Structural patterns in globular proteins

Michael Levitt

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Cyrus Chothia

Service de Biochemie Cellulaire, Institut Pasteur, 75724 Paris, France

A simple diagrammatic representation has been used to show the arrangement of α helices and β sheets in 31 globular proteins, which are classified into four clearly separated classes. The observed arrangements are significantly non-random in that pieces of secondary structure adjacent in sequence along the polypeptide chain are also often in contact in three dimensions.

ONE of the central problems in molecular biology is the formation of the native structure of a protein from the newly synthesised unfolded "structureless" polypeptide chain. In suitable conditions this process occurs spontaneously and the final conformation is determined solely by the amino acid sequence. As the random search of all possible conformations of the whole molecule would take an impossibly long time, this process probably involves intermediate structures that allow the protein to find its native conformation rapidly. Considerable experimental and theoretical effort has been devoted to trying to establish the nature of these intermediates.

In this article we first use a simple two-dimensional representation to illustrate the known conformations of 31 proteins. After classifying these known protein structures into four classes, we show that there is a strong tendency for pieces of secondary structure that are close together along the sequence also to be in close contact in the final three-dimensional structure. Such locally ordered regions, which are referred to here as folding units, associate to form the whole protein molecule, or in the case of some of the larger proteins, to form domains.

Diagrammatic representation of protein structure

Protein conformations are very complicated: it is not easy to comprehend the three-dimensional structure of a single protein, let alone compare many such structures. We have chosen to use a schematic two-dimensional representation similar to that used by certain other workers[1,2], and referred to here as topology/packing diagrams. The following rules were used in preparing our topology/packing diagrams for each protein. First, the α-helical and β-sheet chain segments are identified. As a β sheet can be formed from pieces of chain that are distant along the sequence, we use 'β' strand to refer to a single piece of chain that forms one strand of a β sheet. Second, a viewing direction is defined so that most segments of secondary structure are viewed end on. Third, the protein is rotated about the line of sight until the β strands lie in a horizontal plane (the twist of the β sheet is removed by flattening the sheet), and the front end of each β strand is drawn as a rectangle in the diagram. (This step is omitted if there is no β sheet.) Fourth, a circle representing the front end of each α helix is drawn, taking care to put segments that are close in space close together in the diagram. Finally, the segments are connected by bold or thin arrows (from the N to C terminal) that indicate whether the connection is at the near or far end, respectively, of the α helix or β strand. The scale of the real protein is preserved by making the separation of interacting α helices 10 Å, of hydrogen-bonded β strands 5 Å, and of other interacting β strands 10 Å. The diameter of the α-helix circle is 5 Å, and the β-strand rectangle is 4 × 5 Å.

Not all known protein conformations are included in the topology/packing diagrams given here (Figs 1–4). In some cases a particular structure is omitted as it is closely related to a protein that is shown: we show only one example of the immunoglobulin family[8–11], the trypsin family[41,42] and the haemoglobin family[43,44]. For a few proteins we could not find sufficiently clear pictures in the literature to be able to produce the diagrams (high potential iron protein[45], cytochrome c_2 (ref. 46), the catalytic part of alcohol dehydrogenase[30], malate dehydrogenase[47], ferredoxin[48], rhodenase[49], carbonic anhydrase[50] and soya bean trypsin inhibitor[51]). As these omitted proteins seem quite normal and fall into the present classification their omission should not affect our study.

Four clearly defined classes

The topology/packing diagrams of the 31 proteins are arranged into four classes defined as follows: (I) all-α proteins have only α-helix secondary structure (Fig. 1); (II) all-β proteins have mainly β-sheet secondary structure (Fig. 2); (III) α+β proteins have α-helix and β-strand secondary structure segments that do not mix but tend to segregate along the polypeptide chain (Fig. 3), and (IV) α/β proteins have mixed or approximately alternating segments of α-helical and β-strand secondary structure (Fig. 4).

Class I proteins are built up from α helices: more than 60% of the residues adopt the helical conformation. Because strongly interacting α helices are not always parallel (especially if the helices are short), the topology/packing diagrams do not show the packing of helices very accurately (Fig. 1). Of the three α-helical proteins shown, myohaemerythrin is most accurately represented as all the helices are parallel to the line of sight. The same packing of almost parallel α helices has

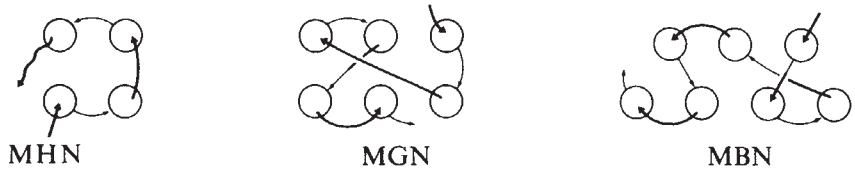


Fig. 1 Topology/packing diagrams of the following three all-α proteins (three-letter abbreviations have been assigned to all the proteins presented here): myohaemerythrin (MHN)[3]; myogen (MGN)[4], and myoglobin (MBN)[5,6]. As the axes of adjacent α helices are not always parallel and are sometimes at right angles, the diagrams cannot show the three-dimensional arrangement of helices very well.

©1976 Nature Publishing Group

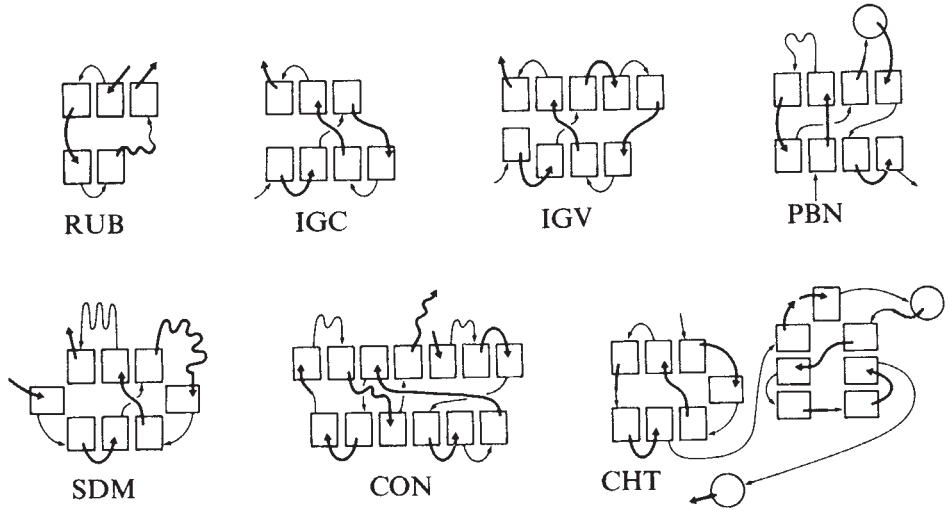


Fig. 2 Topology/packing diagrams of the following seven all-β proteins in order of increasing size: rubredoxin (RUB)[7]; immunoglobulin constant region (IGC)[8,9]; immunoglobulin variable region (IGV)[10,11]; prealbumin (PBN)[12]; superoxide dismutase (SDM)[13]; concanavalin A (CON)[14,15], and chymotrypsin (CHT)[16,17]. If the rectangles representing two β strands are very close together then the strands are hydrogen-bonded together. In some cases the associations are less clear in the diagrams: the final β strand of IGV also hydrogen bonds to the first strand; the first strand of SDM hydrogen bonds to both the second and final strands, and the fifth strand hydrogen bonds to both the fourth and sixth strands; in each domain of CHT, the second strand hydrogen bonds to both the first and third strands. The approximate twofold symmetry relationship between the upper and lower β sheets of the immunoglobulin regions and superoxide dismutase is clear (the twofold axis lies horizontally between the sheets).

been found in low resolution X-ray studies of tobacco mosaic virus coat protein[52,53] and electron microscopy of purple membrane protein[54].

Class II proteins (Fig. 2) are built from β sheets stacked to form a layered structure. Although the number of β strands in each sheet varies from two (in rubredoxin) to seven (in concanavalin A), there are always two layers of β sheet. Adjacent β strands in these proteins run in opposite directions to form antiparallel sheets. All these class II proteins, except chymotrypsin, have only one domain, and in chymotrypsin the two domains have identical topological connections. Often the strands that occur near the ends of the polypeptide chain are positioned in the middle of the β sheet so that they are well stabilised with hydrogen-bond associations to two neighbouring β strands. In three cases, the β strand at the edge of one β sheet also hydrogen bonds to the other β sheet closing one edge of the double layer (in the immunoglobulin variable region, superoxide dismutase and chymotrypsin), and in one of these (superoxide dismutase) both ends of the double layer are closed to form a barrel of β strands. The close similarity of the chain fold in the immunoglobulin variable region and superoxide dismutase, which has been pointed out before, is clear in Fig. 2.

Class III, the α+β proteins (Fig. 3), consist of a mixture of all-α and all-β regions within the same polypeptide chain. Often there is a cluster of helices at one or both ends of the β sheet, which is almost always built up from antiparallel strands. Only the three largest α+β proteins can be split into two relatively stable domains, and in each case one domain is mainly α helical while the other is mainly β sheet (T4 lysozyme, papain and thermolysin).

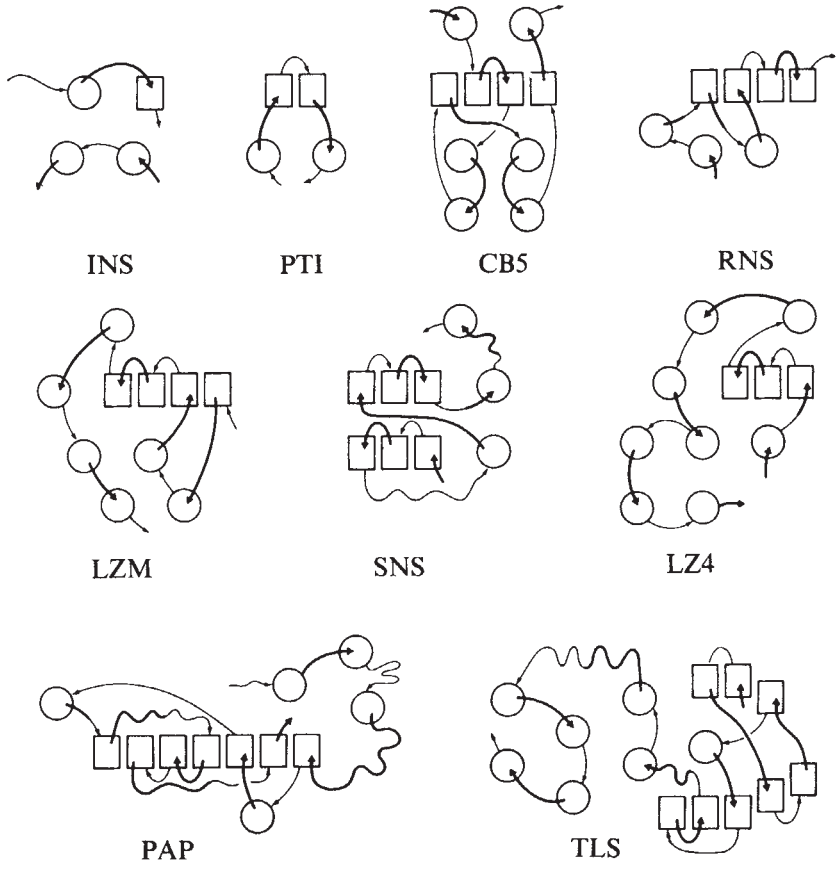


Fig. 3 Topology/packing diagrams for ten α+β proteins in order of increasing size: insulin (INS)[18]; pancreatic trypsin inhibitor (PTI)[19]; cytochrome b_5 (CB5)[20]; ribonuclease (RNS)[21]; hen lysozyme (LZM)[22]; staphyloccocal nuclease (SNS)[23]; T4 lysozyme (LZ4)[24]; papain (PAP)[25], and thermolysin (TLS)[26]. The diagrams show that the two lysozyme structures are more similar than thought previously[24]; analogy with the mammalian enzyme supports the idea[24] that the active residues of T4 lysozyme must be a glutamic acid at the end of the α helix preceding the β sheet and an aspartic acid at a bend in the β sheet (Glu 11 and Asp 20, respectively).

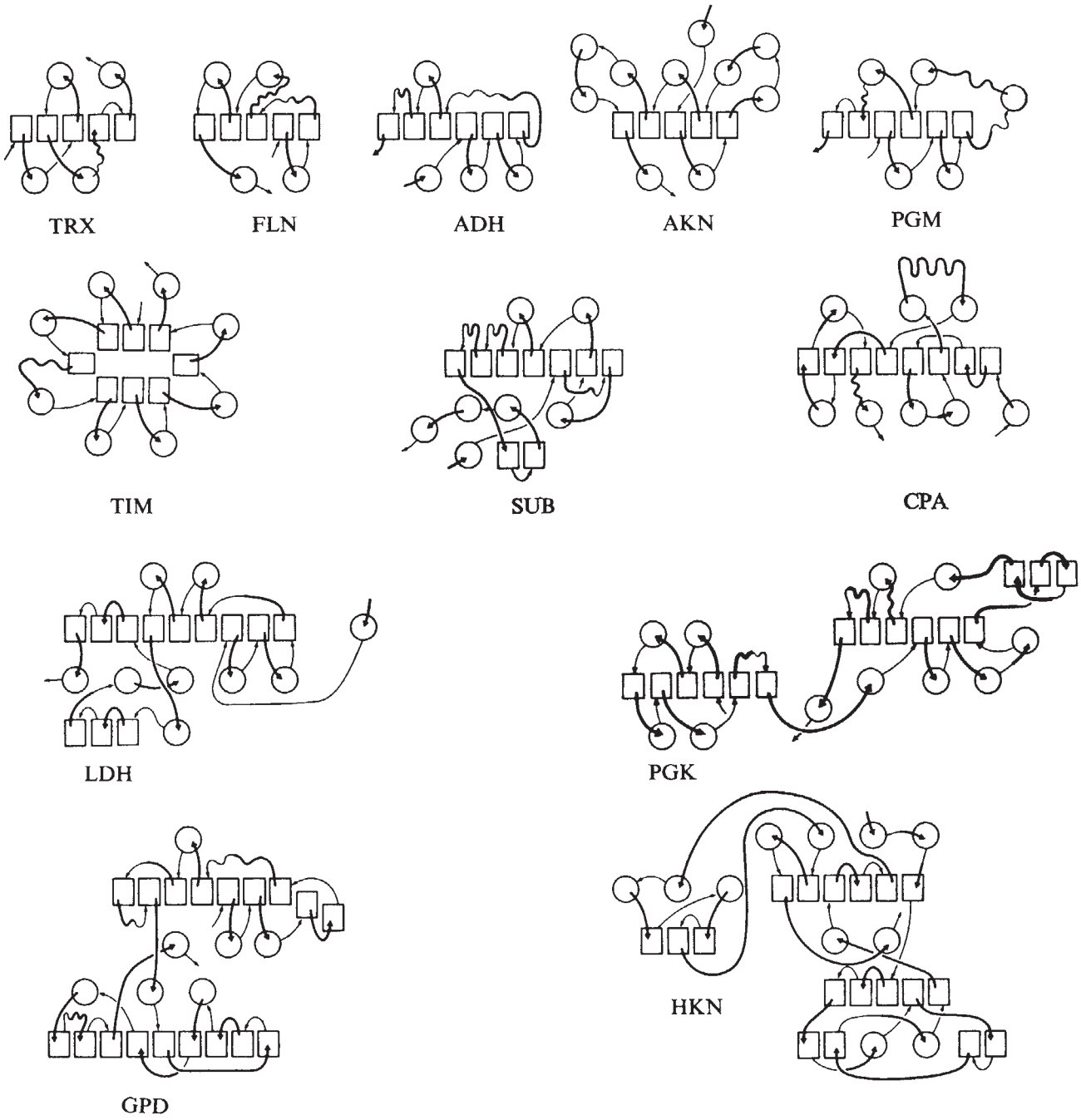
©1976 Nature Publishing Group

Class IV proteins, the α/β proteins (Fig. 4), have α helices and β strands that occur one after the other so that most α helices are separated by β strands along the sequence and vice versa. Most of these proteins have a single sheet surrounded by α helices, but in some of the larger proteins there are extra, smaller β sheets. The α helices pack on both sides of the β sheet, with α helices that follow one another along the chain, often on the same side of the β sheet. The main β sheet of each protein has between five and nine strands and consists mainly of parallel strands. The extra β sheets are smaller, often antiparallel, and more like the β sheets of the α+β proteins in class III (SUB and LDH in Fig. 4). Only the five largest of these α/β proteins (more than 300 residues) can be separated clearly into two or three domains. A common feature of these proteins is the order of the β strands in the β sheet: in most cases the first strand is in the middle of the sheet, with the following strands along the chain added first to the right and then to the left of the first strand. Because most of the α/β proteins shown here are enzymes in the glycolytic pathway and have to bind a common coenyzyme, NAD, some of the common features in Fig. 4 may be a result of these proteins having evolved from a common ancestor[2].

General features of known proteins

The interior hydrophobic core so characteristic of globular proteins is generally formed by contacts between and among α helices and β strands. The pieces of chain (often quite short) that connect these regions of secondary structure are generally

Fig. 4 Topology/packing diagrams of 12 α/β proteins in order of increasing size: thioredoxin (TRX)[27]; flavodoxin (FLN)[28,29]; alcohol dehydrogenase coenzyme domain (ADH)[30]; adenyl kinase (AKN)[32]; phosphoglycerate mutase (PGM)[31]; triose phosphate isomerase (TIM)[33]; subtilisin (SUB)[34]; carboxypeptidase (CPA)[35]; lactate dehydrogenase (LDH)[36]; phosphoglycerate kinase (PGK)[37,38]; D-glyceraldehyde-3-phosphate dehydrogenase (GPD)[39]; and hexokinase (HKN)[40]. Because the upper and lower domains of GPD and HKN are not directly under the upper domains, the interactions are not as strong as implied by the figures. The arrangement of the lower domain of HKN must be regarded as only tentative as it was determined from a mono drawing of this very large protein[40].



©1976 Nature Publishing Group

exposed to solvent and contain most of the hydrophilic and bend-promoting residues. Most proteins can be considered crudely as a layered sandwich structure, with each layer consisting entirely of either α helices or a β sheet. In the all-α proteins or all-β proteins, there are only two such layers. In the α+β proteins there are also often two layers, and often one layer consists of α helices and the other is a β sheet, though more mixed arrangements occur. In the α/β proteins, there are three layers, with a layer of α helices on each side of the central β sheet. Triose phosphate isomerase may seem an exception in that there are four layers with a central β-sheet sandwich surrounded on both sides by α helices, but this protein can also be considered as a layer of β sheet and a layer of α helices rolled into a cylinder so that every β strand hydrogen bonds to two others. Although β strands are sometimes surrounded on all sides by other segments of secondary structure, α helices are almost always only partially buried. It is also rare to find an α helix that is not in contact with at least one other α helix.

Very often several secondary structure segments that are adjacent along the polypeptide chain also interact strongly in three dimensions to form what is defined here as a 'folding unit'. There are 12 possible combinations of α helices and β strands into sequences of two or three adjacent segments of secondary structure: (αα), (αβ), (βα), (ββ), (ααα), (ααβ), (αβα), (βαα), (αββ), (βαβ), (ββα) and (βββ). Not all these secondary structure segments interact to form a stable complex. Those combinations containing a single β strand must be excluded as that strand can neither interact strongly with an α helix nor hydrogen bond to another β strand in the folding unit to form a β sheet. The most important folding units, therefore, consist of the following groups of adjacent secondary structure segments: (αα), (ββ), (βββ) and (βαβ). The (αα) folding unit consists of a pair of α helices adjacent in sequence that are arranged with their axes approximately antiparallel and are in van der Waals' contact (Fig. 5*a*). The (ββ) folding unit consists of two β strands that fold back and hydrogen bond together into a β sheet with two antiparallel strands (Fig. 5*b*). The (βββ) folding unit is just an extension of the (ββ) folding unit with an extra β strand forming a β sheet with three antiparallel strands arranged in a simple zigzag. The (βαβ) folding unit consists of a β sheet with two parallel strands in van der Waals' contact with an α helix antiparallel to these strands (Fig. 5*c*). Three other possible groups, (ααα), (αββ) and (ββα), are rare in globular proteins, difficult to identify unambiguously and can be considered as formed from (αα) or (ββ) folding units.

Table 1 gives the number of times the different folding units occur in all the globular proteins considered here. For some of the α/β proteins (class IV) adjacent (βαβ) folding units have a β strand in common. As folding units are defined here as independent subassemblies of secondary structure, a particular segment of secondary structure should not be part of two folding units. A new type of folding unit, referred to as (βαβ)′, was introduced to take care of the case (βαβαβ), that is, where a (βαβ) folding unit shares a β strand with an adjacent (βαβ) unit. For example, triose phosphate isomerase is formed from eight (βαβ)′, folding units.

In the three all-α proteins there are eight (αα) folding units. If every α helix in these three proteins was in contact with the adjacent helices along the sequence, the maximum possible number of (αα) folding units would also be eight. In the seven all-β proteins there are 20 (ββ) and four (βββ) folding units, respectively; the maximum numbers that could be formed are 24 (ββ) and six (βββ). In the nine α+β proteins, there are 12 (αα), five (ββ) and six (βββ) folding units, respectively; the maximum numbers that could occur for these proteins are 16 (αα), seven (ββ) and six (βββ). In the twelve α/β proteins there are six (αα), 10 (ββ), four (βββ), 18 (βαβ) and 12 (βαβ)′, folding units, respectively; the maximum possible numbers are 12 (αα), 12 (ββ), four (βββ), 23 (βαβ) and 15 (βαβ)′.

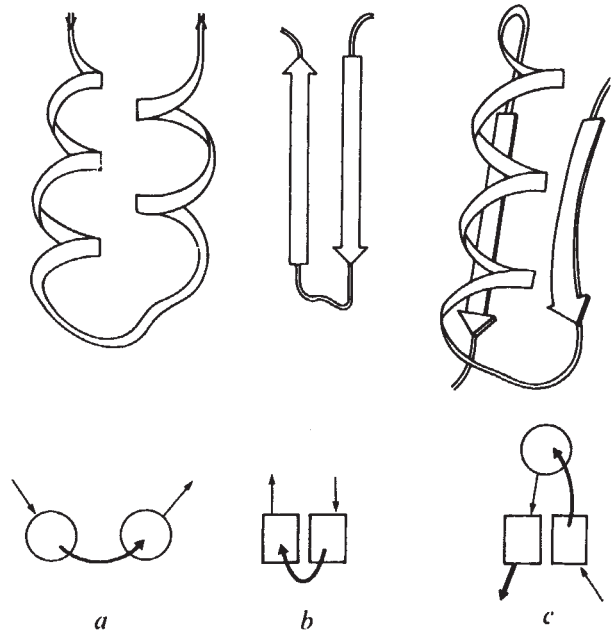
The above results show that for each class of protein the actual number of folding units observed is close to the maximum possible number. For all the 31 proteins together there are 26 (αα), 35 (ββ), 14 (βββ), 18 (βαβ) and 12 (βαβ)′ folding units, respectively. A total of 242 segments of secondary structure belong to one or other of these folding units out of a total of 361 segments (67%). It is interesting that in the all-β proteins more β strands are in (ββ) rather than (βββ) folding units, whereas in the α+β proteins more β strands occur in (βββ) rather than (ββ), folding units.

Rao and Rossmann[55] concluded from an examination of four proteins that a structure consisting of three parallel β strands and two joining α helices was a common structural building block in proteins. (This structure consists of two overlapping (βαβ)′ folding units.) They also noticed that the structure had a unique hand, in that the polypeptide chain is directed from β to α to β in a clockwise sense (Figs 4 and 5*c*). Sternberg and Thornton[56] have confirmed this handedness from a study of many protein structures and concluded that it arises from the twist of the β sheet. In another analysis of β sheets in protein structures, Richardson *et al.*[57] also noticed the high frequency of (ββ) and (βαβ) folding units.

Statistical significance of patterns

Another way to assess the statistical significance of the arrangement of secondary structure in known proteins is to estimate the probability that a particular observed pattern would occur by chance. As the topology/packing diagrams represent the arrangement in space of α helices less well than of β strands, we first considered the statistical significance of the connectivity of the strands in the β sheets of proteins in classes II to IV. A computer program was used to generate random permutations of the order of the different β strands along the sequence while preserving the relative positions of the strands in the β sheets of the particular protein (the direction of the chain in the segments was not taken into account). For each of the 1,000 permutations generated for a particular protein, we counted the number of times a pair of β strands adjacent in the permuted sequence were also in contact in the β sheet (referred to as the number of adjacent contacts). This same

Fig. 5 The folding of the polypeptide chain in the three commonly occurring folding units: (αα), (ββ) and (βαβ). The ribbon illustrates the path of the backbone with the arrows directed from the N to C terminals. Below each drawing of the chain fold is shown the representation of the particular folding unit used in the topology/packing diagrams of Figs 1–4.



©1976 Nature Publishing Group

Table 1 Folding unit counts, β-sheet patterns, β-strand polarity and chance probability of 31 proteins

Class and protein	No. of folding units (αα)	(ββ)	(βββ)	(βα)	(βαβ)′	β-Sheet pattern†	Strand polarity* Np	Nap	Probability
All-α									
Myohaemerythrin	2								0.33
Myogen	3								0.42
Myoglobin	3								0.14
All-β									
Rubredoxin		2				−+− +−		3	0.37
Ig constant		3				−+− −+−+		5	0.14
Ig variable		3	1			−+−+− −+−+		7	0.006
Prealbumin		2				−+−+ +−−+	1	5	0.16
Superoxide dismutase		3				−+− + + −+−		8	0.007
Concanavalin A		2	2			+−+−+−+ +−+−+−		11	0.001
Chymotrypsin		5	1			−+− + + − −+− + + − −		12	0.05
α+β									
Insulin	1								1.0
Trypsin inhibitor		1				+−		1	1.0
Cytochrome b5	2	1				−−+−	1	2	0.93
Ribonuclease	1		1			−+−+		3	0.08
Lysozyme (hen)	2		1			+−+−		3	0.51
Staph.nuclease			2			+−+ +−+		4	0.006
Lysozyme (T4)	3		1			+−+		2	0.33
Papain	1		1			−−+−+−+	1	5	0.54
Thermolysin	3	3				−++ −+++−	3	3	0.17
α/β									
Thioredoxin		1		1		−−−+−	2	2	0.56
Flavodoxin				2	1	−−−−−	4		0.15
Alcohol dehydrog.				2	1	−−−−−−	5		0.21
Adenylade kinase	2					−−−−−	4		0.56
Phosphoglycerate mutase		1		1		−+−−−− −−−	3	2	0.59
Triose phosphate isomerase					8	− − −−−	8		0.001
Subtilisin	1	1		2		−−−−−−− +−	6	1	0.24
Carboxypeptidase		1		1		++−−−−−+	4	3	0.25
Lactate dehydrog.	1		2	2	2	−+−−−−−− −+−	6	4	<0.001
Phosphoglycerate kinase	1			4		−−−−−− +−+ −−−−−+	9	3	<0.001
Glyceraldehyde phosphate dehydrog.		3		3		−−−−−−−−+ ++−+−++−+	9	7	<0.001
Hexokinase	1	2	2	1		+− ++ −−−+− +−+ +−+−−−	5	8	0.003

*The numbers of parallel β-strand associations in the β sheet(s) are given in the column headed *N*p, and the number of antiparallel associations are given in the column headed *N*ap.

†The β-sheet pattern is taken from Figs 2–4 with a plus sign to mark β strand that runs away from the viewer (into the paper), and a minus sign to mark a β strand that runs towards the viewer (out of the paper).

number was obtained for the native protein arrangement from Figs 1–4. As the α helices are ignored, β strands in contact in the sheet but separated in sequence by only one α helix were also counted as adjacent contacts. The statistical significance of the native arrangement was taken as the probability that a randomly generated permutation of strand order in the particular β sheet would have at least as many adjacent contacts as the native arrangement. These probabilities (P) are given in the last column of Table 1 and vary from 1.0 to 0.001 with a geometric mean of 0.05. For the smallest β sheet consisting of only two β strands, any permutation of the strands will leave them in contact in the sheet so the probability of having at least one adjacent contact is 1.0 (for example, trypsin inibitor, Fig. 3). For a β sheet with three strands, there is a one-third chance of forming an antiparallel arrangement with two adjacent contacts (for example, T4 lysozyme, Fig. 3). As the size of the β sheet increases the chances of having many such adjacent contacts decreases. For example, the β sheet in both cytochrome b_5 and ribonuclease consists of four strands (Fig. 3), but the arrangement in the former protein is almost random (one adjacent contact, $P = 0.33$), whereas that in the latter is highly significant (three adjacent contacts, $P = 0.08$). One of the three α/β proteins with a five-stranded β sheet (Fig. 4), flavodoxin has a more significant pattern (three adjacent contacts, $P = 0.15$) than the other two (both thioredoxin and adenylate kinase have two adjacent contacts each and $P = 0.56$). The patterns found in the proteins with the biggest β sheets are very significant, with P values less than 0.01, and in the case of some of the dehydrogenases, less than 0.001. These results show that β strands that are adjacent in

©1976 Nature Publishing Group

sequence are in van der Waals' or hydrogen-bonding contact much more often than expected by chance. Had the connectivity of the α helices also been considered, the observed patterns would be even more significant as then there would be many more segments of secondary structure to permute. The significance of the patterns of the all-α proteins was also estimated in this way but now the helices could be in contact both vertically and horizontally in the topology/packing diagrams. The results of this calculation (Table 1) show that none of these patterns of the all-α proteins is statistically very significant.

Parallel and antiparallel β sheets

Table 1 also shows the polarity of the β strands in the β sheets of the proteins considered here. For the all-β proteins (class II, Fig. 2), only one out of the 50 hydrogen-bonding associations in the β sheets is between a pair of chains that run in the same direction (parallel) rather than in opposite directions (antiparallel); it is in prealbumin. For the α+β proteins (class III, Fig. 2) most of the β strands also form antiparallel rather than parallel associations (23 and five occurrences, respectively). On the other hand, the β sheets of the α/β proteins (class IV, Fig. 4) have many more parallel associations than antiparallel ones (64 and 21 occurrences, respectively). When the β strands are grouped together along the sequence and not separated by α helices (as in the all-β and α+β proteins), (ββ) and (βββ) folding units occur very often leading to antiparallel β sheets. When the β strands are separated by α helices along the sequence (as in the α/β proteins), (βαβ) and (βαβ)′ folding units occur most often leading to parallel β sheets.

Conclusions and implications

In the past the word domain has been used to describe substructures in protein molecules. This use derives from the observation of Phillips[58] who described the structure of lysozyme in terms of a series of "compact globular units". The word domain has also been used to describe the independent globular regions that can occur when a single polypeptide chain is formed by gene fusion or gene duplication (for example, the four domains of the immunoglobulin heavy chains). The use of the same word "domain" to define both types of protein substructure leads to confusion, and we suggest that the phrase 'folding unit' be used to define small assemblies of secondary structure segments that are adjacent in sequence and in van der Waals' or hydrogen-bonding contact with one another. The word domain is then reserved for large subassemblies that would be stable if the polypeptide chain connecting them to the rest of the protein molecule were to be cleaved[59]. Each domain has all the characteristics of a complete globular protein; often the different domains of a multidomain protein have different functions[60].

Our analysis of known protein structures using simplified topology/packing diagrams has shown that the observed arrangements of α helices and β strands are statistically very significant. The classification of 31 protein structures into four clearly separated classes has made it possible to identify common structural patterns, and has shown how the arrangement of segments of secondary structure along the sequence relates to three-dimensional properties such as parallel and antiparallel β sheets. Very often segments of secondary structure adjacent in sequence form a few well defined types of folding units. The high frequency of folding units in globular proteins is probably a result of the kinetic pathway of protein assembly rather than the stability of the final folded form. If certain segments of secondary structure existed with even marginal stability when the native conformation was unfolded, those segments that were close along the polypeptide chain would have a higher probability of diffusing together to form folding units[61]. If these folding units were then to associate rapidly to form the native conformation, which might be in a kinetically determined minimum of the free energy, the high frequency of such units in native proteins would be expected. The idea that segments of secondary structure close in sequence interact to form subassemblies which then associate to form the native structure has been used by Ptitsyn and Rashin to study the folding pathway of myoglobin[62]. Wetlaufer[63] used a similar kinetic argument to explain the domain structure of proteins. Other explanations of the high frequency of folding units in globular proteins do not depend on kinetic arguments: for example, the native conformation could be stabilised if its segments of secondary structure were to be connected up by short pieces of chain that do not cross one another. The chance probability that two evolutionary unrelated proteins have similar arrangements of secondary structure will be less than thought previously[1], for we have shown that all protein conformations are built up from the same three basic folding units and that the conformations fall into four well defined classes.

Thus the picture of protein folding suggested by the final protein structure is as follows: pieces of secondary structure first diffuse together to form folding units that then associate to form the native structure, or in the case of the larger proteins, the domains which subsequently interact weakly to form the native structure.

Received March 18; accepted April 26, 1976.

1 Schulz, G. E., and Schirmer, R. H., *Nature*, **250**, 144–145 (1974).
2 Rossman, M. G., Moras, D., and Olsen, K. W., *Nature*, **250**, 194–199 (1974).
3 Hendrickson, W. A., Klipperstein, G. L., and Ward, K. B., *Proc. natn. Acad. Sci. U.S.A.*, **72**, 2160–2164 (1975).
4 Kretsinger, R. H., and Nuckolds, C. E., *J. biol. Chem.*, **248**, 3313–3326 (1973).
5 Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., and Davies, D. R., *Nature*, **185**, 422–425 (1960).
6 Watson, H. C., in *Prog. Stereochem.*, **4**, 299–333 (1960).
7 Watenpaugh, K. D., Sieker, L. C., Herriott, J. R., and Jensen, L. H., *Acta Crystallogr.*, **B29**, 943–956 (1973).
8 Schiffer, M., Girling, R. L., Ely, K. R., and Edmundson, A. B., *Biochemistry*, **12**, 4620–4631 (1973).
9 Poljak, R. J., *et al.*, *Proc. natn. Acad. Sci. U.S.A.*, **70**, 3305–3310 (1973).
10 Segal, D. M., *et al.*, *Proc. natn. Acad. Sci. U.S.A.*, **71**, 4298–4302 (1974).
11 Epp, C., *et al.*, *Eur. J. Biochem.*, **45**, 513–524 (1974).
12 Blake, C. C. F., Geisow, M. J., Swan, I. D. A., Rerat, C., and Rerat, B., *J. molec. Biol.*, **88**, 1–12 (1974).
13 Richardson, J. S., Thomas, K. A., Rubin, B. H., and Richardson, D. C., *Proc. natn. Acad. Sci. U.S.A.*, **72**, 1349–1353 (1975).
14 Edelman, G. M., *et al.*, *Proc. natn. Acad. Sci. U.S.A.*, **69**, 2580–2584 (1972).
15 Hardman, K. D., and Ainsworth, C. F., *Biochemistry*, **11**, 4910–4919 (1972).
16 Birktoft, J. J., Blow, D. M., Henderson, R., and Steitz, T. A., *Proc. R. Soc.*, **B257**, 67–76 (1970).
17 Birktoft, J. 4., and Blow, D. M., *J. molec. Biol.*, **68**, 187–240 (1972).
18 Adams, M. J., *et al.*, *Nature*, **224**, 491–495 (1969).
19 Huber, R., Kukla, D., Ruhlmann, A., and Steigemann, W., *in Proc. Int. Res. Conf. Proteinase Inhibitors, Munich* 1970 (edit. by Fritz, H., and Tschesche, H.), 56–64 (de Gruyter, Berlin, 1970).
20 Mathews, F. S., Levine, M., and Argos, P., *Nature new Biol.*, **233**, 15–16 (1971).
21 Wyckoff, H. W., *et al.*, *J. biol. Chem.*, **245**, 305–328 (1970).
22 Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R., *Proc. R. Soc.*, **B167**, 365–385 (1967).
23 Arnone, A., *et al.*, *J. biol. Chem.*, **246**, 2302–2316 (1971).
24 Matthews, B. W., and Remington, S. J., *Proc. natn. Acad. Sci. U.S.A.*, **71**, 4178–4182 (1975).
25 Drenth, J., Jansonius, J. M., Koekoek, R., and Wolthers, B. G., *Adv. Protein Chem.*, **25**, 79–115 (1971).
26 Colman, P. M., Jansonius, J. N., and Matthews, B. W., *J. molec. Biol.*, **70**, 701–724 (1972).
27 Holmgren, A., Soderberg, B. -O., Elkund, H., and Branden, C. -I., *Proc. natn. Acad. Sci. U.S.A.*, **72**, 2307–2309 (1975).
28 Watenpaugh, K. D., Sieker, L. C., Jensen, L. H., Legall, J., and Dubourdieu, M., *Proc. natn. Acad. Sci. U.S.A.*, **69**, 3185–3188 (1972).
29 Anderson, R. D., *Proc. natn. Acad. Sci. U.S.A.*, **69**, 3189–3191 (1972).
30 Branden, C. I., *et al.*, *Proc. natn. Acad. Sci. U.S.A.*, **70**, 2439–2442 (1973).
31 Campbell, J. W., Watson, H. C., and Hodgson, G. I., *Nature*, **250**, 301–303 (1974).
32 Schulz, G. E., Elzinga, M., Marx, F., and Schirmer, R. H., *Nature*, **250**, 120–123 (1974).
33 Banner, D. W., *et al.*, *Nature*, **255**, 609–614 (1975).
34 Wright, C. S., Alden, R. A., and Kraut, J., *Nature*, **221**, 235–242 (1969).
35 Quiocho, F. A., and Lipscomb, W. N., *Adv. Protein Chem.*, **25**, 1–78 (1971).
36 Adams, M. J., Ford, G. C., Liljas, A., and Rossmann, M. G., *Biochem. biophys. Res. Commun.*, **53**, 46–51 (1973).
37 Blake, C. C. F., and Evans, P. R., *J. molec. Biol.*, **84**, 585–603 (1974).
38 Blake, C. C. F., in *Essays in Biochemistry*, **11**, 37–79 (edit. by Campbell, P. N. and Aldridge, W. N.) (Academic, London 1975).
39 Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., and Rossmann, M. G., *Proc. natn. Acad. Sci. U.S.A.*, **70**, 3052–3054 (1973).
40 Fletterick, R. J., Bates, D. J., and Steitz, T. A., *Proc. natn. Acad. Sci. U.S.A.*, **72**, 38–42 (1975).
41 Stroud, R. M., Kay, L. M., and Dickerson, R. E., *J. molec. Biol.*, **83**, 185–208 (1974).
42 Shotton, D. M., and Watson, H. C., *Nature*, **225**, 811–816 (1970).
43 Perutz, M. F., *et al.*, *Nature*, **222**, 1240–1244 (1968).
44 Baldwin, J. M., *Prog. Biophys. molec. Biol.*, **29**, 225–320 (1975).
45 Carter, C. W., Jr, *et al.*, *J. biol. Chem.*, **249**, 4212–4225 (1974).
46 Salemme, F. R., *et al.*, *J. biol. Chem.*, **248**, 3910–3921 (1973).
47 Hill, E., Tsernoglou, Webb, L., and Banaszak, L. J., *J. molec. Biol.*, **72**, 577–591 (1972).
48 Adman, E. T., Sieker, L. C., and Jensen, L. H., *J. biol. Chem.*, **248**, 3987–3996 (1973).
49 Bergsma, J., *et al.*, *J. molec. Biol.*, **98**, 637–643 (1965).
50 Kannan, K. K., *et al.*, *Cold Spring Harb. Symp. quant. Biol.*, **36**, 221–231 (1971).
51 Sweet, R. M., Wright, H. T., Chothia, C. H., and Blow, D. M., *Biochemistry*, **13**, 4213–4228 (1974).
52 Champness, J. N., *et al.*, *Nature*, **259**, 20–24 (1976).
53 Holmes, K. C., Stubbs, G. J., Mandelkow, E., and Gallwitz, U., *Nature*, **254**, 192–196 (1975).

©1976 Nature Publishing Group

54 Henderson, R., and Unwin, P. N. T., *Nature*, 257, 28–32 (1975).
55 Rao, S. T., and Rossmann, M. G., *J. molec. Biol.*, 76, 241–256 (1973).
56 Karplus, M., and Weaver D. L., *Nature*, 260, 404–406 (1976); Sternberg, M. J. E., and Thornton, J. M., *J. molec. Biol.* (in the press).
57 Richardson, J. S., Richardson, D. C., Thomas, K. A., Silverton, E. W., and Davies, D. R., *J. molec. Biol.*, 102, 221–235 (1976).
58 Phillips, D. C., *Proc. natn. Acad. Sci. U.S.A.*, 57, 484–495 (1967).
59 Goldberg, M. E., *J. molec. Biol.*, 46, 441–446 (1969).
60 Veron, M., Falcoz-Kelly, F., and Cohen, G. N., *Eur. J. Biochem.*, 28, 520–527 (1972).
61 Light, A., Taniuchi, H., and Chen, R. F., *J. biol. Chem.*, 249, 2285–2299 (1974).
62 Ptitsyn, O. B., and Rashin, A. A., *Biophys. Chem.*, 3, 1–20 (1974).
63 Wetlaufer, D. E., *Proc. natn. Acad. Sci. U.S.A.*, 70, 697–701 (1974).

Identification of an enzyme in platelet microsomes which generates thromboxane A_2 from prostaglandin endoperoxides

P. Needleman, S. Moncada*, S. Bunting & J. R. Vane
Wellcome Research Laboratories, Beckenham, Kent BR3 3BS, UK

M. Hamberg & B. Samuelsson
Department of Chemistry, Karolinska Institutet, S-104 01 Stockholm 60, Sweden

The microsomal fraction of horse and human platelets contains an enzyme which converts prostaglandin cyclic endoperoxides (PGG_2 or PGH_2) to a substance which is much more potent in contracting strips of rabbit aorta. This substance has the same characteristics as thromboxane A_2, and can be distinguished from other products of arachidonic acid metabolism by differential bioassay.

PIPER and Vane[1] have detected the release of a substance during anaphylaxis in isolated lungs from sensitised guinea pigs which they called "rabbit aorta-contracting substance" or RCS. The half life of RCS was <2 min at room temperature[2] and its release was blocked by aspirin-like drugs. RCS contracts all isolated vascular tissues tested from several species[2] and is released from lung tissue also by bradykinin, RCS–RF[2] SRS–C[3], arachidonic acid[3,4] and mechanical agitation or stimulation[2,4]. RCS is also generated by mechanical stimulation of rabbit spleen slices[5] and by incubation of arachidonic acid with rabbit spleen microsomes[5]. In addition, platelet aggregation is accompanied by the generation of an RCS[6–10]. RCS release has also been demonstrated *in vivo*, either by challenging sensitised guinea pigs with antigen or by infusion of bradykinin[4].

The release of RCS by the prostaglandin precursor, arachidonic acid, and the prevention of release by inhibition of prostaglandin biosynthesis led to the suggestion that RCS is an intermediate in prostaglandin biosynthesis[3,5]. A cyclic endoperoxide of arachidonic acid was postulated as the unstable intermediate in the biosynthesis of prostaglandins[11,12]. Endoperoxides have been generated and isolated after incubation of arachidonic acid with sheep vesicular glands[13,14]. The prostaglandin endoperoxides (now most frequently designated PGG_2 and PGH_2; Fig. 1) contract rabbit aorta, gastrointestinal and tracheal smooth muscle[15] and at low concentrations induce aggregation of platelets[10,16]. Aggregation induced by collagen, thrombin, adrenaline and arachidonic acid is also accompanied by release of endoperoxides[10,16,17]. The inhibitory effect of aspirin on the second phase of platelet aggregation[18] could therefore be due to inhibition of biosynthesis of platelet endoperoxides.

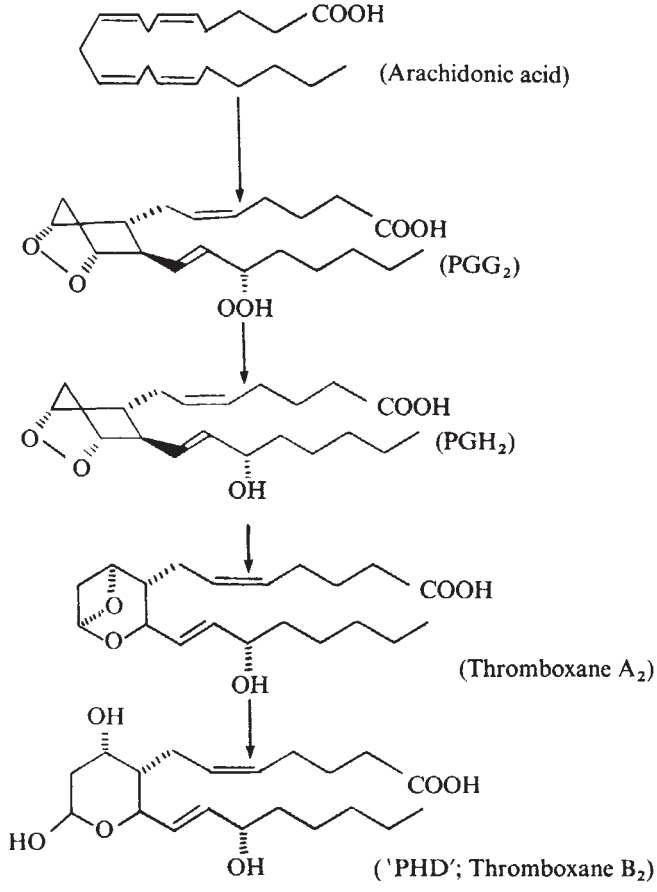
PGG_2 and PGH_2 are unstable, but their half lives are longer ($t_{\frac{1}{2}} \simeq 5$ min) than that of RCS[10,16]. Furthermore, quantitative analysis of the endoperoxide released from either guinea pig lung effluent or human platelets showed that only a minor part of the rabbit aorta-contracting activity was due to prostaglandin endoperoxides[10].

Samuelsson's group[19] has discovered a labile intermediate in the conversion of the endoperoxide PGG_2 into the hemiacetal derivative PHD (Fig. 1). They named the intermediate thromboxane A_2 (TXA_2) because of its structure and potent aggregating activity on platelets. This substance also strongly contracts rabbit aorta strips and its half life in aqueous solution is approximately 30 s. These properties of TXA_2 resemble very closely the properties of RCS, and it has been concluded that the activity of RCS is mainly due to TXA_2 (ref. 19).

Washed human platelets convert PGG_2 to TXA_2 (ref. 19). In the work reported here, we isolated from the same source and also from horse platelets an enzyme system which converts endoperoxides to a substance with the characteristics of TXA_2 and compared the relative contractile effects of endoperoxides and this substance on rabbit aorta.

Two litres of fresh citrated (3.8% w/v) blood from a horse were spun at 200*g* for 10 min to remove red blood cells. The

Fig. 1 Structures of some metabolites of arachidonic acid.



*To whom reprint requests should be addressed.

©1976 Nature Publishing Group