Hypothesis

POLYPEPTIDE CHAIN FOLDING THROUGH A HIGHLY HELICAL INTERMEDIATE AS A GENERAL PRINCIPLE OF GLOBULAR PROTEIN STRUCTURE FORMATION

V. I. LIM

Institute of Protein Research, Academy of Sciences of the USSR, 142292 Poustchino, Moscow Region, USSR

Received 27 February 1978

1. Introduction

General features of the three-dimensional structures of globular proteins can be explained if it is assumed that at early stages of protein folding the major part of the polypeptide chain is in the α-helical conformation [1]. Analysis of the energy, kinetics and diffusion factors as well as of the general features of the primary structure of globular proteins also suggests the formation of highly helical intermediate globule as a general principle of protein folding. In the present paper the results of such an analysis are briefly summarized. It is deduced that:

(i) Short α-helices bearing hydrophobic clusters on their surfaces must be formed first.
(ii) These must then be united into a highly helical intermediate globule.
(iii) A native structure must be formed by the subsequent transition of α-helices of the highly helical intermediate globule into different types of secondary structure (this transition is not accompanied by any considerable shifts of the material of α-helices relative to each other).

The stereochemical simulation of the formation of a highly helical intermediate and its subsequent transition into another conformation permits one to predict, in good agreement with experiment, three-dimensional structures of globular proteins with different ratios of α-helices and β-sheets [2–6]. The suggested principle of protein folding is a development of Pauling and Corey's ideas [7] that in the absence of long-range interactions the α-helix is the most preferable conformation for a polypeptide chain.

2. Structural prerequisites

2.1. *Characteristic distribution of the hydrophobic and hydrophilic residues along the polypeptide chain in globular proteins*

If the whole polypeptide chain of a natural globular protein were coiled into an α-helix the major portion of the hydrophobic side groups would be found in clusters on the helix surface. The total length of these hydrophobic clusters along the helix axis will be 60–80% of the whole length of the α-helix, and the length of each cluster will extend through 2–7 helix turns. This striking feature is observed in all globular proteins regardless of the content of the α- and β-forms in their native structures (see fig.1, [2,3,6]).

An α-helix with one hydrophobic cluster will be described here as an s-helix and the corresponding fragment of the polypeptide chain as an s-fragment (fig.1a). Let us assume that the N-end of an s-helix can be located in positions $i-4$, $i-3$, $i-2$, $i-1$, and the C-end in positions $j+1$, $j+2$, $j+3$ and $j+4$, where the i and j are the positions occupied by the first and the last hydrophobic residue of the cluster, respectively.

2.2. *Properties of s-helices*

2.2.1. Property 1

It follows from the stereochemistry of the polypeptide chain that the number of hydrogen bonds and van der Waals' contacts will decrease [7] and the hydrophobic cluster will break down into smaller clusters when an s-helix is transformed into any other sterically allowed conformation. This means that the

Elsevier/North-Holland Biomedical Press

α-helix is energetically the most stable conformation for a single s-fragment if it does not yet interact with other parts of the polypeptide chain and does not contain a large amount of identically charged side groups. Moreover, in any other conformation the s-fragment will be in contact with a larger amount of water molecules, and thus the entropy of water will decrease when an s-helix is transformed into any other conformation.

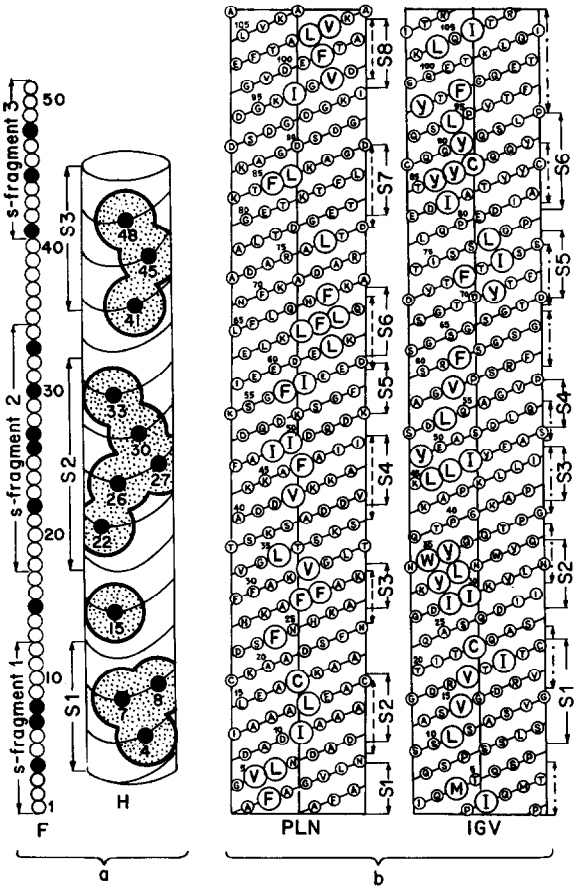

Fig.1. s-Helices. (a) F is the polypeptide chain of a globular protein with a typical distribution of hydrophobic (•) and hydrophilic (○) residues. H is an α-helix built from chain F. Dotted regions on the α-helix surface are hydrophobic clusters. Helical regions S1, S2, S3 are s-helices corresponding to s-fragments 1, 2, 3 of the F chain. (b) Double α-helical nets of parvalbumin (PLN) (helical protein [8]) and of the variable part of immunoglobulin (IGV) (β-structural protein [9]). Large circles denote bulky hydrophobic residues. S1, S2, S3 ... are s-helices. The ends of these s-helices can be prolonged by 1–4 residues (determination of the s-helix termini, section 2.1.). (←– – –→), (←–·–·–·→) are helical and β-structural regions in the native structures. A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

2.2.2. Property 2

It has been shown experimentally that the time of α-helix formation is about 10^{-7} s [10,11]. It is much shorter than the time (10^{-2} s and more) necessary for forming the elements of protein tertiary structure [12,13]. In particular, it is much shorter than the time of β-structure formation which is measured in minutes and even hours [14–18]. Hence, s-helices must be formed much faster than other conformations.

2.2.3. Property 3

s-Fragments of globular protein polypeptide chains contain about 10–20 residues (fig.1b, [2,3,6]). Then s-helices of such a length are small, compact globules with an axial ratio ranging from 1:1 to 2:1 (fig.2). It is natural that such compact s-helices will diffuse in solution much faster than any other conformation of the s-fragment.

2.2.4. Property 4

The s-helix is a rigid structure. There are two regions on its surface, the hydrophobic and the hydrophilic ones, and they are divided by a clear boundary. This means that s-helices possess rigid specific patterns of their surfaces thus providing conditions for selective and cooperative interactions between different parts of the polypeptide chain.

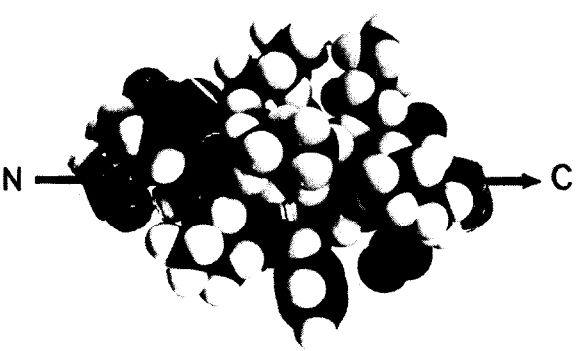

Fig.2. An α-helical fragment assembled from CPK atomic models (12 amino acid residues). The arrow shows the α-helix axis directed from the N- to the C-end of the α-helix.

3. Stages of protein folding

3.1. *Formation of s-helices within the polypeptide chain as the first stage of protein folding*

I postulate that at the first stage of protein folding (the initial stage when long-range interactions are absent) the α-helix is the most probable structure for s-fragments which comprise the major proportion of the polypeptide chain length (section 2.1.). The basis for the postulate is that s-helices are the most preferable energetically (property 1), and can be formed many times before the appearance of long-range interactions (property 2).

If it is taken into account that the s-helices are the most quickly diffusing structures in solution (property 3), it can be concluded that they must be the elements for the first long-range interactions resulting in formation of a highly helical globule.

3.2. *Packing of s-helices into a globule as the second stage of protein folding*

It is natural to think that the highly helical globules formed at this stage must satisfy the main principles of structural organization which are observed in native structures of globular proteins. Numerous experimental data on three-dimensional structures of globular proteins permit one to suggest three main principles of their structural organization [4–6, 19–21].

1. Most hydrophobic side groups form a tightly packed core, while most hydrophilic side groups form a polar shell to the protein globule.

2. Polar atoms of the polypeptide chain, which are found within the internal part of the globule, always form intramolecular hydrogen or salt bonds.

3. Hydrophobic side groups in the core form double hydrophobic layers.

The latter principle of structural organization is also observed in crystals of organic compounds [22] and in crystals of amino acids [23]. The hydrophobic layers in proteins are formed by β-structural sheets, by wide hydrophobic clusters of α-helices and by two α-helices whose hydrophobic cluster borders are drawn together according to the knob-hole principle (fig.3).

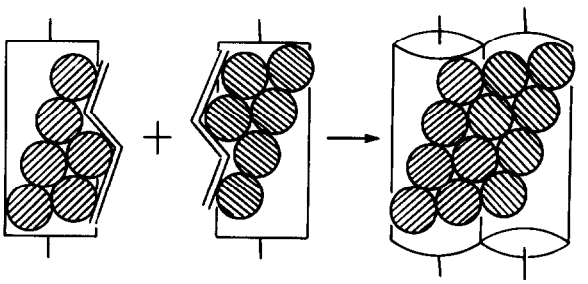

Fig.3. Hydrophobic layer formed from the hydrophobic clusters of two s-helices according to the knob-hole principle [21]. (,) Hydrophobic groups on the s-helix surface.

Taking into account the three principles of structural organization mentioned, we built highly helical globules from s-helices of polypeptide chains of several globular proteins with known tertiary structures [6]. We began the globule construction with a search for the s-helices which could form hydrophobic layers according to the knob-hole principle (fig.3), and we finished the construction by uniting the hydrophobic layers with each other and with other s-helices. At each stage of such an assembly, the selectivity and cooperativity of the interaction between s-helices (property 4) decreased the number of alternative packing combinations. It was possible to satisfy all the three principles of structural organization only in strictly defined, unique globules. Consequently, at the second stage, the random search of different packing combinations of s-helices resulted in unique highly helical globules.

Each of the highly helical globules must be built from three to eight s-helices. In the bi-s-helical structure it would be impossible to shield a considerable part of the hydrophobic side groups from contact with water. Nine or more s-helices would be difficult to pack into one compact globule without shielding the hydrophilic side groups of some s-helices by hydrophobic clusters of other s-helices. Eight s-helices could still be packed into a globular structure according to the scheme: six s-helices are assembled in a bunch, on each butt of which one more s-helix is located. The length of an s-fragment is about 10–20 residues (fig.1a, [2,3,6]). Consequently, to build one highly helical globule it is necessary to have a chain consisting of 50–150 residues, and in those

cases where the chain length is several hundred residues, multidomain folding must take place.

The folding of proteins which are highly helical in their native state must be terminated at this second stage. This happens when the highly helical globule formed at this stage is stable enough to remain unchanged. With multidomain folding of the polypeptide chain the globules first formed can then be united into one large highly helical globule.

3.3. *Intramolecular rearrangements in highly helical globules as the third stage of protein folding*

This stage must follow the previous one when folding of the polypeptide chain of a protein of low helical content takes place. Such a protein can result from transformation of s-helices of a highly helical globule into the β-structural and/or irregular conformations. In the cases of folding of long polypeptide chains, the globules of low helical content formed in this way can be united into one large globule, in particular, into a globule of highly β-structural content.

As a rule, there are hydrophilic side groups on the surface of s-helices. That is why s-helices cannot be immersed into the highly helical globule and must be located on its surface. Therefore the transformation of s-helices into other conformations will not meet strong steric hindrance. Energy loss due to breakdown of the s-helix can be compensated by long-range interactions. For example, the transformation of s-helices into other conformations can result in:

(i) A more tightly packed and/or larger hydrophobic core.
(ii) Formation of additional intraglobular bonds such as disulfide bonds and/or coordinate bonds with cofactors.
(iii) Better contacts with other globules and/or chain regions which have not been involved in the formation of a highly helical globule.
(iv) A smaller globule surface which gives an entropy gain at interaction with water.

A detailed analysis of the transition conditions is given in [5,6].

The transition of the s-helix into another conformation cannot be accompanied by a distance shift of the s-fragment on the surface of the hydrophobic core of a globule, since this process would be accompanied by a considerable damage to the core and its unshielding. Consequently, the s-fragments in a globule of low helical content must be located relative to each other approximately in the same manner as at the preceding stage of the highly helical globule, i.e., the three-dimensional structure of proteins of low helical content has been determined by interactions between s-helices in intermediate highly helical globules.

4. Testing the hypothesis

Recently our group has developed a completely a priori stereochemical theory of the three-dimensional structure of globular proteins [4–6] based on the above-mentioned principle of protein folding. Using the algorithm developed for the prediction of native structure we tested our theory on a number of protein molecules with known tertiary structure, such as globins and carp parvalbumin (helical proteins), the variable part of immunoglobulin (β-structural protein), carboxypeptidase (a large protein with both α- and β-structures), trypsin inhibitor (a protein with disulfide bridges). These molecules represent the main types of three-dimensional structures of globular protein known at present [19]. A comparison of theoretical and X-ray three-dimensional structures of these proteins has shown that there is practically complete agreement between them both in the localization of helical and β-structural regions and in the spatial orientation of these regions relative to each other. The following peculiarity has been also observed: all helical regions and most of the β-strands in native structures are localized on s-fragments (fig.1b), an exception being the β-strands which are located on the borders of β-structural sheets and have few hydrophobic residues. It is just this peculiarity that must be expected from our principle of protein folding.

The available experimental evidence on the folding pathway of proteins [24–26] (reviewed [27–30]) is also in good agreement with our theory. For example, the order of disulfide bridge formation in the trypsin inhibitor molecule determined theoretically [6,31] and experimentally [25] completely coincide. Recently, a folding scheme of α-lactalbumin deduced from the three-state denaturation mechanism has been proposed [26]. According to this scheme a helical globular intermediate is formed first, and its

transition into the native structure is accompanied by a decrease of the helix content and the simultaneous appearance of the β-structure.

5. Discussion

It should be noted that for β-structural proteins it would be reasonable to propose an alternative pathway of native structure formation where the β-structure is formed at the first stage of folding [32]. Indeed, the β-structure is the only one capable of competing energetically to a certain extent with the α-helix [32]. Nevertheless, this alternative scheme seems to me hardly probable for the following reasons:

(1) The parallel β-structure is formed by drawing together of the sections which are distant from each other along the polypeptide chain. The same is valid for the antiparallel β-structure observed in globular proteins since their β-structural sheet is assembled, as a rule, from β-structural hairpins which are also distant from each other along the chain [19].

(2) The formation of each hydrogen bond in the β-structural hairpin is accompanied by the fixation of two amino acid residues. The fixation of two residues by one hydrogen bond also takes place on formation of the parallel β-structure when two free strands are united.

In contrast to the β-structure, the α-helix is formed by drawing together of the residues which are located close to each other along the chain, and one hydrogen bond fixes only one residue.

In my opinion, these distance and entropy differences in formation of α- and β-structures must result in an inability of the β-structure to compete effectively with the α-helices and their complexes in the first stage of protein folding (property 2, section 2.2.2.).

Acknowledgements

The author thanks Professor A. S. Spirin and Drs M. B. Garber, A. L. Mazanov and A. V. Efimov for useful discussions.

References

[1] Lim, V. I. (1975) Dokl. Akad. Nauk SSSR 222, 1467–1469.
[2] Lim, V. I. and Efimov, A. V. (1976) FEBS Lett. 69, 41–44.
[3] Lim, V. I. and Efimov, A. V. (1977) FEBS Lett. 78, 279–283.
[4] Lim, V. I., Mazanov, A. L. and Efimov, A. V. (1978) Molekul. Biol. 12, 206–213.
[5] Lim, V. I. (1978) Molekul. Biol. 12, 214–218.
[6] Mazanov, A. L. and Lim, V. I. (1978) Molekul. Biol. 12, 219–232.
[7] Pauling, L. and Corey, R. B. (1951) Proc. Natl. Acad. Sci. USA 37, 235–282.
[8] Kretsinger, R. H. and Nockholds, C. E. (1973) J. Biol. Chem. 248, 3313–3326.
[9] Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M., Huber, R. and Palm, W. (1974) Eur. J. Biochem. 45, 513–524.
[10] Cummings, A. L. and Eyring, E. M. (1975) Biopolymers 14, 2107–2114.
[11] Zana, R. (1975) Biopolymers 14, 2425–2428.
[12] Anfinsen, C. B. (1972) Biochem. J. 128, 737–749.
[13] Anfinsen, C. B. (1973) Science 181, 223–230.
[14] Snell, C. R. and Fasman, G. D. (1973) Biochemistry 12, 1017–1025.
[15] Peggion, E., Cosani, A., Terbojevich, M. and Romanin-Jacur, L. (1974) J. Chem. Soc. Chem. Commun. 9, 314–316.
[16] Hartman, R., Schwaner, R. C. and Hermans, J., jr (1974) J. Mol. Biol. 90, 415–429.
[17] Auer, H. E. and Patton, E. (1976) Biophys. Chem. 4, 15–21.
[18] Arfmann, H. A., Labitzke, R. and Wagner, K. G. (1977) Biopolymers 16, 1815–1826.
[19] Levitt, M. and Chothia, C. (1976) Nature 261, 552–558.
[20] Lim, V. I. (1974) J. Mol. Biol. 88, 857–872.
[21] Efimov, A. V. (1977) Dokl. Akad. Nauk. SSSR 235, 699–702.
[22] Kitaigorodsky, A. I. (1955) Organic Crystallochemistry (in Russian) Akad. Nauk SSSR, Moscow.
[23] Gurskaya, G. C. (1966) Structure of Amino Acids (in Russian) Nauka, Moscow.
[24] Robson, B. and Pain, R. H. (1976) Biochem. J. 155, 331–344.
[25] Creighton, T. E. (1977) J. Mol. Biol. 113, 275–293.
[26] Kuwajima, K. (1977) J. Mol. Biol. 114, 241–258.
[27] Wetlaufer, D. B. and Ristow, S. (1973) Annu. Rev. Biochem. 42, 135–158.
[28] Anfinsen, C. B. and Scheraga, H. A. (1975) Adv. Protein Chem. 29, 205–301.
[29] Schulz, G. E. (1977) Angew. Chem. 16, 23–32.
[30] Richards, F. M. (1977) Annu. Rev. Biophys. Bioengin. 6, 151–243.
[31] Mazanov, A. L., Falk, P. and Lim, V. I. (1978) in preparation.
[32] Ptitsyn, O. B. and Finkelstein, A. V. (1978) Bioorg. Khim. in press.