L-alanine is concluded, by analogy to the configuration of L-lactic acid, to be the one shown in the next figure. Thick horizontal bonds point to the front of the paper and thin vertical bonds

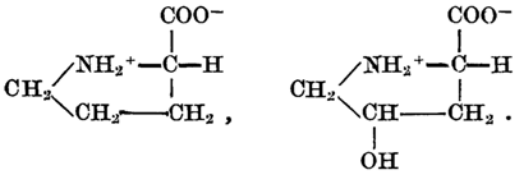


(8) M. L. Huggins, *Chem. Rev.*, **32**, 195 (1943).
(9) R. B. Corey, *Chem. Rev.*, **26**, 227 (1940).
(10) R. B. Corey, *J. Am. Chem. Soc.*, **60**, 1598 (1938).
(11) G. Albrecht and R. B. Corey, *J. Am. Chem. Soc.*, **61**, 1087 (1939).
(12) H. A. Levy and R. B. Corey, *J. Am. Chem. Soc.*, **63**, 2095 (1941).
(13) K. Freudenberg, "Stereochemie," 662 (1932).
(14) W. Kuhn, *Naturwiss.*, **26**, 289, 305 (1938).

point to the back of it. For amino acids other than L-alanine proper radicals are to be substituted in place of CH_3. In this configuration one H atom is connected to the C_α atom and three to the N atom.



(3) When R becomes longer and is connected to the N atom to form a ring, we



obtain proline and hydroxyproline. Here one H atom is attached to the C_α and two to the N atom.

III. Polypeptide Chain.—In fibrous proteins it is considered that α-amino acids constitute the polypeptide chain, after dehydration in the interaction of the COO^- radical of an amino acid molecule and the NH_3^+ radical of another.

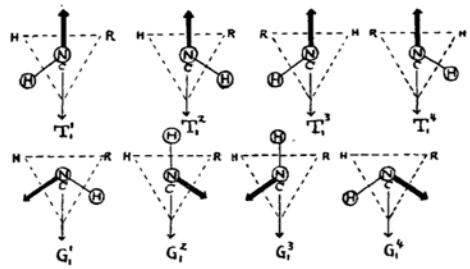


Fig. 1.—Bond direction types in N—C_α

In the polypeptide chain thus formed we shall assume for the various bond lengths and angles the values given in Table 1. The intramolecular rotation about the C-C single bond in dihalogenoethane, CH_2X-CH_2X, has been studied by Mizushima, Morino, *et al.*[13] In analogy to their papers, we can consider a great many possibilities of bond orientations in our polypeptide chain. In our case, however, the reduced symmetry makes a rigorous treatment extremely difficult, so we may, at least qualitatively, assume the bond directions shown in Figs. 1, 2 and 3 as the most probable. The N or C atom enclosed by a circle is the atom nearer to the reader in a N—C_α or a C_α—C bond seen in its direction, and its other bond which will be used in the chain formation is shown by a thick arrow. In a T-type configuration the main bonds (chain bonds) are in the *trans* form. A G-type is obtained from a T-type by rotating the upper or the lower atom with its arrow-indicated bond through a proper angles ($< 180°$) about the central bond as an axis in either direction. In our figures we show the cases of the upper atom turned. For the N—C_α bond (Fig. 1) we obtain as the T-Type T_1^1 and T_1^2, but we must exclude T_1^3 and T_1^4 as they correspond to the cases of D-amino acids. Hereafter D-type bond directions shall all be omitted. For the G-type G_1^1 and G_1^2 are derived from T_1^1, and G_1^3 and G_1^4 from T_1^2. Fig. 2 shows bond directions of C_α and C atoms in a C_α—C bond. T-type: only T_2. G-types: turning the upper atom

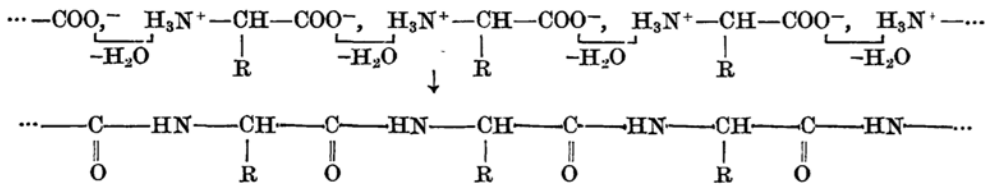


Table I

Bond Lengths and Bond angles*

Bond Lengths*		Bond Angles
N—C_α	1.47 Å.	
C_α—C	1.54 „	: Tetrahedral configuration.
C—N	1.47 „	
N—H	1.00 „	Each angle 109°28'
C_α—H	1.09 „	
C_α—R	1.54 „	: Plane triangular configuration.
C—O	1.22 „	Each Angle 120°
C_α : Carbon atom in the α-position.		: Pyramidal configuration.
C: Carbon atom in the carbonyl radical		Each Angle 109°

* L. Pauling: "The Nature of the Chemical Bond," p. 164 (1940).

through 120° we get G_2^1 and G_2^2, by 90° rotation G_2^3 and G_2^4, by 60° rotation G_2^5 and G_2^6. In Fig. 3 bond directions of C and N atoms in C—N are shown. T-types: T_3^1, T_3^2. G-types: turning the upper atom of T_3^1 type through 120° we get G_3^1 and G_3^2, by 90° rotation G_3^3 and G_3^4, by 60° rotation G_3^5 and G_3^6. Similarly we get G_3^7, G_3^8, G_3^9, G_3^{10}, G_3^{11} and G_3^{12} from the T_3^2 type.

IV. Polypeptide Structure of α-Keratin.—Protein of mammalian hair or wool is keratin. Important amino acids constituting it are leucine (11.5 g.), glutamic acid (12.9 g.), cystine (13.1 g.), arginine (10.2 g.) and so on (in 100 g. wool). In addition there is a little content of glycine (0.6 g.). Proline is also present in a relatively small amount (4.4 g.).[15] Therefore amino acids of the kind described in I (2) takes a greater part. An X-ray diffraction study of keratin was carried out by Astbury and Street.[1] Keratin in an ordinary state is α-keratin, and it has a rectangular cell (a=27 Å., b=10.3 Å., c=9.8 Å.). The half of b, i.e., 5.15 Å. is called the identity distance, and c=9.8 Å. is termed the side chain spacing.

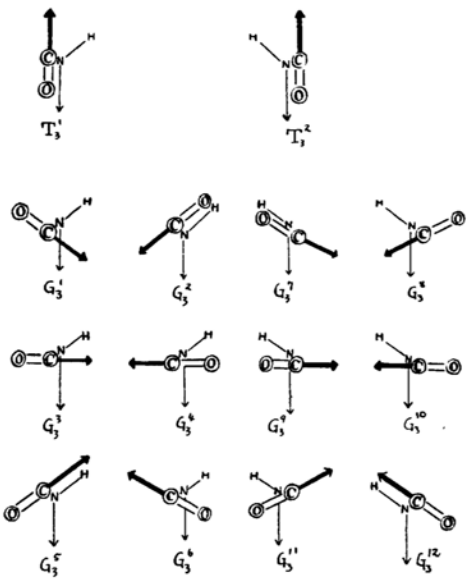


Fig. 3.—Bond direction types in C—N

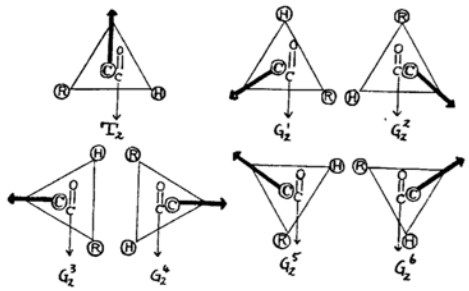


Fig. 2.—Bond direction types in C_α—C

The structure of the polypeptide chain of α-keratin must suffice all the conditions described in II (2) and III. Also it must explain the identity distance 5.15 Å., and at the same time must be arranged on the principle of no steric hindrance. Among many different combinations of bond directions shown in Figs. 1, 2 and 3 the only structure satisfying all of the above conditions is the one given in Fig. 4, in which the chain structure is shown projected on three planes. The main chain is not contained in one plane. Carbonyl radicals are attached alternately up and down. The C=O radical attached to the 1 or 7 C atom points down forwards, and that attached to the 4 or 10 C atom up backwards. As amino acids are all of the L-type, C_α—R points down backwards from C_α (6), and from C_α (3 or 9) it points up forwards. When C_α—H makes a

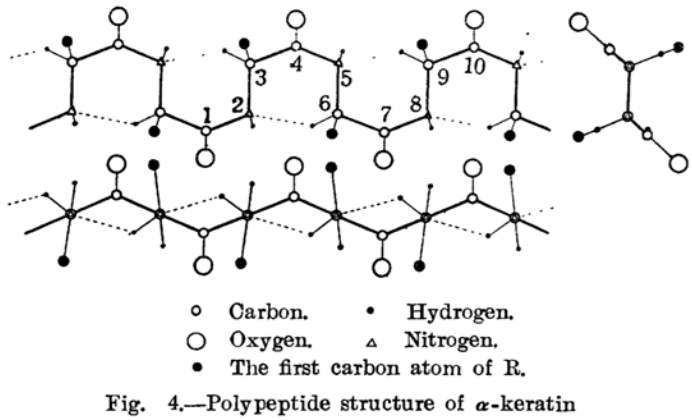


Fig. 4.—Polypeptide structure of α-keratin

Table 2

← C (1)— N (2) →	G_3^1
← N (2)— C_α (3) →	T_1^1
← C_α (1)— C (4) →	G_2^1
← C (4)— N (5) →	G_3^1
← N (5)— C_α (6) →	T_1^1
← C_α (6)— C (7) →	G_2^1
← C (7)— N (8) →	G_3^1
← N (8)— C_α (9) →	T_1^1
← C (9)— C (10) →	G_2^1

(15) C. Rimington, *Biochem. J.*, **23**, 41 (1929).

tetrahedral angle with C_α—R, an N atom approaches it and a C–H···N hydrogen bond is formed. In many literatures (for examples (3), (8), (16)), N–H···O hydrogen bonds are assumed as the intramolecular hydrogen bonds in α-keratin, but if that kind of hydrogen bonds were included in our prerequisites, we would obtain no stable main chain and the identity distance would become much smaller than 5.15 Å. Evidences that the C–H vibration frequency 2960 cm^{-1} is observed in the infrared absorption of proteins[8] seems unfavorable for our C–H···N bond. However, for this frequency we would rather assume that a C–H vibration in side chains is responsible, such as in the CH_3 radical of alanine residue. That N–H···O bonds actually are present in quantity in proteins has been confirmed by infrared studies,[8] but there seems to be no evidence that they must be contained in the main chain. Therefore we want to assume C–H···N hydrogen bonds in the polypeptide structure of α-keratin. The circumstance that a single H atom is attached to C_α in the α-amino acid molecule seems very favorable for the formation of C–H···N bonds.

Now let us look at the mutual orientation of various bonds in our main chain (Fig. 4). If we pick up two consecutive atoms arbitrarily in the chain and place the atom labeled with a smaller number in the position of the upper atom in one of the figures in Figs. 1, 2 and 3, we obtain Table 2. We see here only the series T_1^1—G_2^1—G_3^1 repeats. If we took G_1^1 for T_1^1, and some other types in places of G_2^1 and G_3^1, steric hindrances would come in, and even if no steric hindrance occurs the identity distance deviates from 5.15 Å. greatly. The calculated value of the identity distance of the structure shown in Fig. 4 is 5.20 Å., the C···N distance in C–H···N hydrogen bond 2.60 Å. and the N···H distance 1.93 Å.

Out of the bond direction types shown in Figs. 1, 2 and 3 those leading to steric hindrances, such as T_1^2 and G_3^7, must be excluded. All other types derived from these two, i.e., G_1^3, G_1^4, T_3^2, G_3^8, G_3^9, G_3^{10}, G_3^{11} and G_3^{12}, are excluded. Next, if we interchanged R and H in every amino acid constituent in our structure, we would obtain a chain from T_1^3 etc., which is made up solely of D-types. But obviously this structure cannot arise because of steric hindrances necessarily introduced. It can be concluded that the polypeptide chain of α-keratin derivable solely from L-amino acids must have the structure shown in our Fig. 4.

Naturally we might imagine a chain structure which is exactly a mirror image of the one shown in Fig. 4. It would be composed solely of D-amino acid constituents and yet has no steric hindrance. However, this structure is unnatural on the experimental ground repeatedly stated above, inasmuch as we exclude the chance that all the D-amino acids in proteins infallibly convert themselves into their antipodes in the process of hydrolytic decomposition.

Resumo.

(1) Aminoacidoj estas klasigitaj en tri klasojn.

(2) Pri polipeptida ĉeno valuoj de atomvalentaj longecoj kaj anguloj, kaj atomvalentaj direktoj montritaj en Tabelo 1, Fig. 1, 2 kaj 3 estas supozitaj.

(3) Nova polipeptida strukturo de α-keratino derivebla eksklusive el L-aminoacidoj estas proponita, en kiu nur la uniĝtipa serio T_1^1– G_2^1– G_3^1 ripetas, kaj samtempe tre probabla apero de C–H···N hidrogena uniĝo en la polipeptida ĉeno de α-keratino estas emfazita.

In conclusion, the author wishes to express his hearty thanks to Prof. S. Imanishi for the kind guidance and valuable discussions, and also to the Ministry of Education for grants in aid out of the Scientific Research Expenditure.

Chemical Institute, Faculty of Science, Kyūsyū University, Fukuoka.

(16) W. T. Astbury, *Trans. Faraday Soc.*, **36**, 871 (1940); A. E. Alexander and E. K. Rideal, *Nature*, **147**, 541 (1941).