THE FOLDING PATHWAY FOR GLOBINS

V. I. LIM and A. V. EFIMOV

Institute of Protein Research, Academy of Sciences of the USSR, Poustchino, Moscow Region, USSR

Received 5 April 1977

1. Introduction

Previous investigations have successfully predicted the three-dimensional structures of trypsin inhibitor [1,2], myoglobin [3,4], TMV protein [5] and rubredoxin [2]. Using mechanisms of protein chain folding which one of us [5,6] suggested, we can find a common folding pathway for the globins which are included in the 'Atlas of Protein Sequence and Structure' [7]. In the present paper this common folding pathway is illustrated by the examples of haemoglobin α- and β-chains [8], lamprey globin [9] and insect erythrocruorin [10].

2. Method and results

Let us consider an α-helix of a particular type which will be called the s-helix and a di-s-helical structure which will be called the F-structure.

The s-helix is an α-helix in which most of the hydrophobic side chains form a hydrophobic cluster of an arbitrary width extending from one end of the helix to the other. One border of the hydrophobic cluster will be called the right border (RB), and the other the left border (LB) (see fig.1). All or most of the hydrophobic residues forming each cluster border must be located on the portion of the α-helical cylindrical surface generated by a sector of the cross-section with a sector angle not exceeding ~ 120°.

Proline can be present in the internal s-helical turns if this arrangement causes an elongation of the hydrophobic cluster of an s-helix and if proline is not located on the surface of this cluster. An analysis with CPK models shows that the presence of proline in the internal turns does not produce a large change in the geometry of the α-helix.

The F-structure is a di-s-helical structure in which the s-helices approach each other in such a way that the N-end of one s-helix is near the C-end of the other. Interactions between s-helices are realized by the hydrophobic and hydrophilic side chains which are located along the RBs of both s-helices. Hydrophobic side chains of one s-helix intercalate into the space between the hydrophobic side chains of the other s-helix. The hydrophilic side chains form two types of interhelical hydrogen and salt bonds: (1) side chain with side chain; (2) side chain with water molecule(s) or ion(s) with side chain. The whole F-structure is a body which has hydrophobic and hydrophilic surfaces. In an aqueous medium the F-structure is more advantageous energetically than any other di-s-helical structure [5]. A schematic representation of the s-helix and the F-structure is given in [5].

Following our mechanism, the folding of a globin polypeptide chain was performed in two stages: (1) the formation of F-structures from s-helices; (2) the joining of F-structures with other F-structures and with s-helices.

If the α-helix had a long hydrophobic cluster and did not satisfy the definition of the s-helix, we broke it into two helices both of which could be considered as s-helices.

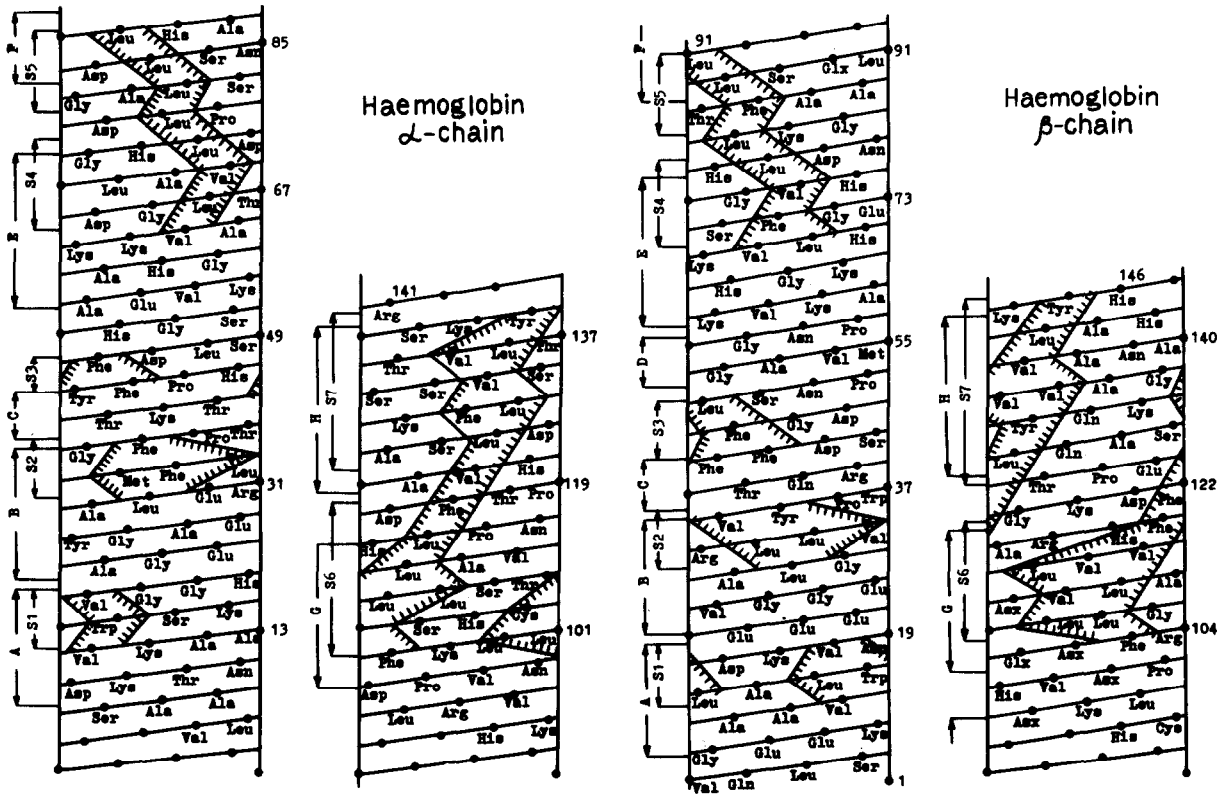
The F-structures composed of long s-helices were formed first, since long F-structures are more stable than the short ones. Sterically two right borders (RB) of the same sense (right-handed or left-handed) interact better than two RBs of the opposite sense, and this fact was taken into account when we formed the F-structures (especially the long ones). The

F-structures and s-helices were joined at their hydrophobic surfaces.

It is known that almost all polar protein groups form intramolecular hydrogen or salt bonds or contact with water molecules [11]. Therefore we considered only such F-structures which could be joined to each other and (or) to s-helices without intercalation of hydrophilic side chains into the hydrophobic core and without shielding of hydrophilic groups from water molecules. Earlier [5] we rejected the F-structures mainly with the help of RB–RB interactions of two s-helices. Such an approach did not allow us to obtain the only possible set of F-structures.

Figure 1 shows that each of the four molecules which we studied has seven regions (S1, S2... S7). These regions are s-helices if they are wound in the α-helical conformation. The s-helices S6 and S7 are the longest in all four cases. These s-helices always have right-handed RBs. There is only one s-helix S4 which is comparable in length to the S6. Such an s-helix S4 is found in the erythrocruorin molecule. However, this s-helix has a left-handed RB. Therefore the F-structure (S6 · S7) formed from the s-helices S6 and S7 is the longest and most stable F-structure in all four molecules.

Since S4 is the longest of the s-helices (S1, S2...S5) we form the second F-structure using S4. The resulting F-structures will be (S4 · S1), (S4 · S2), (S4 · S3) and (S4 · S5). The combinations (S4 · S3) and (S4 · S5) are not of interest, as S3 and S5 have the position *i*, occupied by a hydrophilic side chain, while the positions *i*+1 and *i*+4 are occupied by hydrophobic side chains which form RB s-helices (see fig.1). The presence of such 1–2–5 'triangles' is not possible on the surface of a short F-structure, which must be joined to another F-structure because there is a hydrophilic side chain in position 1 of the triangle. This side chain will penetrate into the inter-F-structural hydrophobic core whenever two F-structures are joined. Such hydrophilic groups are the OH-group of Tyr 42 and the side chain of Asp 82 in the haemoglobin α-chain, Ser 44 and Thr 84 in the haemoglobin β-chain,



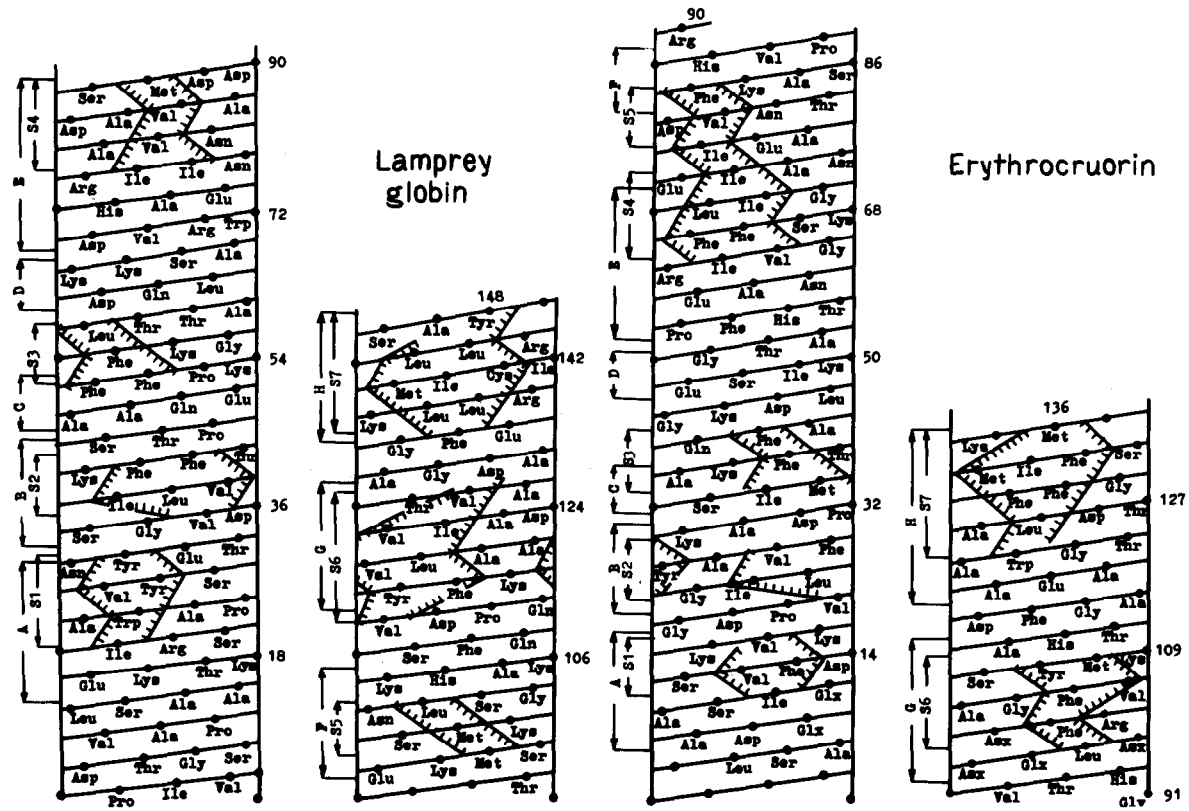


Fig.1. Amino acid sequences of α- and β-chains of haemoglobin, lamprey globin, erythrocruorin plotted on the α-helical nets. A, B, C...H are helical regions in the native structure of the globin molecule. S1, S2...S7 are the s-helices. Regions of the chains having only two hydrophobic residues in positions i and i+3 are not considered to be s-helices. ( ) is the RB of the s-helices, ( ) is the LB of the s-helices.

Lys 54 and Ser 96 in lamprey globin, Lys 37 and Asp 79 in erythrocruorin (see fig.1).

It should be pointed out that F-structures may also be rejected for other topological reasons. For example, two s-helices cannot form an F-structure if along the chain there are one or two residues between them. Three hydrophobic side chains in positions *i*, *i*+3, *i*+6 of one s-helix cannot simultaneously take part in interhelical interactions with hydrophobic side chains of the second, etc.

It is necessary to examine only the F-structures (S4 · S1) and (S4 · S2) and to see how they can be joined to (S6 · S7). The F-structure (S6 · S7) should not be attached to the F-structure (S4 · S1) or (S4 · S2) by its terminal moiety formed by N- and C-terminal parts of S6 and S7, respectively. Otherwise the Asn 97 in the α-chain and Asx 102 in the haemoglobin β-chain, Tyr 114 in lamprey globin and Glx 96 in erythrocruorin would be immersed in the hydrophobic core. These hydrophilic residues are a part of the triangle 1–2–5 in which the hydrophobic residues forming the RB of the s-helix S6 occupy positions 2 and 5. Therefore, (S4 · S1) or (S4 · S2) must be shifted to the other terminus (S6 · S7) formed from the C- and N-terminal parts of S6 and S7, respectively (see fig.2a).

The F-structure (S6 · S7) should be joined crosswise to (S4 · S1) or (S4 · S2). The (S6 · S7) hydrophobic cluster is longer than that of (S4 · S1) and (S4 · S2). Therefore, if we do not join the F-structures crosswise the polar atoms will be situated on the hydrophobic surface of (S6 · S7). These atoms will be the polar atoms of terminal peptide groups of S1, S2 and S4 and of hydrophilic side chains and peptide

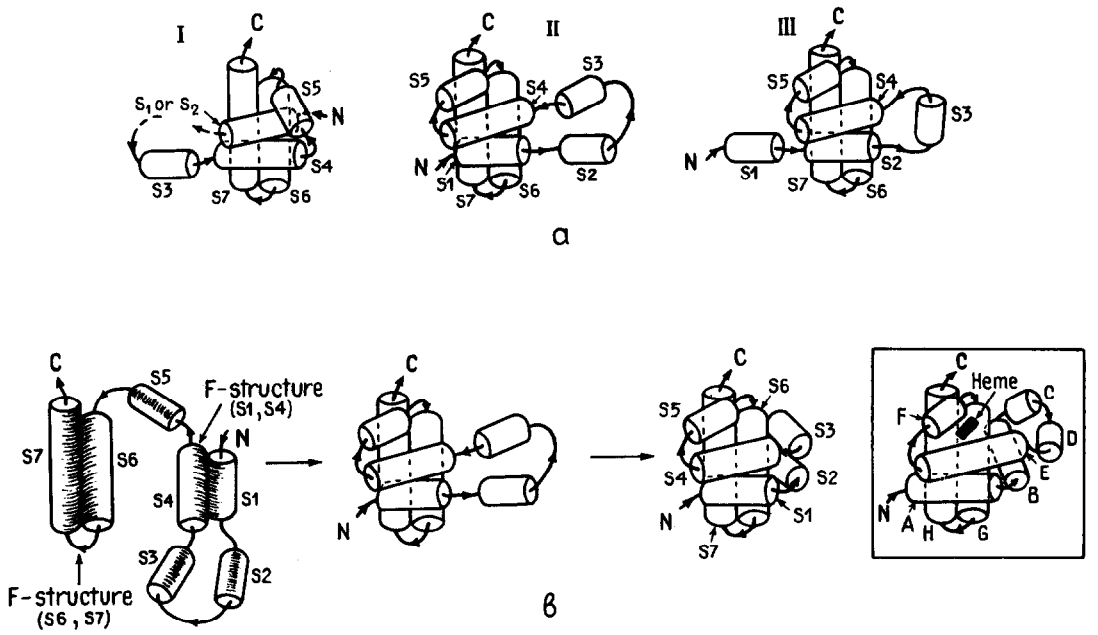


Fig.2. Packing of globin s-helices. (a) I, II, III are different variants of the di-F-structure formation. (b) Pathway of globin polypeptide chain folding. The globin structure determined experimentally is shown in the rectangle.

groups not belonging to S1, S2 and S4 but located near their terminals. Such a situation will be observed even if S1, S2 and S4 are elongated.

The F-structure (S6 · S7) must be joined to (S4 · S1) or (S4 · S2) according to variant II or III, respectively (see fig.2a). Variant I is less advantageous, as in this case the chain fragment connecting S4 and S6 will be situated on the hydrophilic surface of the di-F-structure.

In variant II the hydrophobic side chains of S2 can interact with the hydrophobic side chains of S3 (see fig.2a). In variant III such an interaction between S1 and S3 is sterically impossible. In all four cases the average angular width of the hydrophobic cluster S6 is greater than or equal to 200 degrees. Therefore, some of the hydrophobic side chains of the s-helix S6 lie on the surface of the di-F-structure. The above results lead to the following: when S2 and S3 are joined to the di-F-structure (fig.2b), a compact structure is formed more favourable in hydrophobic interactions than the structure formed by joining S1 and S3 to the di-F-structure.

Thus we predict that all the four globins should have a common folding pathway leading to a structure which is in good agreement with the native structure (see fig.2b). There are differences (~ 2 turns) only in the localization of helical ends.

3. Discussion

According to the mechanism of protein chain folding [5,6], the three-dimensional distribution of the protein chain material in native conformation is dictated by highly helical intermediate structures. In turn, the formation of highly helical intermediate structures is controlled mainly by such characteristics of the primary structure as the number and distribution of s-helices along the protein chain, their length and hydrophobic cluster geometry (s-helical characteristics of a polypeptide chain). Within the framework of this mechanism, globin folding is a particular case. Herein a highly helical intermediate structure is a native structure. The results of the present work corroborate, on the example of globins, the main assumption of our mechanism that the s-helical characteristics of a polypeptide chain code the protein folding pathway. In conclusion it should be noted

that our results are in good agreement with the conclusions of Perutz, Kendrew and Watson [12] that s-helices play an important role in globin structure formation.

Acknowledgements

The authors express their gratitude to Dr M. B. Garber and Professor O. B. Ptitsyn for critical remarks.

References

[1] Levitt, M. and Warshel, A. (1975) Nature 253, 694–698.
[2] Kuntz, I. D., Crippen, G. M., Kollman, P. A. and Kimelman, D. (1976) J. Mol. Biol. 106, 983–994.
[3] Ptitsyn, O. B. and Rashin, A. A. (1975) Biophys. Chem. 3, 1–20.
[4] Rashin, A. A. (1976) Bioorg. Khim. 10, 1373–1380.
[5] Lim, V. I. and Efimov, A. V. (1976) FEBS Lett. 69, 41–44.
[6] Lim, V. I. (1975) Dokl. Akad. Nauk 222, 1467–1469.
[7] Dayhoff, M. O. (1972) in: Atlas of Protein Sequence and Structure, Vol. 5, National Biomedical Research Foundation, Silver Spring.
[8] Perutz, M. F. (1969) in: Encyclopedia of Polymer Science and Technology, Vol. 11, pp. 646–677. John Wiley and Sons, New York, London, Sydney, Toronto.
[9] Hendrickson, W. A. and Love, W. E. (1971) Nature New Biol. 232, 197–203.
[10] Huber, R., Epp, O., Steigemann, W. and Formanek, H. (1971) Eur. J. Biochem. 19, 42–50.
[11] Blow, D. M. and Steitz, T. A. (1970) Ann. Rev. Biochem. 39, 63–100.
[12] Perutz, M. F., Kendrew, J. C. and Watson, H. C. (1965) J. Mol. Biol. 13, 669–678.