

REFERENCES

- Alfrey, T., Agron, P., Bohner, J., Haas, H. & Wechsler, H. 1948 *J. Poly. Sci.* **3**, 157.
Alfrey, T. & Goldfinger, G. 1944 *J. Chem. Phys.* **12**, 205, 322.
Alfrey, T. & Price, C. C. 1947 *J. Poly. Sci.* **2**, 101.
Arlman, E. J., Melville, H. W. & Valentine, L. 1949 *Rec. Trav. chim. Pays-Bas*, **68**, 945.
Bickel, A. F. & Melville, H. W. 1949 *Trans. Faraday Soc.* **45**, 1049.
Burnett, G. M., Valentine, L. & Melville, H. W. 1949 *Trans. Faraday Soc.* **45**, 960.
Cohen, S. G., Ostberg, B. E., Sparrow, D. B. & Blout, E. R. 1948 *J. Poly. Sci.* **3**, 269.
Degens, P. N. & Gouverneur, P. 1950 *Anal. Chim. Act.* (in the Press).
Grassie, N. & Melville, H. W. 1950 *J. Poly. Sci.* (in the Press).
Masson, C. R. & Melville, H. W. 1949 *J. Poly. Sci.* **4**, 337.
Mayo, F. R. & Lewis, F. M. 1944 *J. Amer. Chem. Soc.* **66**, 1594.
Mayo, F. R., Lewis, F. M. & Walling, C. 1947 *Disc. Faraday Soc.* No. 2, 287.
Melville, H. W., Noble, B. & Watson, W. F. 1947 *J. Poly. Sci.* **2**, 229.
Melville, H. W. & Valentine, L. 1950a *Proc. Roy. Soc. A*, **200**, 353.
Melville, H. W. & Valentine, L. 1950b *Trans. Faraday Soc.* **46**, 210.
Price, C. C. 1947 *Disc. Faraday Soc.* No. 2, 307.
Rodebush, W. H. & Feldman, I. 1946 *J. Amer. Chem. Soc.* **68**, 897.
Schütze, M. 1939 *Z. anal. Chem.* **118**, 241.
Simha, R. & Wall, L. A. 1948 *Bur. Stand. J. Res., Wash.*, **41**, 521.
Smakula, A. 1934 *Z. angew. Chem.* **47**, 777.
Unterzaucher, J. 1940 *Ber. dtsh. Chem. Ges.* **73B**, 391.
Walling, C. 1949 *J. Amer. Chem. Soc.* **71**, 1930.

Polypeptide chain configurations in crystalline proteins

BY SIR LAWRENCE BRAGG, F.R.S., J. C. KENDREW AND M. F. PERUTZ

Cavendish Laboratory, University of Cambridge

(Received 31 March 1950)

Astbury's studies of α -keratin, and X-ray studies of crystalline haemoglobin and myoglobin by Perutz and Kendrew, agree in indicating some form of folded polypeptide chain which has a repeat distance of about 5.1 Å, with three amino-acid residues per repeat. In this paper a systematic survey has been made of chain models which conform to established bond lengths and angles, and which are held in a folded form by N—H—O bonds. After excluding the models which depart widely from the observed repeat distance and number of residues per repeat, an attempt is made to reduce the number of possibilities still further by comparing vector diagrams of the models with Patterson projections based on the X-ray data. When this comparison is made for two-dimensional Patterson projections on a plane at right angles to the chain, the evidence favours chains of the general type proposed for α -keratin by Astbury. These chains have a dyad axis with six residues in a repeat distance of 10.2 Å, and are composed of approximately coplanar folds. As a further test, these chains are placed in the myoglobin structure, and a comparison is made between calculated and observed F values for a zone parallel to the chains; the agreement is remarkably close taking into account the omission from the calculations of the unknown effect of the side-chains. On the other hand, a study of the three-dimensional Patterson of haemoglobin shows how cautious one must be in accepting this agreement as significant. Successive portions of the rod of high vector density which has been supposed to represent the chains give widely different projections and show no evidence of a dyad axis.

The evidence is still too slender for definite conclusions to be drawn, but it indicates that a further intensive study of these proteins, and in particular of myoglobin which has promising features of simplicity, may lead to a determination of the chain structure.

1. INTRODUCTION

Proteins are built of long chains of amino-acid residues. Amino-acids unite to form a chain (figure 1) in which the R groups are of some twenty-three different kinds, varying from a hydrogen atom in glycine to moderately complex groups (e.g. the linked five- and six-membered rings of tryptophan). As examples, insulin of molecular weight 12,000 has, according to Chibnall (1945), 106 residues, in four chains linked by six disulphide bridges of cystine (Sanger 1948). Myoglobin, 17,000, has about 146 residues; haemoglobin, 67,000, has about 580 residues.

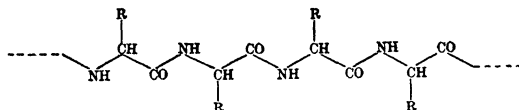


FIGURE 1. Polypeptide chain.

In this paper an attempt is made to glean as much information as possible about the nature of the chains from X-ray studies of crystalline proteins, and to survey the possible types of chain which are consistent with such evidence as is available. The configuration and arrangement of the R groups is not discussed; we are merely concerned with the configuration of the $-\text{CO}-\text{CH}-\text{NH}-$ chain to which they are attached.

Certain features of all protein structures deserve special mention.

(a) The mean molecular weight of the residue $\begin{array}{c} \text{R} \\ | \\ \text{NH}-\text{CH}-\text{CO} \end{array}$ is much the same in most proteins, ranging between 110 and 120 (e.g. haemoglobin 112.5). Since the molecular weight of the chain element $-\text{CO}-\text{CH}-\text{NH}-$ is 56, one-half the weight of the protein is in the chain and one-half in the side groups, excluding associated water. To put this in another way, the average side-chain contains about four atoms other than hydrogen.

(b) All the amino-acids occurring in proteins (except glycine which is non-enantiomorphous, and a very few amino-acids found in primitive organisms such as bacteria) have the steric configuration about the central carbon atom conventionally termed 'laevo'.

(c) X-ray determinations of the structures of simple amino-acids or dipeptides yield consistent information about the interatomic distances and bond angles in their crystals, and presumably these angles and distances will not be very different in a long polypeptide chain.

2. PREVIOUS SPECULATIONS ABOUT THE CONFIGURATION OF THE POLYPEPTIDE CHAIN

Astbury and his co-workers in their pioneer investigations have made an exhaustive study of the fibrous proteins such as the keratin of hair and wool. Their most important result, in the present connexion, is their inference that the marked

5.1 Å repeat along the fibre axis which is shown prominently by X-ray photographs of α -keratin and its analogues corresponds to an element of folded chain containing three amino-acid residues. Briefly, Astbury (private communication, 1949) summarizes the evidence as follows:

(a) The β -keratin X-ray diagram with its strong meridional 3.4 Å reflexion represents a system of extended polypeptide chains as in figure 2(b), for which the average length per residue must be about 3.4 Å to correspond to the accepted bond lengths and angles.

(b) The β configuration is approximately twice as long as the folded α configuration, since the reversible extensibility of wool and hair is approximately to twice the normal length. The average length per residue in the α chain is therefore 1.7 Å, giving three residues per repeat of 5.1 Å.

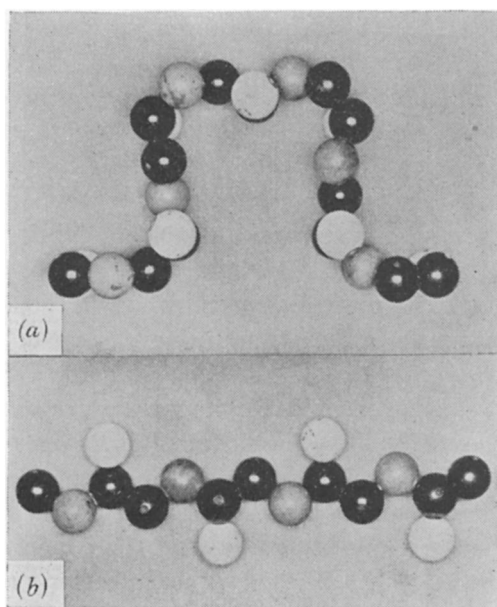


FIGURE 2. Chain configurations proposed by Astbury (1949a) for (a) α -keratin, (b) β -keratin.

The model of the folded chain in figure 2 (a) is proposed for α -keratin by Astbury as that which at present seems to fit the facts best, other indications in favour of this structure being obtained from considerations of side chain packing. The complete repeat is at distances of 10.2 Å, each loop occupying 5.1 Å and containing three residues as indicated in the diagram. It will be referred to again below.

Huggins (1943) has made an extensive review of possible types of polypeptide chains. He gives numerous examples of chains in both extended and shrunk forms, and of ways in which these may be linked together, showing how many the possibilities are until further evidence is obtained which restricts them.

Recently a type of chain has been proposed by Ambrose & Hanby (1949; see also Ambrose, Elliott & Temple 1949) based on measurements of the dichroism of

α -keratin, myosin and tropomyosin in the infra-red. This is a folded chain with two residues in a repeat of 5.1 Å, and Ambrose and his collaborators picture these chains as occurring in pairs so that there are four residues in the α -keratin unit.

3. POLYPEPTIDE CHAINS IN CRYSTALLINE HAEMOGLOBIN AND MYOGLOBIN

The X-ray diffraction pictures given by haemoglobin, in particular horse methaemoglobin, have been studied in detail by Perutz and his collaborators (Boyes-Watson, Davidson & Perutz 1947; Perutz 1949), and Kendrew (1950) has published an account of work with myoglobin. The evidence for the existence and features of chains in haemoglobin and myoglobin has been very fully discussed in these papers, but it may be useful to recapitulate here the nature of the problem and the way in which an attack on it has been made. Protein crystals give a wealth of diffracted beams, extending out to angles which correspond to spacings as small as 2 Å, but the interpretation of these photographs as a complete picture of the arrangement of the thousands of atoms in the unit cell is a task greater by several orders of complexity than the most complex crystals yet successfully analyzed. It is always possible, however, to represent the experimental observations as a Patterson or 'vector' map. Broadly speaking, if there are in the actual crystal two atoms a and b with 'weights' m_a and m_b , the Patterson map has a peak or lump of density proportional to the product $m_a m_b$ at a point ' ab ' such that the line joining ' ab ' to the origin of the map is equal and parallel to the line drawn between a and b in the crystal. If the crystal contains n atoms, there are n^2 such points in the Patterson map corresponding to the product $(a + b + c + d + \dots)(a + b + c + d + \dots)$. The n terms $a^2, b^2, c^2, d^2, \dots$ have zero vectors and superimpose at the origin to form a large concentration. The remainder are distributed throughout the cell. (The meaning of Patterson maps is explained in greater detail by Kendrew & Perutz (1949).) It is common practice in crystal analysis to start by forming two- or three-dimensional Patterson summations and then to seek to decipher the significance of their more prominent peaks in terms of the atomic arrangement. However, the Patterson projection of a protein represents some millions of vector peaks. It can only be deciphered if there is some regular underlying arrangement throughout the crystal structure, such as would be the case if the polypeptide chains are regular in form, straight, and parallel to each other in the molecule. It has already been remarked that half the atoms (hydrogen can be neglected on account of its low scattering power) are in the chain element CO—C—N. Further, if some model of the chain is adopted, the first carbon atom of each side-chain is also in a definite position, so that some five-eighths of all the atoms form part of a regularly recurring pattern. The remainder are in the side groups and will have a wide variety of configurations. The vectors drawn between atoms inside any one chain, and those between its atoms and the atoms of neighbouring parallel chains, will form a large proportion of all the vectors in the Patterson and will be of a relatively few constantly repeated types. Owing to this regularity they may be prominent in the Patterson projection in spite of the irregular welter of other vectors on which they are superimposed. This possibility is the greater because the parallel straight chains would be regions of relatively high density. The atoms in any one

4. CLASSIFICATION OF CHAIN STRUCTURES

In this and the following sections we attempt to survey systematically all those types of folded polypeptide chain configurations which satisfy certain conditions, established by experiment or plausible on general grounds.

It cannot be assumed as certain that the polypeptide chain has the same configuration in all crystalline proteins, or that a similar configuration occurs in fibrous proteins such as α -keratin. It is, however, not unreasonable to expect that haemoglobin and myoglobin contain chains of the same type, because these proteins appear to be closely related in several ways; and furthermore, the repeat distance, the inter-chain distance, and the number of residues per repeat, are similar in these two proteins to the corresponding features of α -keratin. It will therefore be assumed as a working hypothesis that the chain configurations in large classes of proteins resemble one another closely, while bearing in mind that this hypothesis is based on slender evidence and may have to be abandoned when further experimental data are available.

In our survey of chain configurations we have adopted the following conditions:

(a) Interatomic distances and bond angles

Some relevant data have been obtained by X-ray analysis of the structures of amino-acids and small peptides. Glycine has been analyzed by Albrecht & Corey (1939), DL-alanine by Levy & Corey (1941) and β -glycylglycine by Hughes & Moore (1949). An addition compound of cysteyl-glycine has recently been analyzed in the Cavendish Laboratory by Dyer, and progress has been made with the analysis of tripeptides and tetrapeptides.

In an earlier paper Huggins (1943) has given the following summary of the

distances and angles to be expected in the chain unit

$$\begin{array}{c} \text{H} \quad \text{O} \\ | \quad || \\ -\text{N}-\text{C}-\text{C}'- \\ | \quad | \\ \text{H} \quad \text{R} \end{array}$$

	(Å)		(°)
N—C	1.41	$\widehat{\text{NCC}'}$	112
C—C'	1.52	$\widehat{\text{CC}'\text{N}}$	118
C'—O	1.25	$\widehat{\text{C}'\text{NC}}$	118
C'—N	1.33		

The experimental results, including the most recent, agree generally with these figures, though with some individual variations (for summary see Corey 1948). For

example, in cysteyl-glycine Dyer has found $\widehat{\text{C}'\text{NC}}$ considerably greater than the figure given. All the structures contain hydrogen bonds N—H...O which vary in length between 2.6 and 2.85 Å.

For the construction of models we have adopted the following values, which agree with those given by Huggins or with generally recognized standards:

- (i) Covalent distances: C—C, 1.52 Å; C—N, 1.36 Å; C—O, 1.24 Å.
(ii) Hydrogen bond distances: N—O (in N—H...O) 2.85 Å.
(iii) Bond angles: C: tetrahedral distribution.
N: tetrahedral (interbond angle 109° 28') or planar (120°).*
N—H...O=C<: we have placed no restriction on $\widehat{\text{NHO}}$ or on $\widehat{\text{HOC}}$, though we have generally attempted to make NHOC as nearly collinear as possible.

(b) *Optical configuration of amino-acids*

We have assumed that all the amino-acids (except glycine which is optically inactive) belong to the *laevo* series.

(c) *Symmetry of the chain*

Huggins (1943) has made the following two general points about the symmetry of a stable chain configuration:

(i) '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 or the one that he has most recently proposed for collagen, would tend to bend it continuously in the same direction.'

(ii) '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.'

We have accepted Huggins's first criterion throughout; indeed, we have used it as a basis for classifying types of configuration. In other words, all the structures examined possess a screw axis of symmetry (not necessarily restricted to crystallographic types of screw symmetry, e.g. a fivefold axis would be permissible).

Huggins's second criterion, that each element of the chain should be in a similar relation to neighbouring elements, we have not regarded as essential, and some of the chains described below do not obey this rule.

(d) *The role of hydrogen bonding*

We have made the plausible (but still unproved) assumption that the chain is held in a folded condition by hydrogen bonds between >NH and >CO groups of nearby amino-acid residues. In other words, the folded chain is thrown into a series of rings which must be ruptured at the hydrogen bond before unfolding can take place. Formally speaking, these rings may be of two types, illustrated diagrammatically in figure 4(a).

* No complete structure analysis of any compound containing nitrogen bound analogously to nitrogen in proteins of the α -keratin type is yet available. We have therefore thought it better to leave this question open for the present, and to examine, in each type of structure, the effect of the two configurations given.

It will be evident that the number of configurations formally possible is rather large, and even though many of the formal schemes are found on examination not to be sterically possible it is still true that the number of configurations which should be examined in a comprehensive survey is considerable. We regard those structures in which all NH and CO groups are hydrogen bonded ($q = 1$) as inherently the more probable, because their free energy is presumably lower. We have therefore examined structures of this type in considerable detail and have made efforts to build all possible models provided they conformed to the conditions outlined above.

We have also tried to build all the possible structures having only a fraction of the NH and CO groups hydrogen bonded, determining at least the repeat distance of the molecular pattern and the number of amino-acid residues which that pattern contains. This survey was less thorough, so that some possible structures may have been overlooked, and only a few of the structures were examined in detail. Among the latter was $2_{13} \cdot \frac{1}{3}$ because of the evidence which Astbury produced in its favour, and $2_{14} \cdot \frac{1}{3}$ because of its close similarity with $2_{13} \cdot \frac{1}{3}$.

5. STRUCTURES EXAMINED IN DETAIL

Within the limits discussed in the last two paragraphs we have attempted to make models of as many types of structure as possible, considering in turn each type of chain symmetry and each ring size. In some instances no structure could be devised for steric reasons. In general, very small or very large rings may be excluded; small rings on account of the strains involved in closing them, large ones because they involve either excessively long hydrogen bonds or else chains of very large cross-sectional area containing many residues per repeating unit.

These structures examined in detail are listed in table 1, classified in accordance with their symmetry and ring size. It is to be noted that a given chain configuration can in general give rise to two types of structure depending on the way the side groups are attached, since the chain itself is enantiomorphous; broadly speaking, these may be distinguished as *laevo* groups in a left-handed chain, or *laevo* groups in a right-handed chain—structures which are *not* mirror images of each other and which are designated by suffixes 'a' and 'b'. In some cases steric considerations rule out one of these configurations.

We now proceed to discuss briefly each of the structures listed in table 1, leaving over for more detailed treatment later those which we regard as displaying the most promising agreement with the experimental data. Those not further considered have generally been excluded because they do not satisfy one or both of the following criteria:

(a) *Repeat distance.* The structure must account for an apparent 'repeat' distance of just over 5 Å. This 'repeat' may be the true repeat distance between identical points, or a fractional translation produced by the action of the screw axis.

(b) *Number of residues per repeat.* We have preferred structures in which the repeat (or pseudo-repeat) contains three amino-acid residues, for the reasons given above (p. 325). We do not feel, however, that we can entirely exclude the possibility that the repeat contains four residues, and two such structures are considered in

TABLE 1

screw axis of symmetry	no. of atoms in ring	repeat distances (Å)	no. of residues per repeat	illustrations (figure no.)	comments
twofold	7 <i>a</i>	5-5.6	2	5	structure proposed by Huggins (1943)
	7 <i>b</i>	5-5.6	2	6	structure proposed by Zahn (1947) and Ambrose <i>et al.</i> (1949); readily folds in pairs; see §7, 8
	8	4.6-4.8	2	7	structure proposed by Huggins (1943); only one configuration possible; repeat distance too short
	13	10.2	6	8	structure proposed by Astbury & Bell (1941); see §7, 8
	14	10.2	6	9	see §7, 8
threefold	7	7.5	3	—	repeat distance too short
	8	5.4	3	10	hydrogen bonds mutually perpendicular; see §7
	10	5.2	3	11	structure proposed by Taylor (1941) and Huggins (1943); hydrogen bonds oriented nearly parallel to the chain direction; see §8
fourfold	11	—	—	—	} no possible structures
	13	—	—	—	
	14	—	—	—	
	7	—	—	—	} no possible structures
	8	—	—	—	
	10	—	—	—	
	11	5.4	4	—	
13	5.6	4	12	a possible structure	
14 or greater	—	—	—	no possible structures	
fivefold and higher symmetries	—	—	—	—	all such structures contain more than four amino-acid residues per repeat unit

detail (2_{7*a*} and 4₁₃). It appears highly improbable that the number of residues in the 5 Å repeat is greater than four, and this criterion excludes symmetries which are more than fourfold; a fivefold chain repeating at 5 Å, for example, would necessarily contain at least five amino-acid residues per repeat.

The following diagrams illustrating the types of chain were drawn from optical projections of models, and are only intended to illustrate the structure of the chain. The co-ordinates of the atoms given in the accompanying tables are accurate to about 0.1 Å.

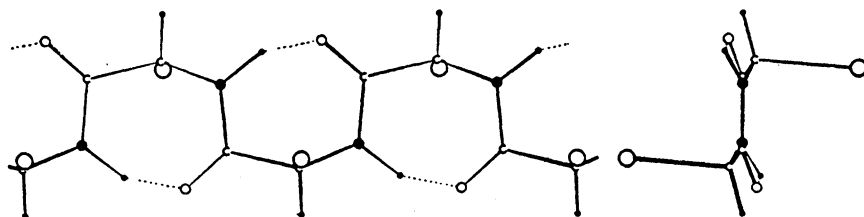
(i) *Twofold structures*

*Seven-atom ring (2_{7*a*}, 2_{7*b*})*

One form of this ring (2_{7*a*}) was proposed by Huggins (1943), whose drawings of it are reproduced in figure 5. The chain itself forms a more or less flat ribbon while R groups project alternately up and down from the plane of the ribbon. The repeat distance is 5 to 5.6 Å and contains two residues.*

* The precise repeat distances in 2₇ and 2₈ cannot be predicted since the N—H—O—C bonds deviate widely from a straight line, and precise experimental evidence is not yet available for the N—O distance in such a case.

The other form (2_7b), differing only in that the R groups and H atoms on the α -carbon atoms are interchanged, was suggested by Zahn (1947) and described in more detail by Ambrose & Hanby (1949) and Ambrose *et al.* (1949). As shown in figure 6 the R groups now lie approximately in the plane of the ribbon, instead of normal to it. This is an attractive feature of the structure, since it enables the chain



αC $\bullet N$ $\circ O$ $\circ R$ $\bullet H$

FIGURE 5. Projections of 2_7a chain, after Huggins (1943). Repeat distance about 5 Å

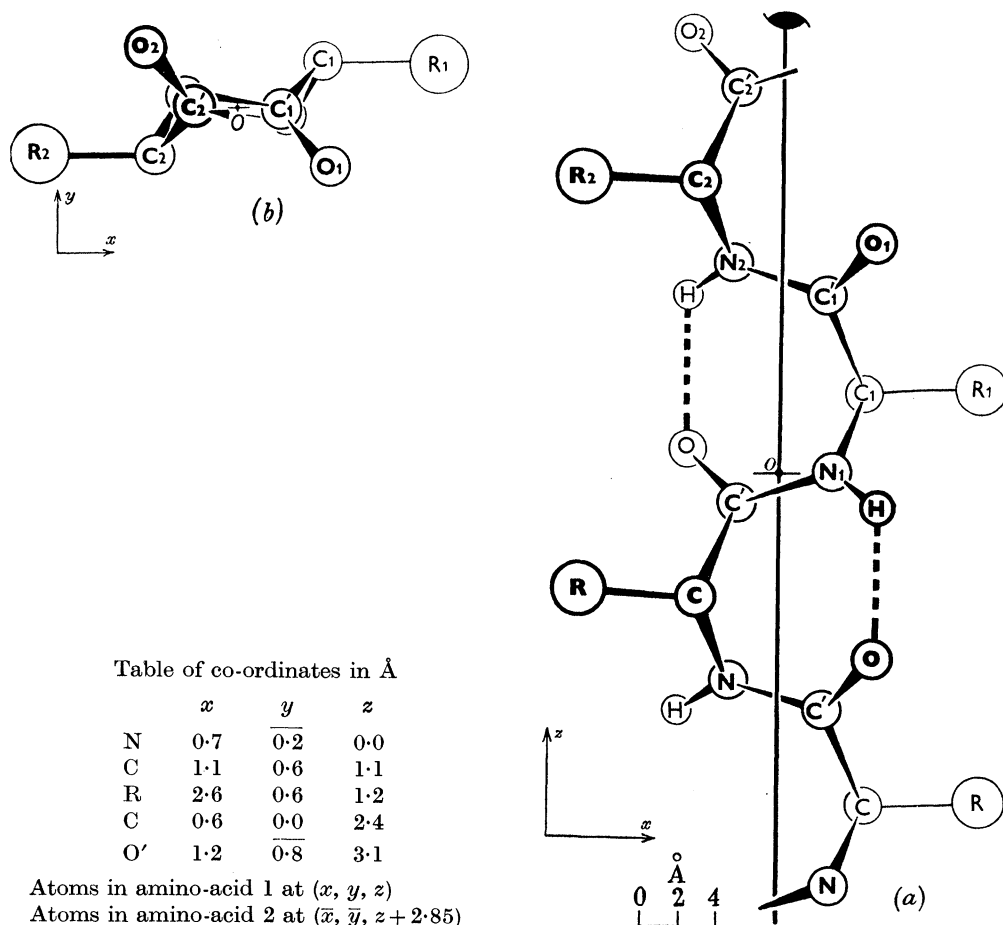


Table of co-ordinates in Å

	x	y	z
N	0.7	0.2	0.0
C	1.1	0.6	1.1
R	2.6	0.6	1.2
C	0.6	0.0	2.4
O'	1.2	0.8	3.1

Atoms in amino-acid 1 at (x, y, z)

Atoms in amino-acid 2 at ($\bar{x}, \bar{y}, z + 2.85$)

FIGURE 6. Projections of 2_7b chain.

to bend back on itself about an axis in the ribbon plane, without mutual interference of R groups or rupture of hydrogen bonds; the two limbs then lie about 4 Å apart.

Astbury (1949*b*) has criticized 2_7b on several grounds, one being that a chain having a true repeat of 5 Å and a screw dyad axis would give a weak 010 (5 Å) and a strong 020 (2.5 Å) reflexion, whereas in fact in α -keratin it is the 5 Å meridional reflexion which is overwhelmingly strong. This objection would apply to both forms of 2_7 , and to overcome it Huggins (1943) introduced 'the additional assumption that alternate R groups are much more potent X-ray scatterers than the intermediate ones'. There is, however, no other evidence that such an arrangement exists.

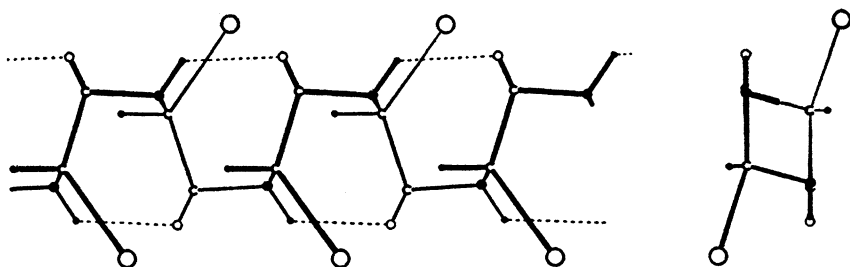


FIGURE 7. Projections of 2_8 chain after Huggins (1943). Repeat distance about 5 Å. Symbols as for figure 5.

2_7b has also been criticized by Darmon & Sutherland (1949) on grounds connected with the interpretation of infra-red absorption spectra given by Ambrose *et al.* (1949).

Another serious objection to both forms of 2_7 is that it contains only two amino-acid residues per repeat of 5.5 Å. Alternatively, 2_7b can be arranged so that two close-spaced chains (see above) run through a single subcell of $5 \times 10 \times 10$ Å, giving *four* residues per subcell. We have given reasons for preferring models containing *three* residues per subcell; but these reasons are not conclusive ones, and so we have given more detailed consideration (see below) to 2_7b , in particular in relation to haemoglobin and myoglobin.

Both structures can be built using either a planar or a tetrahedral configuration of bonds about the nitrogen atom. We have preferred a planar configuration, since this gives a more nearly linear arrangement of the hydrogen bond N—H—O, and because the ring is flatter than with the tetrahedral configuration, which prevents the packing of neighbouring chains to as close a distance as the myoglobin data would appear to demand (see p. 350).

Eight-atom ring (2_8)

This model was first suggested by Huggins (1943), whose projections of it are reproduced in figure 7. In the configuration of its R groups it resembles 2_7b rather than 2_7a ; no second form, corresponding to 2_7a , is possible for steric reasons; in such a structure R would lie far too close to C'.

The 2_8 structure is not immediately attractive since, like 2_7 , it contains only two residues in the repeat distance and has other disadvantages not shared by 2_7 :

(a) The repeat distance is distinctly less than 5 Å; 4.6 with planar, and 4.7 to 4.8 Å with tetrahedral nitrogen atoms.

(b) The angle HOC in the hydrogen bonds is only about 100° , which seems unlikely on general grounds, since it brings the N and O atoms very close to one another.

(c) The whole chain is much more rigid than 2_7b , leaving no possibility of folding about an axis in the plane of the ribbon. In any case the side chains would interfere with such a fold.

For all these reasons we regard 2_8 as an unlikely structure, and do not consider it further.

Thirteen-atom ring ($2_{13} \cdot \frac{1}{3}$)

This is the well-known structure proposed for α -keratin by Astbury & Bell (1941). We have constructed models in conformity with the diagrams published by these authors; our projections are illustrated in figure 8. The repeat distance is just more than 10 \AA , and contains six amino-acid residues; side chains project alternately up and down. The structure satisfies all the experimental data and is discussed in more detail below.

We found it impossible to construct a $2_{13} \cdot \frac{1}{3} \cdot b$ configuration. On the other hand, we obtained a quite distinct version of $2_{13} \cdot \frac{1}{3}$ in which the side-chains project alternately up and down as in the Astbury version, and the repeat distance is 9 \AA , containing six amino-acid residues. The ring is somewhat strained and the structure seems in general an unlikely one; we mention it merely as an illustration of the fact, which we have observed here and in $2_{14} \cdot \frac{1}{3}$, that these large ring structures have considerable flexibility and their detailed configurations can be altered within wide limits. At the present stage there is clearly nothing to be gained by studying all such minor variations in detail.

Fourteen-atom ring ($2_{14} \cdot \frac{1}{3}$)

This structure, which we believe has not been previously described, is illustrated in figure 9. It possesses all the attractive features of $2_{13} \cdot \frac{1}{3}$: the ring is not strained and is almost planar; R groups project alternately up and down from the ring plane (though they are not close-packed, their distances apart varying from 4 to 6 \AA); the repeat distance is 10.2 \AA and contains six residues. Accordingly, we discuss the structure in detail below.

Both $2_{14} \cdot \frac{1}{3} \cdot a$ and $2_{14} \cdot \frac{1}{3} \cdot b$ are possible, and each is susceptible of minor variations, but the ring is so flexible that we have only thought it worth while to discuss one typical, and on general grounds highly probable, structure.

(ii) *Threefold structures*

In all these structures (and those of higher symmetry) the R group and H atom on the α -carbon atoms can easily be interchanged, with little effect on the general properties of the structure. We have in each case considered the version in which the side-chains would protrude most nearly normal to the axis of symmetry.

Seven-atom ring (3_7)

This structure would appear to be possible sterically, and we constructed a model without difficulty. We have not considered it further, since its repeat distance is 7.5 \AA (containing three amino-acid residues).

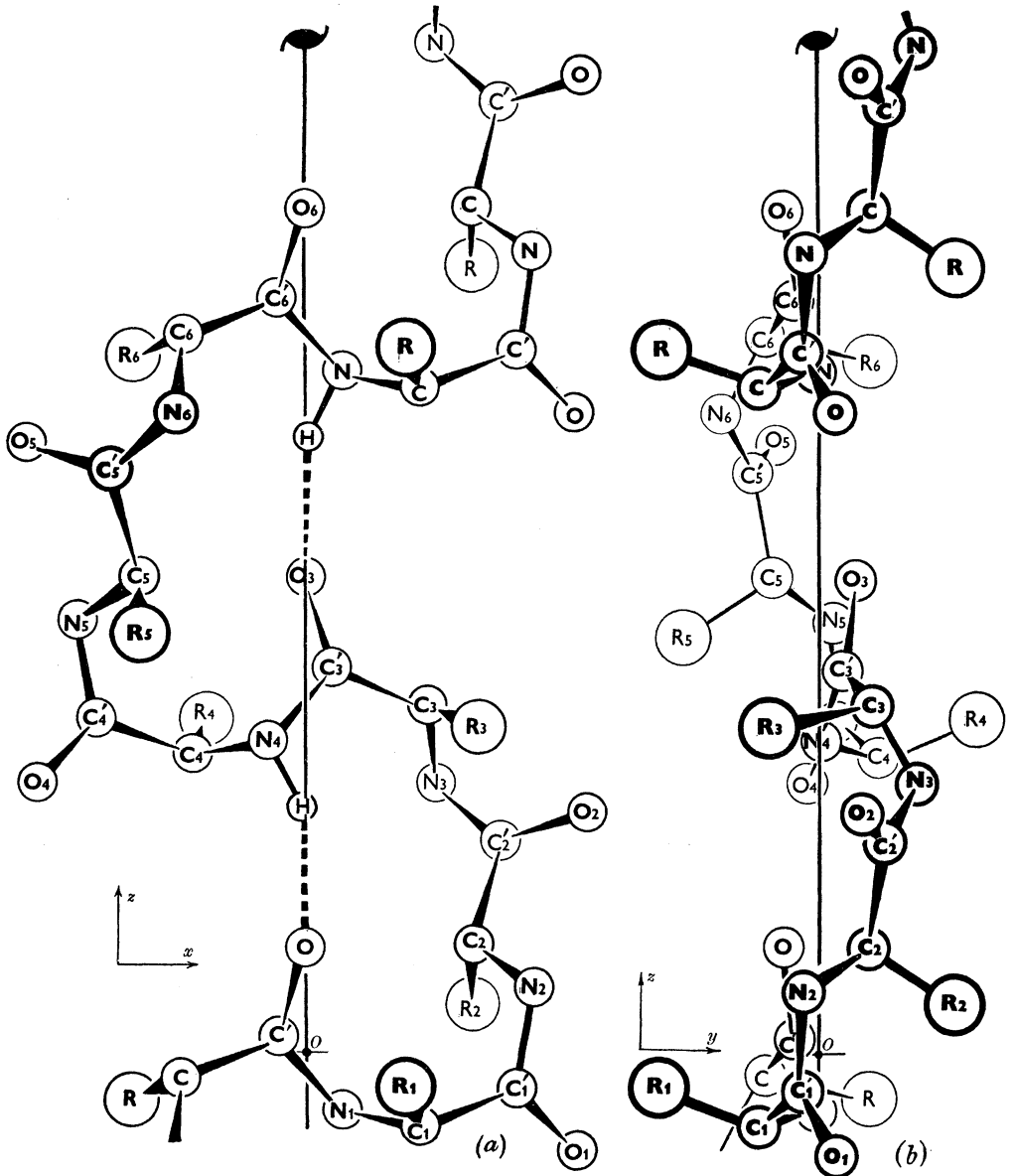


Table of co-ordinates in Å

	x	y	z
N_1	0.5	0.0	0.8
C_1	1.6	0.8	1.0
R_1	1.4	2.2	0.5
C'_1	2.9	0.2	0.5
O_1	3.7	0.2	1.4
N_2	3.1	0.2	0.9
C_2	2.3	0.7	1.5
R_2	2.3	1.9	0.7
C'_2	2.7	0.9	2.9
O_2	3.8	0.6	3.3
N_3	1.8	1.4	3.7
C_3	1.7	0.7	4.8
R_3	2.4	0.7	4.5
C'_3	0.4	0.3	5.3
O_3	0.0	0.5	6.5

Atoms in amino-acid residues 4, 5, 6 at $(\bar{x}, \bar{y}, z \pm 5.1)$

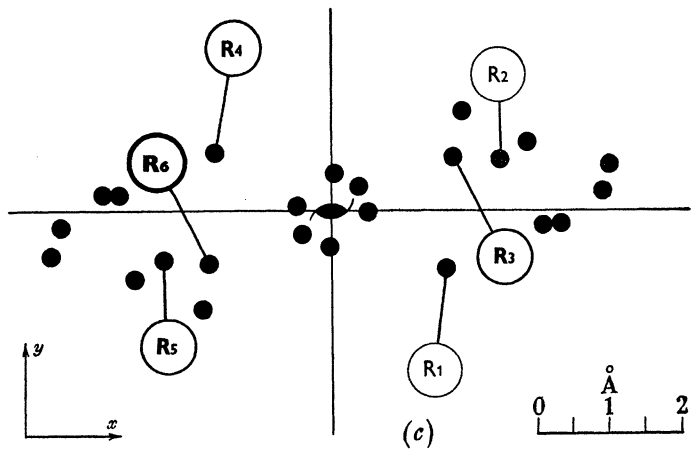


FIGURE 8. Projections of $2_{13} \cdot \frac{1}{2}$ (Astbury) chain.

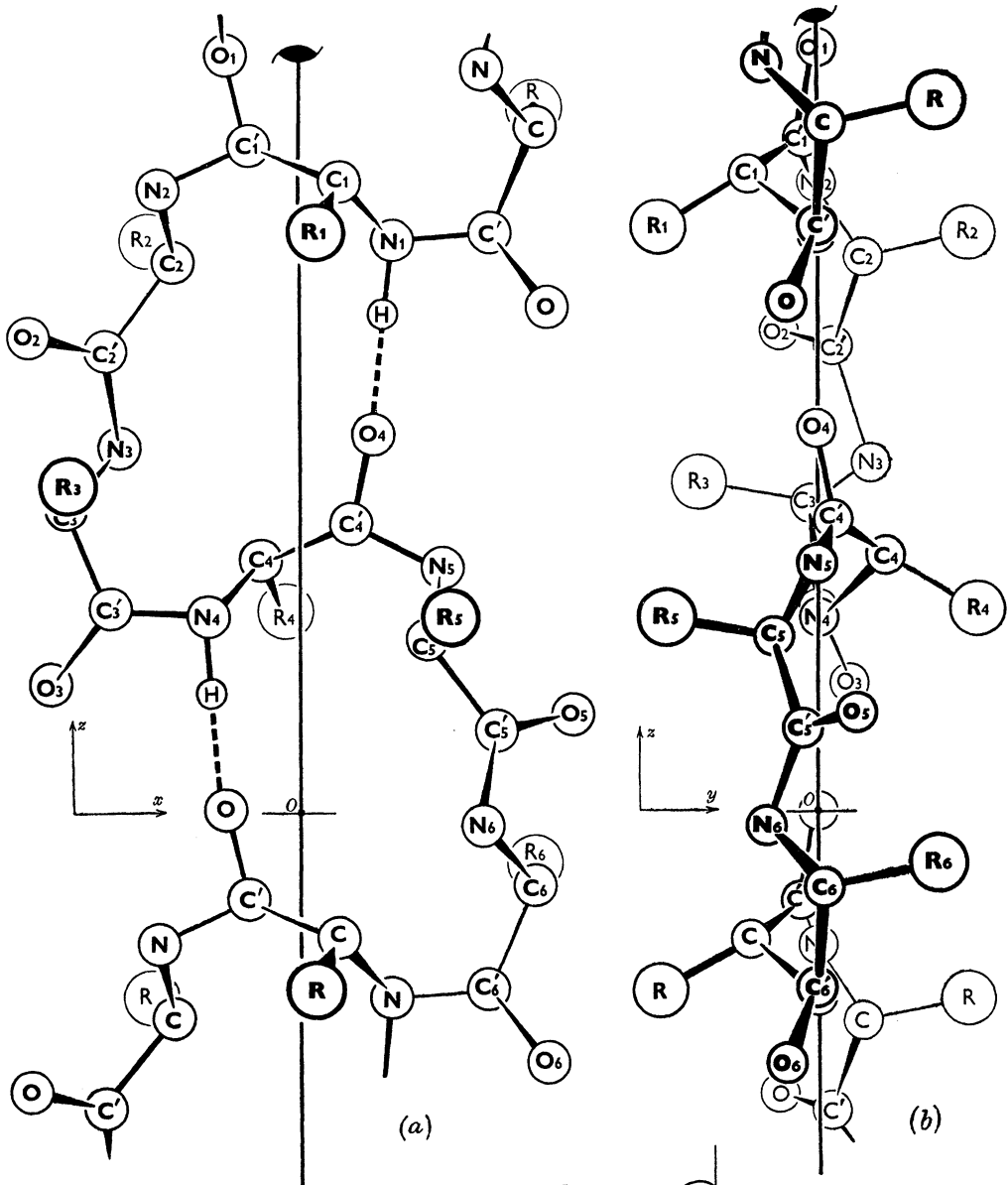


Table of co-ordinates in Å

	1			2		
	\bar{x}	\bar{y}	\bar{z}	\bar{x}	\bar{y}	\bar{z}
N	1.2	0.0	7.7	1.9	0.0	8.4
C	0.5	0.9	8.5	1.7	0.6	7.4
R	0.2	2.1	7.8	2.0	2.0	7.7
C'	0.7	0.2	9.0	2.6	0.2	6.2
O	1.0	0.0	10.2	3.6	0.5	6.4

	3		
	\bar{x}	\bar{y}	\bar{z}
N	2.4	0.7	4.9
C	3.1	0.1	4.1
R	3.1	1.6	4.4
C'	2.5	0.0	2.7
O	3.3	0.4	1.7

Atoms in amino-acid residues 4, 5, 6
at $(\bar{x}, \bar{y}, \bar{z} \pm 5.1)$

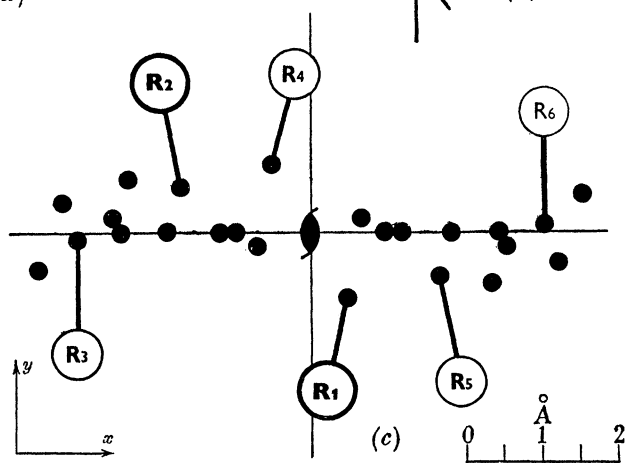


FIGURE 9. Projections of $2_{14} \cdot \frac{1}{3}$ chain.

Eight-atom ring (3₈)

We have not encountered any earlier description of this structure. Our model of it is illustrated in figure 10. The repeat distance is about 5.4 Å, and of course comprises three amino-acid residues. It is a feature of this structure that the three

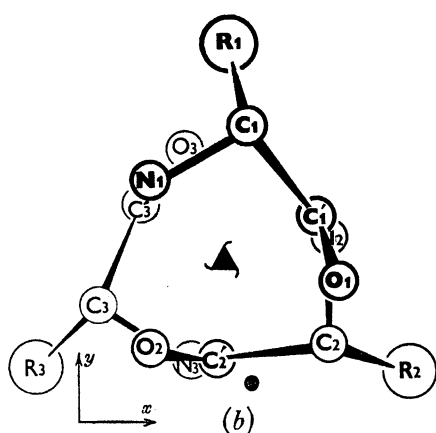
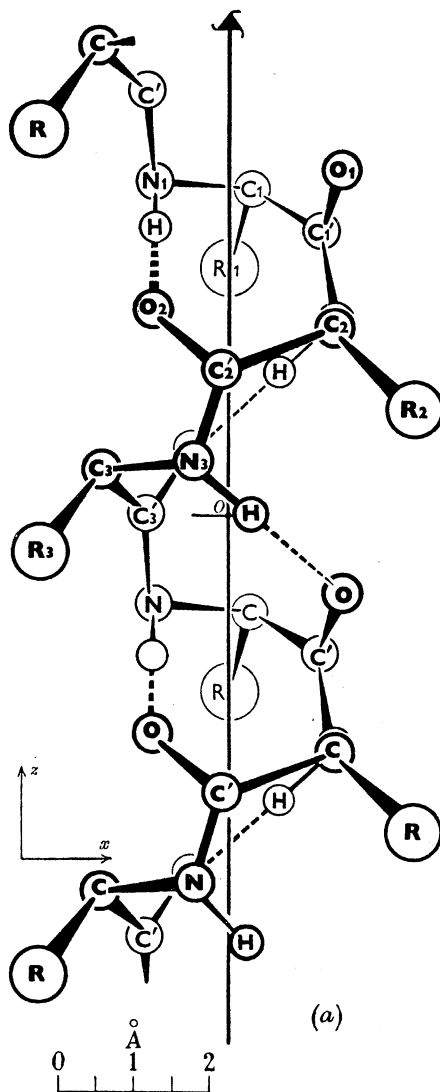


Table of co-ordinates in Å

	x	y	z
N ₁	1.0	1.0	4.4
C ₁	0.3	1.8	4.3
R ₁	0.0	2.8	3.2
C' ₁	1.2	0.6	3.7
O ₁	1.5	0.3	4.5
N ₂	1.4	0.3	2.5
C ₃	1.4	1.1	2.4
R ₂	2.4	1.4	1.4
C' ₂	0.1	1.3	1.9
O ₂	1.0	1.2	2.7
N ₃	0.4	1.4	0.7
C ₃	1.7	0.6	0.6
R ₃	2.5	1.4	0.5
C' ₃	1.1	0.7	0.0
O ₃	0.5	1.4	0.8

FIGURE 10. Projections of 3₈ chain.

N—H...O bonds in each repeating unit are mutually perpendicular and therefore could not possibly be responsible for any infra-red pleochroic effects which might be observed in the crystal or fibre. However, we regard the whole interpretation of these effects as still equivocal, and hence, though we shall not discuss this model further (owing to its close similarity to 3₁₀), we cannot definitely reject it at this stage.

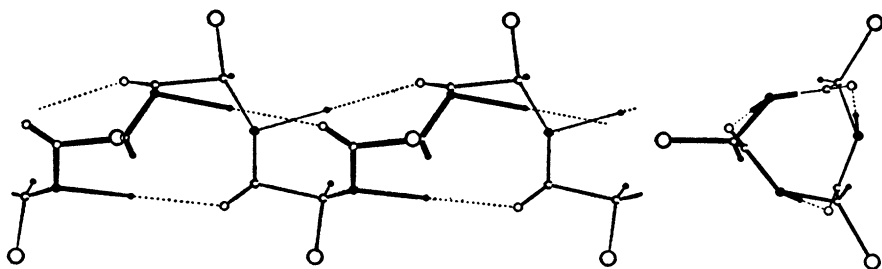


FIGURE 11. Projections of 3_{10} chain, after Huggins (1943). Repeat distance about 5.1 Å. Symbols as for figure 5.

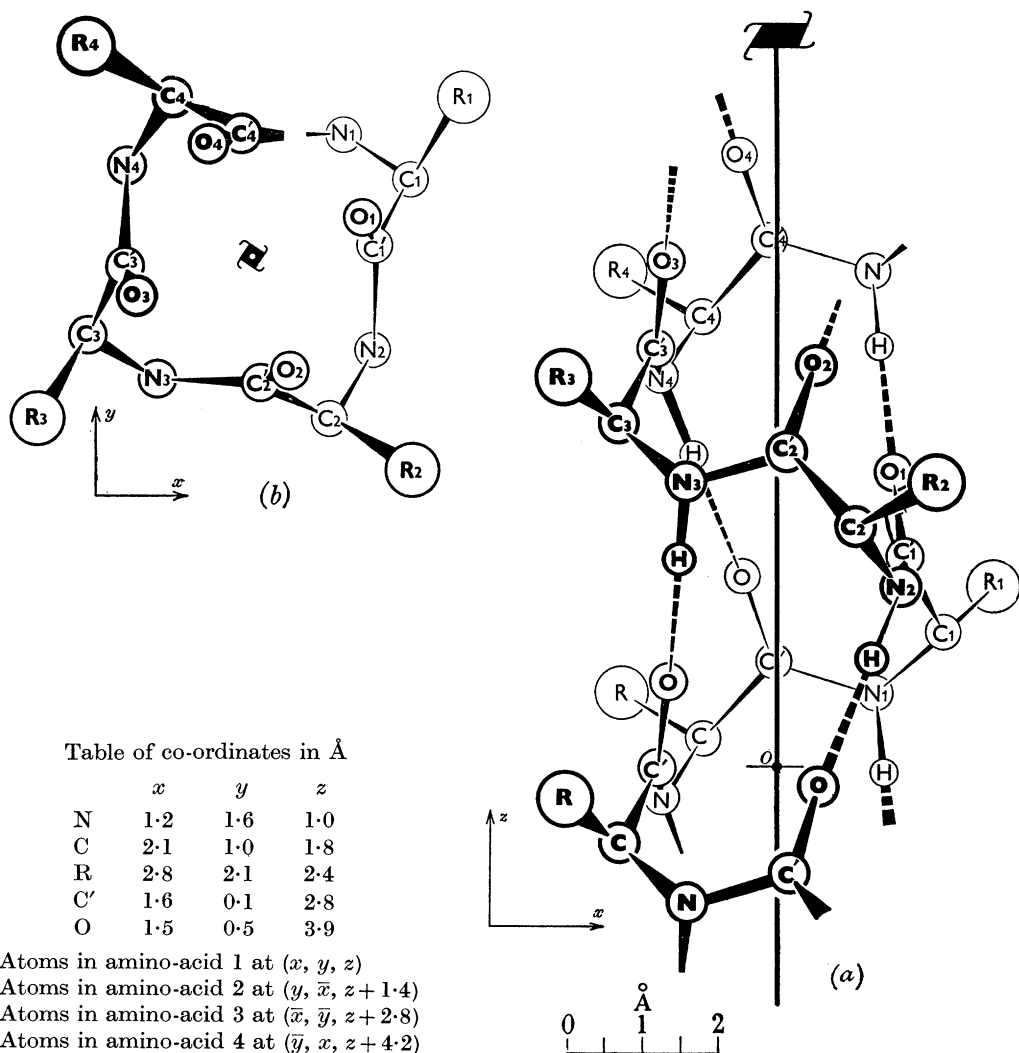


FIGURE 12. Projections of 4_{13} chain.

Ten-atom ring (3_{10})

This structure was first proposed by Taylor (1941); it was later discussed in detail by Huggins (1943), whose projections of it are reproduced in figure 11. Like 3_8 its repeat distance is rather over 5 \AA , and contains three residues; it has the further attractive feature that the hydrogen bonds are oriented nearly parallel to the triad axis and could therefore contribute to the infra-red pleochroism of the molecule as a whole, granted a suitable orientation of the chains.

We discuss this structure in detail below.

Other threefold structures

We have found it impossible to build threefold structures containing rings of 11, 13 or more members; structures with these large rings invariably assume a symmetry higher than threefold.

(iii) *Fourfold structures*

We were able to build no fourfold structures having rings of fewer than eleven or more than thirteen members.

Eleven-atom ring (4_{11})

A structure of this type can be built, though neighbouring pairs of oxygen atoms fall abnormally close to one another (2.2 \AA instead of the normal minimum of 2.8 \AA). Because of this feature, and because the repeat unit (of 5.4 \AA) contains *four* amino-acid residues, the structure is regarded as improbable.

Thirteen-atom ring (4_{13})

This version seems to be more plausible than 4_{11} since its rings are less strained; projections are given in figure 12. The repeat distance is again rather over 5 \AA and, of course, contains four amino-acid residues. This structure will be further referred to in §7.

6. STRUCTURES NOT EXAMINED IN DETAIL

(a) *Structures of higher symmetry*

We have not examined in detail structures with fivefold, sixfold or higher symmetry since they would inevitably contain more than four residues per repeat unit. There is, nevertheless, no difficulty in building models of such structures. For example, there is a hexad structure 6_{20} , with a repeat distance of about 5 \AA and containing six residues per repeat; it is an open helix whose internal diameter is between 7 and 8 \AA .

(b) *Structures in which only part of the CO and NH groups are hydrogen bonded*

We mentioned in §4 above that structures of this type were considered comparatively unlikely, because they would be expected to have a higher free energy than those in which all CO and NH groups are bonded. Nevertheless, a general survey has been made. The list of possible structures is not as formidable as might appear at first sight, because the repeat distances along the fibre axis become too long when the fraction of bonded NH or CO groups falls below $\frac{1}{3}$, and often already when it falls

below $\frac{1}{2}$. Some structures were found to be sterically impossible, and others again, though capable of shortening to give the desired repeat of $n \times 5 \text{ \AA}$, contained no bond that would keep them in this shortened configuration.

The structures are described in table 2, which includes only three models having the correct repeat distance. Though these three, as well as some of the other structures listed in table 2, may have any type of screw symmetry owing to the rotational freedom of the individual rings, we decided for simplicity's sake to describe them all in terms of twofold screw axes. We append some comments on the three structures having approximately correct repeat distances.

TABLE 2

screw axis of symmetry	no. of atoms in ring	fraction of NH (and CO) groups which are hydrogen bonded		no. of residues per repeat	no. of illustrations (figure no.)	comments
		fraction of	true repeat			
twofold	7	$\frac{1}{2}$	11.6	4	13	two side-chains above and two below plane of main chain; free rotation between rings
	7	$\frac{1}{3}, \frac{1}{4}, \text{etc.}$	≥ 10	6, 8, etc.	—	'repeat' too long
	8	$\frac{1}{2}$	9.8	4	14	free rotation between rings
	8	$\frac{1}{3}, \frac{1}{4}, \text{etc.}$	≥ 10	6, 8, etc.	—	'repeat' too long
	10	$\frac{1}{2}$	8.2	4	—	'repeat' too short
	10	$\frac{1}{3}$	12	6	—	'repeat' too long
	11	$\frac{1}{2}$	8.0	4	—	'repeat' too short
	11	$\frac{1}{3}$	9.6	6	15	possible structure; properties similar to 2_{13} and 2_{14} , though rings are not so flat; side chains point alternately up and down from the plane of the rings; free rotation between rings
	11	$\frac{1}{4}$	16.5	8	—	'repeat' too long
	13	$\frac{1}{2}$	—	—	—	no possible structure
	13	$\frac{1}{3}$	10.2	6	8	described in table 1
	13	$\frac{1}{4}$	≥ 10	8	—	'repeat' too long
	14	$\frac{1}{2}$	—	—	—	no possible structure
	14	$\frac{1}{3}$	10.2	6	9	described in table 1
14	$\frac{1}{4}$	≥ 10	8	—	'repeat' too long	
three-fold	10	$\frac{1}{2}$	13.4	6	—	'repeat' too long
	11	$\frac{1}{2}$	13.0	6	—	'repeat' too long

(i) *Seven-atom ring. One-half of CO and NH groups hydrogen bonded ($2_7 \cdot \frac{1}{2}$)*

In its fully extended state this structure has a repeat distance of 11.6 \AA , with four amino-acid residues per repeat, and pairs of side-chains protruding alternately above and below the plane of the main chains in the *a* form. In the *b* form pairs of side-chains protrude alternately right and left within the plane of the main chain. Neighbouring side-chains are thus brought improbably close together. This proximity can be avoided in a variety of ways, however, by departing from the twofold screw symmetry. Such departure produces a slight shortening of the repeat distance. In end-on projection a structure of the type $S_7 \cdot \frac{1}{2}$ with high-order screw symmetry is

likely to have an appearance similar to that of 3_8 , 3_{10} or 4_{13} . The structure with twofold screw symmetry is illustrated in figure 13, which shows that neighbouring rings are free to rotate about the $C'-C-N$ bonds.

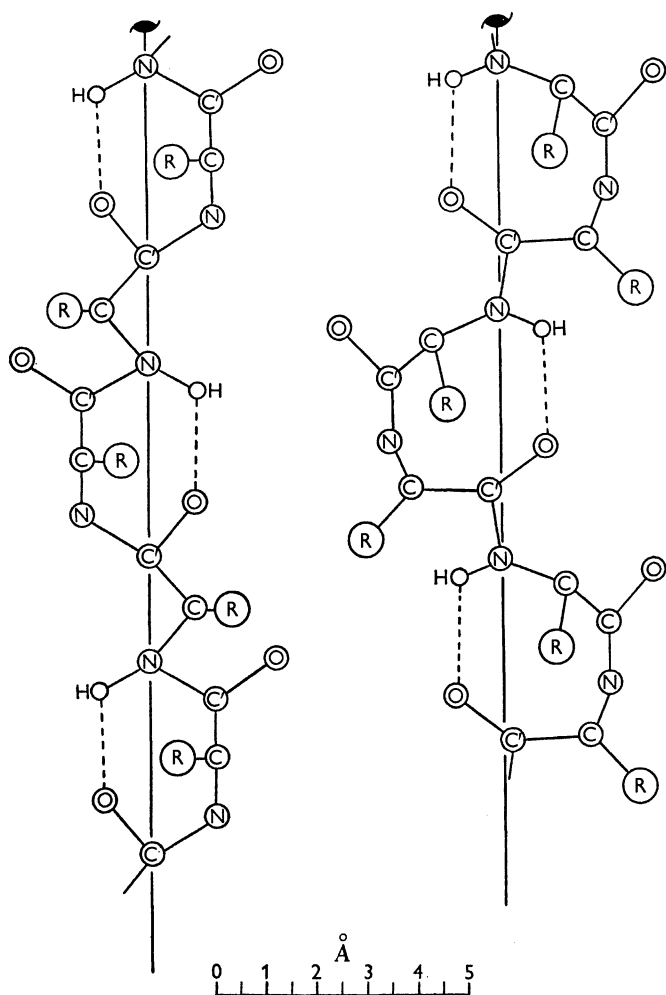


FIGURE 13

FIGURE 14

FIGURE 13. Sketch of the $2_{7.5}$ chain.

FIGURE 14. Sketch of the $2_{8.5}$ chain.

(ii) *Eight-atom ring. One-half of CO and NH groups hydrogen bonded ($2_{8.5}$)*

This structure has a repeat distance of 9.8 \AA with four amino-acid residues and is shown in figure 14. Neighbouring rings are free to rotate about the $N-C'$ bond, without appreciably affecting the length of the repeat. Side-chains protrude alternately parallel and perpendicular to the plane of the rings.

(iii) *Eleven-atom ring. One-third of NH and CO groups hydrogen bonded ($2_{11.3}$)*

This structure has a repeat distance of 9.6 \AA with six amino-acid residues, and is shown in figure 15. Side-chains protrude alternately up and down from the plane

of the rings. The properties of this model are very similar to those of 2_{13} and 2_{14} except that the individual rings are free to rotate about the C'—N bond. Viewed in end-on projection with the resolving power of the Pattersons of haemoglobin and myoglobin this structure would be indistinguishable from 2_{13} or 2_{14} .

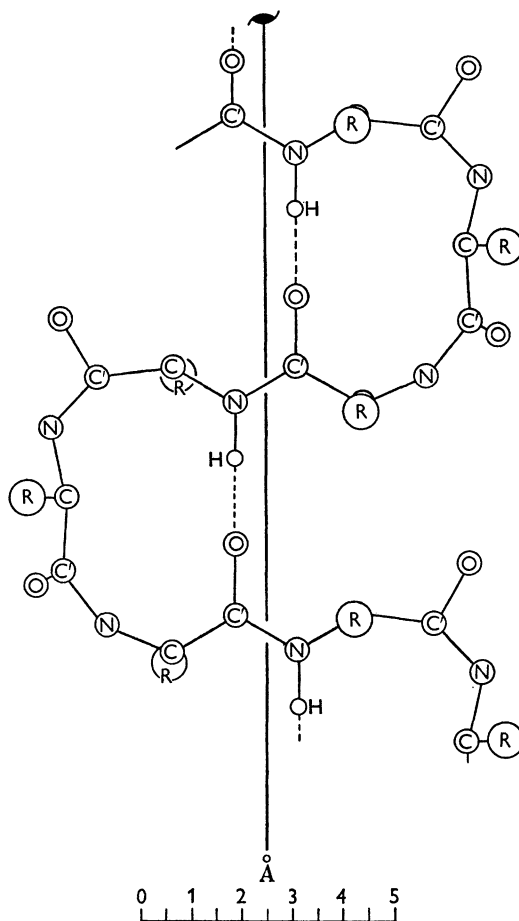


FIGURE 15. Sketch of the $2_{11.1/3}$ chain.

It is seen that in each of the three structures listed above the hydrogen bonded rings are free to rotate relative to each other. None of the models can be tied down to one specific configuration, so that definite parameters, on which a comparison of observed and calculated intensities might be based, cannot be assigned. Depending on the nature of the screw symmetry, end-on projections of these models would resemble one or the other of the types of structure described in detail in § 5. With the resolving power available no distinction can be made between models whose end-on projections are roughly similar.

(c) *Structures with secondary folds*

We may distinguish structures of a higher degree of complexity in which the 'asymmetric unit' which the screw axis repeats consists not of one hydrogen-

bridged ring alone, but of more than one ring or of some combination of rings and chain. Some of the structures in this category are similar to those described in the preceding paragraph and suffer from the same drawbacks. Examples of such structures are those recently discussed by Mizushima, Simanouti, Tsuboi, Sugita & Kato (1949); for example, these authors illustrate one in which the 'asymmetric unit' is formed of two 2_7b rings and one amino-acid residue which is not part of a ring; this asymmetric unit is operated upon by a screw dyad to form the chain.

Another structure of this type was proposed by Huggins (1943) and is illustrated in figure 16; here the asymmetric unit consists of three linked 2_7a rings, and the whole chain has screw dyad symmetry. Using space-filling models we have found the repeat distance of this chain to be 8 Å, not 10.2 Å as stated by Huggins. We ourselves have found it possible to coil a 2_7b structure into helical forms, with, for example, a screw hexad symmetry. The structure has an asymmetric unit consisting of a single seven-membered ring; the internal diameter of the helix is about 5 Å and its pattern repeats at intervals of 6 Å.

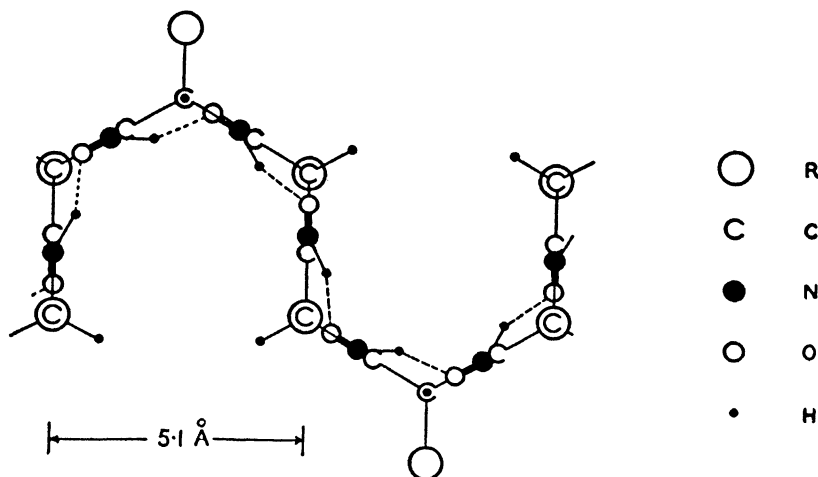


FIGURE 16. Structure with secondary folds built up of 2_7a rings, after Huggins (1943).

We have not made an exhaustive study of structures of this general type, because they all suffer from what we believe is a serious objection, namely, that, although the rings are held together by hydrogen bonds as in all the configurations mentioned earlier, the structure as a whole is not so held together. For example, there seems no obvious reason why Huggins's structure referred to in the last paragraph should not stretch to the ordinary 2_7b structure, since such stretching would involve only the overcoming of van der Waals attractions, and not the rupture of hydrogen bonds. It seems to be generally accepted that the α - β transformation in keratin involves the breaking of hydrogen bonds at least, though it must be admitted that up to now studies of the energetics of this transformation have not given any clear indication of the nature of the processes involved.

To sum up, after excluding the types of chain which are incompatible with three or four residues in a sub-cell of $10 \times 10 \times 5$ Å, or are improbable for other reasons,

we are left with 2_7b , $2_{13} \cdot \frac{1}{3}$, $2_{14} \cdot \frac{1}{3}$, 3_8 , 3_{10} and 4_{13}^* . In the next section an attempt is made to narrow the possibilities still further by comparing the vector projections of these chains with the Patterson projections of haemoglobin and myoglobin. If, as is believed, the chains are parallel in their crystals, the vectors between atoms of the same chain should have a deciding influence on the form of the Patterson around the origin when the projection is upon a plane at right angles to the chain direction. This is a more favourable basis of comparison than the Patterson projections shown in figure 3 (*a*) and (*d*), because the chain vectors are crowded into a small area and their effects are superimposed.

7. VECTOR PROJECTIONS OF CHAIN MODELS

(*a*) End-on projections of the chains

A series of vector projections on a plane at right angles to the chain are shown in figure 17. The atoms C, N, O have been assigned equal weights, and vectors have been drawn between every pair of atoms in one repeat of the chain. These are transferred to a common origin, and their ends are plotted as projected on a plane. For instance, in the $2_{14} \cdot \frac{1}{3}$ chain which has six amino-acid residues in a repeat distance of $10 \cdot 2 \text{ \AA}$ or 30 atoms per repeat, there are 900 vectors, 30 of which constitute the peak at the origin. The projection is divided into squares of 1 \AA side, and the number of vector ends in each square is counted. The density distribution can then be represented by contours at convenient levels. The vectors are distributed irregularly, and considerable smoothing is necessary in drawing the contours, but these contours serve to represent sufficiently accurately the type of Patterson diagram which the chains would give if arranged regularly in a simple unit cell with a wide spacing between the chains so that there is little overlap of the origin peaks. The correct representation would be attained by calculating the F 's of this arrangement, and constructing a Patterson projection with them. Such a calculation has been carried out for certain chains in the myoglobin cell in §8 below.

Figure 17 (*a*) is the vector projection of the $2_{14} \cdot \frac{1}{3}$ chain illustrated in figure 9. The atoms lie very nearly in a plane, with the exception of the first atoms R of the side-chains, and the Patterson projection is correspondingly elongated in one direction and narrow in the other. In comparing this projection with the origin peak of myoglobin, it must be remembered that the crystal has a dyad screw axis perpendicular to the chains, so that the part of the Patterson distribution around the origin due to the chains is a combination of (*a*) with its counterpart formed by rotation about the axis. Assuming the axis to be nearly parallel to the plane of the chain, the result would be as shown in figure 17 (*b*), which is formed by superposing (*a*) and its mirror image. Figure 17 (*e*) shows the distribution around the origin in a Patterson projection of myoglobin on a plane perpendicular to the chains, and it will be noted that it has a general similarity to (*b*). Figure 17 (*f*) is a similar projection for horse methaemoglobin. Its form suggests that it may be due to the superposition of

* The chains $2_7 \cdot \frac{1}{3}$, $2_8 \cdot \frac{1}{3}$ and $2_{11} \cdot \frac{1}{3}$ are also possible, but these chains are so ill-defined by the limiting conditions hitherto available that detailed tests of their conformity with the observations cannot be made at the present stage.

Assuming an arrangement of this kind, pairs of chains have been inserted into the projection in the way indicated in figure 21 (*e*). The resulting values of F^2 are compared in table 6 with those observed, and it will be noticed again that the agreement is not so good as in the case of the $2_{13} \cdot \frac{1}{3}$ and $2_{14} \cdot \frac{1}{3}$ configurations. The calculated vector projection (figure 22 (*d*)) fails in the following respects:

(i) the origin peak is too thick and is insufficiently elongated; there is no sign of resolution into three peaks;

(ii) in the middle layer the four peaks are unresolved. The same difficulties are encountered in the 3_8 and 4_{13} structures.

9. DISCUSSION OF RESULTS

If we base our conclusions only on the projections on a plane at right angles to the chain direction, undoubtedly the best agreement with the experimentally determined projections and intensities of diffraction is given by the $2_{13} \cdot \frac{1}{3}$ and $2_{14} \cdot \frac{1}{3}$ chains. Considering that the side-chains and intermolecular liquid have been left completely out of account, the resemblance between the projections (figures 22 (*a*), (*b*), and 20 (*b*)) is very striking. In contrast to this agreement, in the structures shown in figures 22 (*c*) and (*d*) the lack of resolution of peaks in the middle layer would appear to be a decisive criterion for rejecting these structures for myoglobin. The nearest peaks in the middle layer of the observed vector projection are only 4.2 \AA apart; it can be shown by drawing cross-sections of the vector equivalent of the various chains (as illustrated in figure 17) that resolution could be expected only for chains about as narrow, in the x direction, as the $2_{13} \cdot \frac{1}{3}$ or $2_{14} \cdot \frac{1}{3}$ configurations.

The extension of the comparison to the three-dimensional Patterson projection of haemoglobin in §7, however, shows how misleading it may be to base conclusions on reflexions around a single zone. It will be necessary to get a three-dimensional Patterson for myoglobin in order to assess whether these resemblances are significant or fortuitous.

The problem is very complex, and one is forced to rely on a number of items of evidence each of which is very slight. The conclusion that the chains are of a folded coplanar form resembling the $2_{13} \cdot \frac{1}{3}$ or $2_{14} \cdot \frac{1}{3}$ type would be more convincing if there were any indication that the form has obvious advantages over others. In X-ray analysis in general, when a crystal structure has been successfully analyzed and a model of it is built, it presents so neat a solution of the requirements of packing and interplay of atomic forces that it carries conviction as to its essential correctness. In the present case the models to which we have been led have no obvious advantages over their alternatives. Much more evidence must be accumulated before conclusions can be safely drawn. There are certain hopeful features, however, at the present stage of the investigation. Assuming that criteria of the type adopted in this paper are justified, a survey shows that the number of possible forms of chain is very restricted. There appears to be a real simplicity of chain structure in myoglobin, which will perhaps be shown by other favourably built proteins, and which makes it particularly suitable for intensive X-ray investigation. There is hope that the study of such proteins may lead to a reliable determination of the chain structure.

The authors wish to thank Mr C. W. Bunn for his helpful advice on the technique of model building, Mr F. H. C. Crick for his valuable criticisms of a number of points raised in §7, and Miss V. E. Marting for calculating the Fourier transform of the 3_{10} helix. They also thank the Medical Research Council for financial support of two of them (J. C. K. and M.F.P.)

REFERENCES

- Albrecht, G. & Corey, R. B. 1939 *J. Amer. Chem. Soc.* **61**, 1087.
Ambrose, E. J., Elliott, A. & Temple, R. B. 1949 *Nature*, **163**, 859.
Ambrose, E. J. & Hanby, W. E. 1949 *Nature*, **163**, 483.
Astbury, W. T. 1949*a* *Brit. J. Radiol.* **22**, no. 259.
Astbury, W. T. 1949*b* *Nature*, **164**, 439.
Astbury, W. T. & Bell, F. O. 1941 *Nature*, **147**, 696.
Boyes-Watson, J., Davidson, E. & Perutz, M. F. 1947 *Proc. Roy. Soc. A*, **191**, 83.
Bragg, W. L. 1949 *Nature*, **164**, 7.
Chibnall, A. C. 1945 2nd Procter Memorial Lecture. Croydon: Int. Soc. of Leather Trades' Chemists.
Corey, R. B. 1948 *Advances in protein chemistry*, **4**, 385. New York: Academic Press Inc.
Darmon, S. E. & Sutherland, G. B. B. M. 1949 *Nature*, **164**, 440.
Huggins, M. L. 1943 *Chem. Rev.* **32**, 195.
Hughes, E. W. & Moore, W. J. 1949 *J. Amer. Chem. Soc.* **71**, 2618.
Kendrew, J. C. 1950 *Proc. Roy. Soc. A*, **201**, 62.
Kendrew, J. C. & Perutz, M. F. 1949 *Haemoglobin*, p. 161. London: Butterworths.
Levy, H. A. & Corey, R. B. 1941 *J. Amer. Chem. Soc.* **63**, 2095.
Mizushima, S., Simanouti, T., Tsuboi, M., Sugita, T. & Kato, E. 1949 *Nature*, **164**, 918.
Perutz, M. F. 1949 *Proc. Roy. Soc. A*, **195**, 474.
Sanger, F. 1948 *Nature*, **162**, 491.
Taylor, H. S. 1941 *Proc. Amer. Phil. Soc.* **85**, 1.
Zahn, H. 1947 *Z. Naturforschung*, **2b**, 104.