PRINCIPAL FOLDING PATHWAY AND TOPOLOGY OF ALL-β PROTEINS

O. B. PTITSYN, A. V. FINKELSTEIN and P. FALK (BENDZKO)*
Institute of Protein Research, USSR Academy of Sciences, 142292 Poustchino, Moscow Region, USSR

Received 27 February 1979

1. Introduction

The majority of proteins with mainly β-sheet secondary structure (all-β proteins [1] or simply β-proteins) have the structure of a double β-sheet rolled into an open or closed cylinder [1]. The topology of the β-sheet in these proteins usually includes the 'Greek key' [2] (see fig.1), and the topologies of all such proteins are surprisingly similar [5,6]. This suggests that these proteins have a common folding mechanism which determines the principal folding pathways for β-proteins. Now we propose a folding mechanism for β-proteins which is based on very simple and physically reasonable assumptions. According to this mechanism the folding pathway and the final protein topology depend only on the total number of β-strands and on the localization of the initiating complex in the given protein chain. It reduces the number of possible topologies for β-protein from 10^2 or 10^8 (depending on the number of β-strands) to only a few. Nevertheless the topologies of all known 'Greek key' β-proteins can be obtained on these pathways.

2. Methods

The kinetic [7] and thermodynamic [8,9] consideration of the β-structure formation in unfolded protein chains leads to the conclusion [5,10,11] that the most stable β-hairpins are often formed from long β-strands, each including two, three or even more

* Permanent address: Central Institute of Molecular Biology, GDR Academy of Sciences, Berlin-Buch, GDR

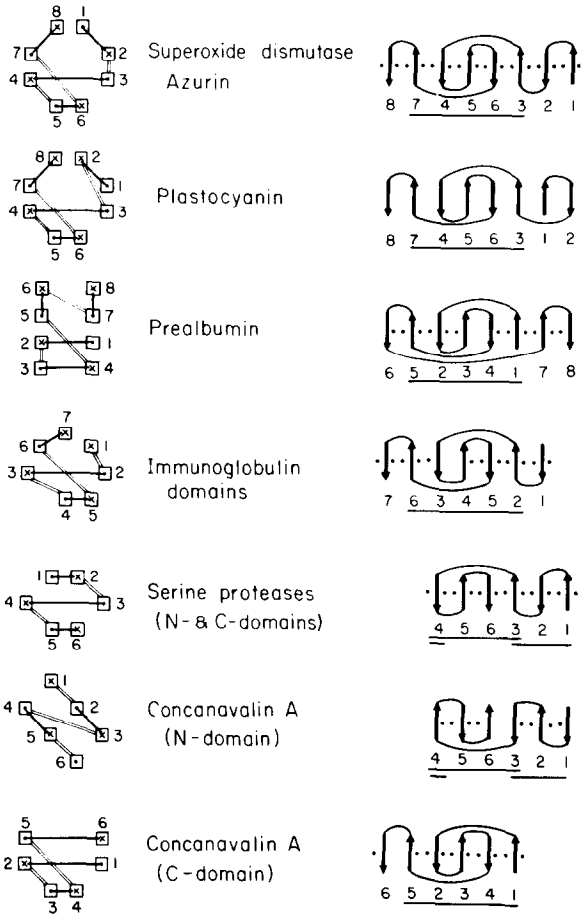


Fig.1. Schematic representation of β-protein structures with the 'Greek key' topology [1–4]. Left: top view of the open or closed cylinder. Right: plane representation of the double β-sheet ('topologies' [2]). Dots show hydrogen bonds between β-strands (hydrogen bonds in plastocyanin have not been reported). β-Strands included in the 'Greek key' (four strands) or the 'double Greek key' (five strands) are underlined.

segments of native β-structure. Computer simulation of pancreatic trypsin inhibitor folding [12] has also shown the existence of folding pathways passing through the stage of such long hairpins (see fig.13 in [12]).

Let us treat the one-center (i.e., the most favourable [13]) folding pathway which starts with the formation of a long β-hairpin from long β-strands, each including two (fig.2), three (fig.3) or more native β-segments. To screen its hydrophobic surfaces this hairpin must break into two, three or more parts at its 'weak' (hydrophilic) sites between the native β-strands [5,10,11] forming the 'embryo' of the double β-sheet. However, we must have in mind that due to the difference in the right-handed twist of the two β-sheets they can pack closely only if the 'top' sheet rotates in a clockwise direction [14,15]. Therefore the breaking of the long hairpin becomes possible only after an additional β-strand adjoins to its end forming the hydrophobic surface of the 'bottom' sheet wide enough to absorb the 'top' part of the hairpin. This leads to the formation of the double-sheet nucleation center which can consist of five (fig.2), seven (fig.3) or more native β-strands.

There are two mirror-image forms of a double β-sheet with the same topology differing in the clockwise or counter-clockwise swirl of the 'Greek key' [2]. The choice between these forms (i.e., the choice between two directions of hairpin breaking) is dictated by the requirement of the most compact nucleation center with the best screening of the hydrophobic core. For this the 'top' and the 'bottom' loops of the initiating hairpin must screen each other (fig.2,3). Together with the clockwise rotation of the 'top' β-sheet [14,15] this requirement leads to the unambiguous choice of the direction of hairpin breaking shown in fig.2,3. An analogous requirement of the most compact nucleation center determines the direction of the second breaking in a triple hairpin (fig.3). As a result the 'Greek key' swirl must be counter-clockwise as viewed from the outside of the double β-sheet (cf. [2]).

Other β-strands must adjoin the boundaries of the growing double β-sheet in a way to ensure minimal screening of hydrophilic surfaces and maximal screening of hydrophobic ones. The first means that the connections between β-strands should neither cross each other [16] nor cross the external surface of β-sheets [5]. Maximal screening of hydrophobic core means that the long internal intersheet connections of the A type (see upper part of fig.3) are always more favourable than the short external intersheet connections of the B type, while the intrasheet connections of the C type are possible only if the connections of other types would intercross after the adjoining of the next β-strand.

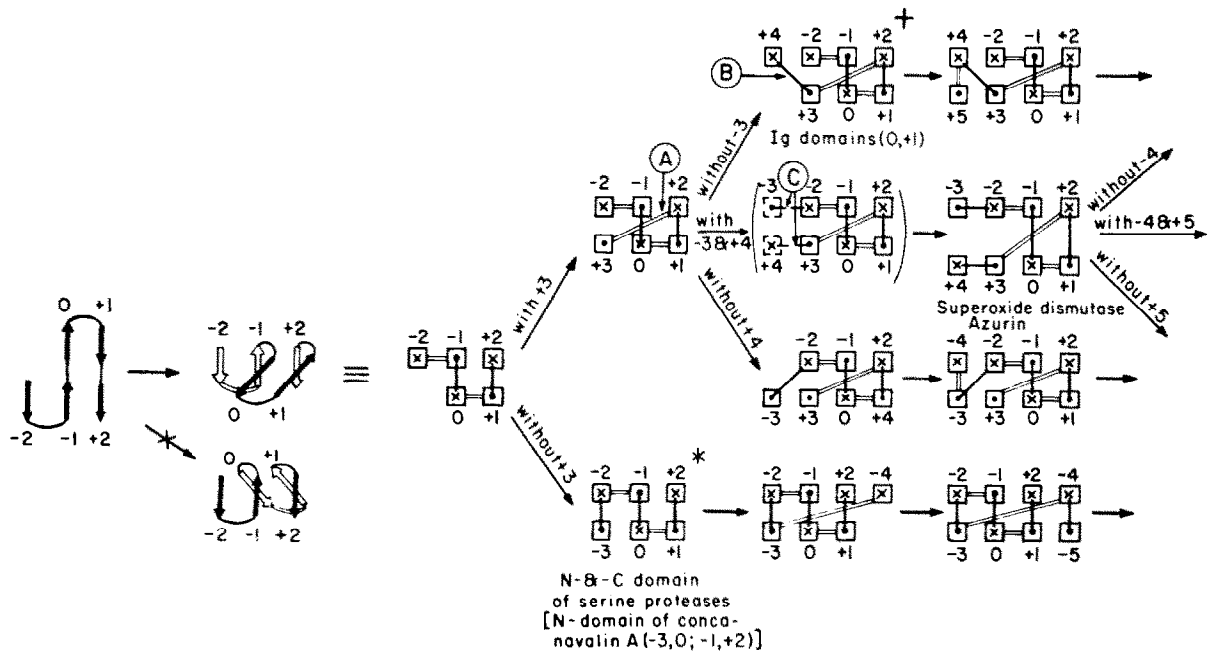
There are no further limitations on the sequence or places of adjoining of β-strands; in particular, they can adjoin both antiparallel or parallel to their neighbours in a β-sheet.

After the adjoining of the last β-strand, the hydrogen bonds between some neighbour β-strands can form or break without changing the topology of the molecule. In particular, the formation of the additional hydrogen bonds which close a double β-sheet into a β-barrel is very probable (especially for wide β-sheets shown in fig.3).

3. Results and discussion

Figures 2 and 3 present all the folding pathways (for β-proteins containing up to eight β-strands) allowed by the principles described above. They show that the folding pathway and the topology of each β-protein is determined unambiguously by its initiation complex and the number of β-strands in the protein. Topologies of β-proteins from nine and more

Fig.2. Folding pathways of β-proteins starting with five-stranded initiating complexes. Signs * and $^{+}$ mark two pairs of identical β-barrels. The examples of long internal intersheet connections (A), short external intersheet connections (B) and intrasheet connections (C) are shown in the upper part of the figure (see text). Structures from seven β-strands which are intermediates on the folding pathways for β-proteins with eight or more β-strands are shown in brackets (alternative β-strands adjoining in these structures are marked by dotted lines). All allowed topologies (excluding those obtained on folding pathways 'without +3' and 'without −3') include the double 'Greek key'. Proteins or domains possessing one of the obtained topologies are named (figures in brackets mean pairs of β-strands not connected by hydrogen bonds). The N-domain of concanavalin A is a mirror image of the presented structure.


Ig domains(0,+1)
Superoxide dismutase
Azurin
N-&-C domain
of serine proteases
[N-domain of conca-
navalin A(-3,0; -1,+2)]

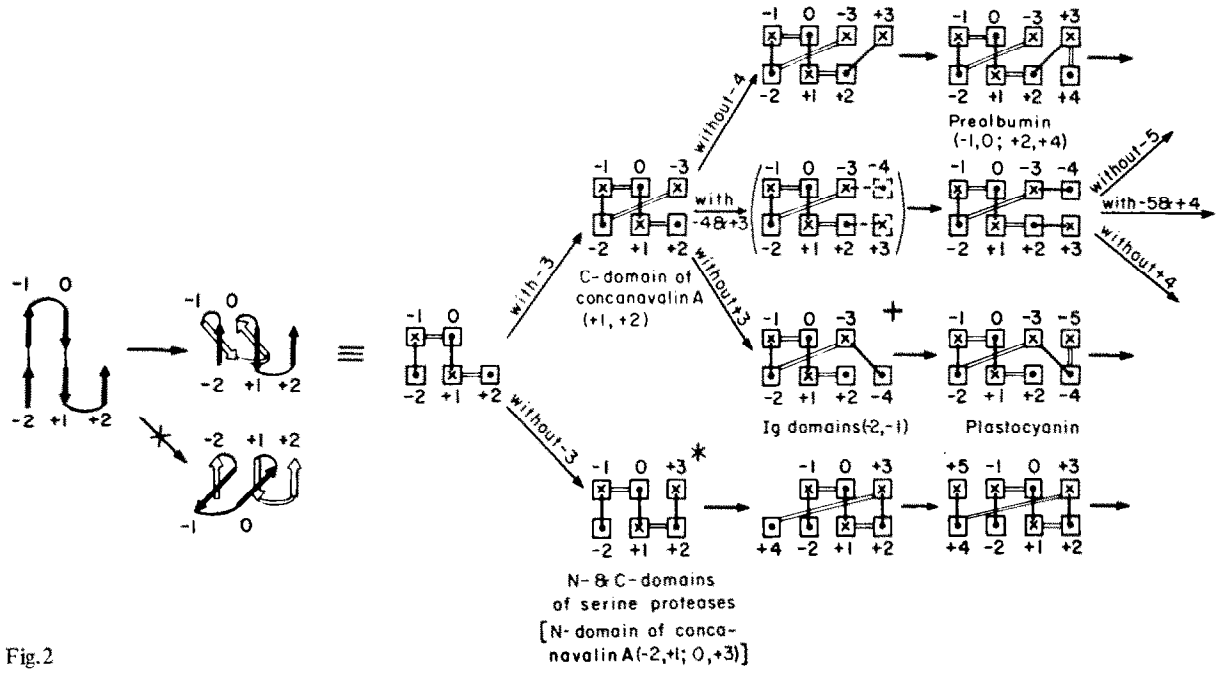

Preolbumin
(-1,0; +2,+4)
C-domain of
concanavalin A
(+1, +2)
Ig domains(-2,-1)
Plastocyanin
N-&C-domains
of serine proteases
[N-domain of conca-
navalin A(-2,+1; 0,+3)]

Fig.2

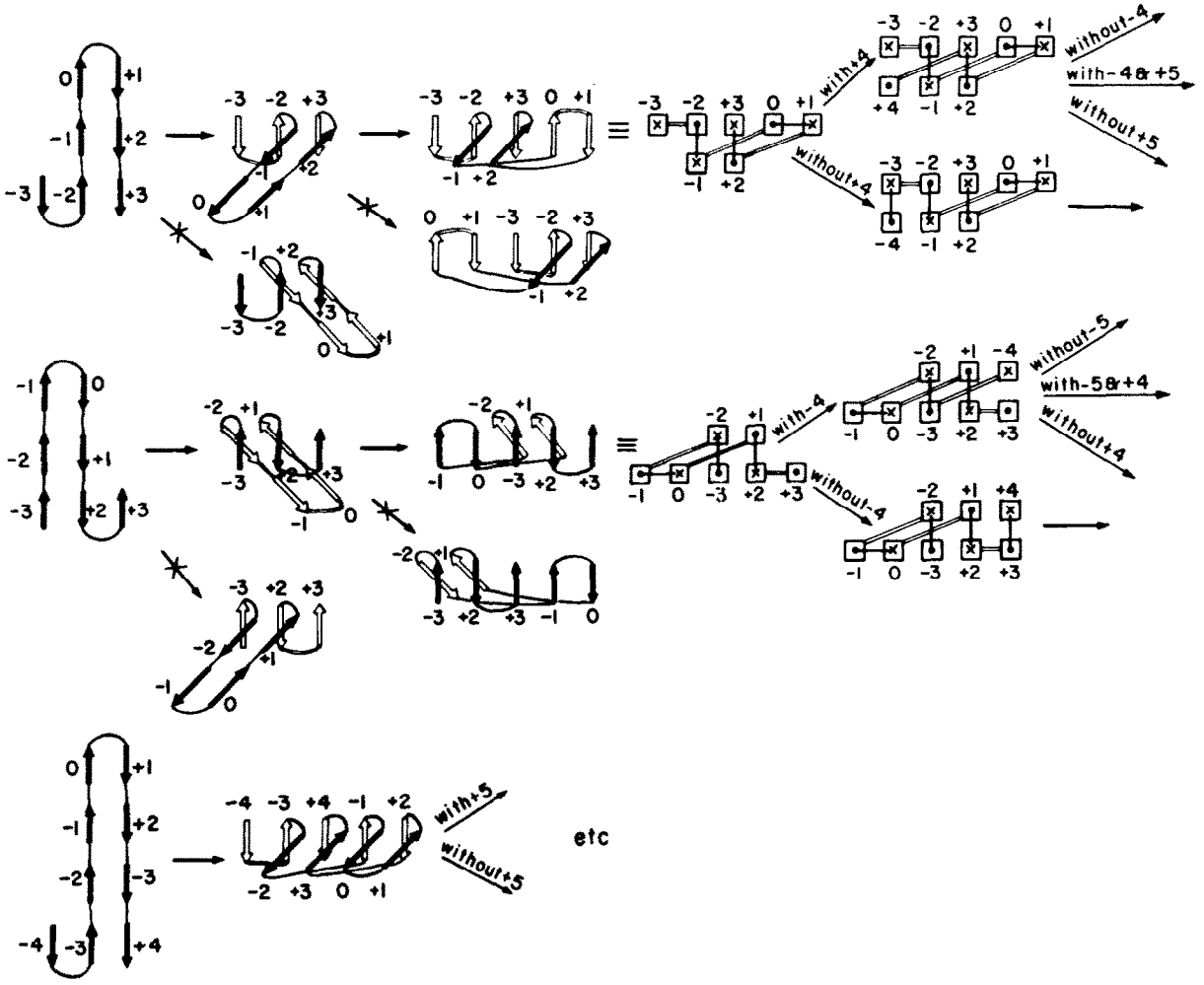


Fig.3. Folding pathways of β-proteins starting with the seven-stranded initiating complex. An example of folding of a longer initiating hairpin is shown in the bottom row.

β-strands can be easily obtained by following these pathways or by treating the longer initiating hairpins in an analogous manner.

Figure 2 shows that all β-proteins with the 'Greek key' topology fit our scheme. Each of them has one of the structures predicted by the proposed folding pathway. It is an essential result because the number of β-barrel topologies n_m allowed by the proposed folding pathway (see table 1) ranges from 0.5% (for five strands) to 0.002% (for eight strands) of the total number of its possible topologies $N_m = (m-1)! \cdot 2^{m-1}$ (m being the number of β-strands). This is a very strong evidence of the existence of directed folding pathways for all-β proteins leading to an unambiguous prediction of their structures.

It is very probable that many other all-β proteins or their domains with still unknown structures will have one of the topologies which are shown in fig.2,3 or can be obtained by a prolongation of the corresponding pathway. Folding pathways initiated by short β-hairpins (from native β-strands) are also possible, but seem to be less probable for all-β proteins. Other

Table 1

Number of β-strands m	5	6	7	8	9	10
Number of allowed structures n_m	2	3	7	12	18	24
Total number of possible structures N_m	384	3840	46 080	645 120	10 321 920	185 794 560
n_m/N_m	$5 \cdot 10^{-3}$	$8 \cdot 10^{-4}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-6}$	$1 \cdot 10^{-7}$

exceptions can be connected with some terminal β-strands (especially hydrophilic ones) which are not the true structural segments in a folding pathway.

The necessary condition for the existence of an initiating hairpin from long β-strands (each including two, three or more native β-segments) is that the lengths of two pairs, triads, etc. of adjacent native β-strands (together with their connections) must be approximately equal. This condition is approximately satisfied in proteins considered here for many adjacent pairs or triads including those which are necessary for the proposed folding mechanism.

It follows that the problem of topology prediction for all-β proteins is reduced now to the prediction of their secondary structure and to the localization of their initiating complexes. The number of allowed topologies of all-β proteins is surprisingly small and the probability for two of these proteins to possess the same or similar topologies is therefore very high, irrespective of their evolutionary origin or their functions. A good example of this is azurin whose primary structure is homologous to that of plastocyanin [17] and whose three-dimensional structure is homologous to that of superoxide dismutase [3].

Acknowledgements

The authors are indebted to A. G. Murzin for fruitful discussion and to A. G. Raiher for improving our English.

References

[1] Levitt, M. and Chothia, C. (1976) Nature 261, 552–558.
[2] Richardson, J. S. (1977) Nature 268, 495–500.
[3] Adman, E. T., Stenkamp, R. E., Sieker, L. C. and Jensen, L. H. (1978) J. Mol. Biol. 123, 35–47.
[4] Colman, P. M., Freeman, H. C., Guss, J. M., Murata, M., Norris, V. A., Ramshaw, J. A. M. and Venkatappa, M. P. (1978) Nature 272, 319–324.
[5] Ptitsyn, O. B. and Finkelstein, A. V. (1979) Proc. 12th FEBS Meet., Dresden, Pergamon, New York, Oxford, in press.
[6] Ptitsyn, O. B., Finkelstein, A. V. and Bendzko, P. (1979) Biofizika 24, 21–26.
[7] Finkelstein, A. V. (1978) Bioorg. Khim. 4, 340–344.
[8] Finkelstein, A. V. (1978) Bioorg. Khim. 4, 345–348.
[9] Finkelstein, A. V. (1979) Proc. Symp. Biomolecular Structure, Conformation, Function and Evolution, Madras, in press.
[10] Ptitsyn, O. B. and Finkelstein, A. V. (1978) Bioorg. Khim. 4, 349–353.
[11] Ptitsyn, O. B. and Finkelstein, A. V. (1979) Proc. Symp. Biomolecular Structure, Conformation, Function and Evolution, Madras, in press.
[12] Levitt, M. (1976) J. Mol. Biol. 104, 59–107.
[13] Ptitsyn, O. B. (1975) Dokl. Akad. Nauk SSSR 223, 1253–1256.
[14] Efimov, A. V. (1977) Dokl. Akad. Nauk SSSR 235, 699–702.
[15] Chothia, C., Levitt, M. and Richardson, D. (1977) Proc. Natl. Acad. Sci. USA 74, 4130–4134.
[16] Lim, V. I., Mazanov, A. L. and Efimov, A. V. (1978) Molekul. Biol. 12, 206–213.
[17] Ryden, L. and Lundgren, J.-O. (1976) Nature 261, 344–346.