

(From Life, Wetlanfer (ed.), The Protein Folding Problem, Westview, 1984, pp. 1-28.

1. What Do the Folds in Proteins Look Like?

#19

Proteins are at the interface between chemistry and biology: on the one hand they are among the largest of chemically well-defined molecules, while on the other hand they are perhaps the smallest systems which display the complexity, quirkiness, and downright intractability of living systems. One reason for the fascination of the protein folding problem is that it represents an unusually concrete and limited case of the whole problem of reductionism. An unfolded protein is clearly a chemical object: a backbone of exactly repeated simple units each with one of an alphabet of 20 possible side chains. Its properties are relatively dull and are quite predictable by summing up the properties of its components. A folded protein, on the other hand, in addition to complexity and unpredictability, has acquired meaning: unity, controlled interaction with other systems, and biologically significant function. In many cases one can watch a protein undergo the spontaneous transition from randomness to directed functionality in the space of a few minutes. There is apparently no extra information hidden within the starting state, so we feel that understanding the rules of the transformation would teach us worthwhile lessons about hierarchical organization, cooperative properties, and exactly how an organic whole becomes so much more than a sum of its parts.

As an introduction to the general problem of protein folding, I would like to summarize some characteristics of the final folded state of proteins, trying to emphasize properties that seem likely to be related to the pathway by which they arrived there.

It is rather misleading, of course, to speak of a single folded state, because the native protein is a dynamic object that fluctuates around one or a few preferred conformations. The largest of those fluctuations are presumably related to the final steps of folding or the first steps of unfolding. Figure 1 shows all non-hydrogen atoms in the preferred conformation of pancreatic trypsin inhibitor (as determined by X-

From The Protein Folding Problem edited by D.B. Wetlanfer, etal. (c) 1984 by Westview Press. Permission to reprint granted by the publisher.

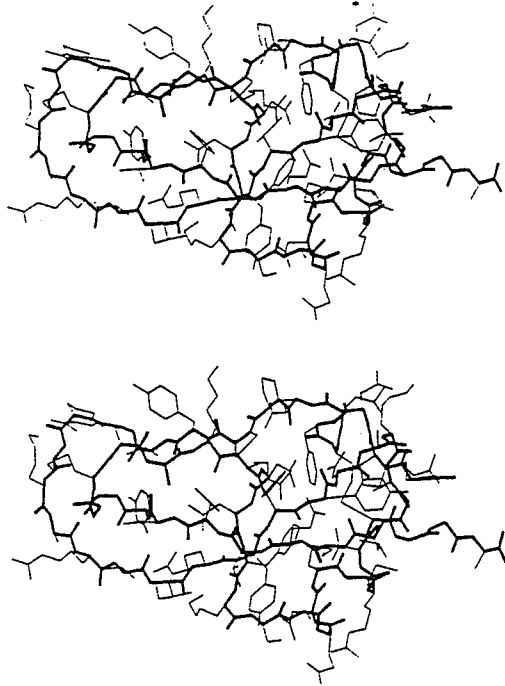


Figure 1. Stereo drawing of all non-hydrogen atoms in the x-ray crystal structure of pancreatic trypsin inhibitor (17). Backbone is in heavy lines and side chains in lighter lines. Coordinates for this and most other figures are from the Protein Data Bank (18).

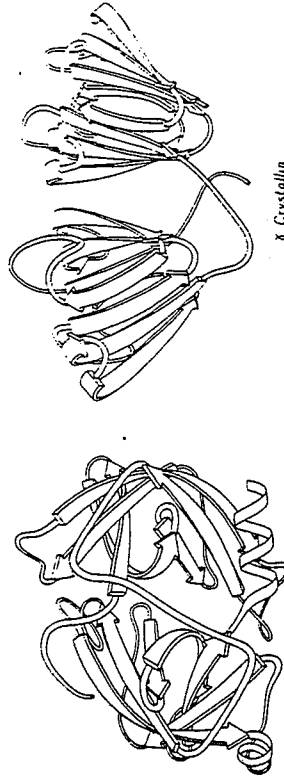


Figure 2. Proteins with two very similar domains. Left: elastase (19), with the domains closely packed into a spherical overall subunit. Right: gamma crystallin (20), with its two domains well separated.

ray crystallography), to illustrate the complexity and asymmetry of the native conformation. The stick-figure model allows one to look through the molecule and to visualize the covalent connectivity which is a crucial part of folding. However, the internal interactions of a protein with itself that constitute the folding process, and the external interactions with the environment that include its functionality, all occur at the surfaces of the atoms. There is a highly complementary interlocking of atom surfaces inside a protein, with only a little open space left. The final readjustments which allow everything to settle down in the position of best fit is presumably one of the last steps in folding, although once such complementarity is achieved it is not necessarily the first thing to be lost on unfolding.

The most obvious, crucial task that must be accomplished by the protein folding process is arrangement of the backbone into an approximately correct tertiary structure. Although those tertiary structures show a wide variety there is considerable order within that variety, some of which may be the result of restrictions on favorable folding pathways. Within the four major categories there are 9 or 10 rather specific structural patterns, which together account for about 3/4ths of the known protein structures (1). We will survey these common structural patterns, trying to recognize the footprints left there by folding requirements.

Large proteins are commonly divided into domains: contiguous stretches of the sequence which are also contiguous in 3 dimensions (2). Domains usually resemble other entire single-domain proteins. They sometimes can move as independent rigid bodies. Domains are often assumed (and in a few cases proven) to fold up independently and thereby to make the folding of large proteins fast enough to be biologically useful. The commonest domain size is 100 to 200 residues, but they vary from about 40 to at least 400. Figure 2 shows schematic backbone drawings of elastase and of gamma crystallin, each of which has two very similar domains. Since the domains within one protein are often completely different, however (as shown for papain in Figure 3), classification of tertiary structure is much more straightforward if done one domain at a time. Within each domain we will pay attention to the type, amount, and organization of secondary structure, the number and shape of the major layers of backbone versus hydrophobic cores, and the topology of connections among the individual chains in a layer. The structures are illustrated by highly simplified schematic drawings of the polypeptide backbone in which alpha-helices are shown as spiral ribbons, beta strands as arrows, non-repetitive loops as smoothed ropes and disulfides as zigzags. All the drawings are to the same scale and within each structural class they are usually shown from an equivalent direction of view.

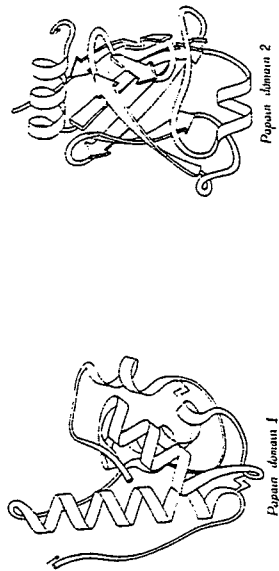


Figure 3. Papain (21), a protein with two entirely different domains. Left: domain 1, which is all helical. Right: domain 2, which is almost entirely a barrel of beta sheet.

248

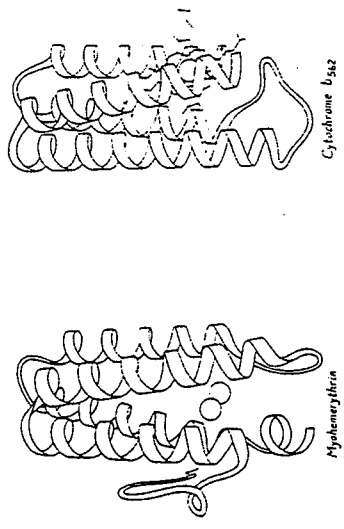


Figure 4. Up-and-down helix bundle structures (antiparallel alpha category): myohemerythrin (22) and cytochrome b562 (23).

The 4 major categories of tertiary structure are:

- I. Antiparallel alpha
- II. Parallel alpha/beta
- III. Antiparallel beta
- IV. Small irregular

It is worth noting, to begin with, that there are no all-alpha (or all-beta) structures in which the elements of each layer are parallel to one another. That would presumably be possible for a 2-layer structure in its final form, but would require the chain to fold by winding first up in one layer and then down in the other, like a ball of string. Apparently that does not happen in such a straightforward form, although later we will find evidence suggesting that more complex elements may fold by winding up as loops.

Antiparallel Alpha Proteins

Within the first major category, the antiparallel alpha, the simplest type of structure is the up-and-down helix bundle: an approximate cylinder of alpha helices with each connected to its nearest neighbor, so that the chain goes up one helix, moves to the next one and down it, over by one and up, etc., all around the cylinder. Many of these structures have only four helices, such as the myohemerythrin and cytochrome b562 shown in Figure 4. Bacteriorhodopsin is a bundle of 7 alpha helices spanning the membrane, and it probably has an up-and-down connectivity (see chapter 5). The prevalence of nearest-neighbor connections between elements of secondary structure, as seen in these up-and-down helix bundles, is one of the most general regularities in protein structure. This has frequently been explained (e.g., 2) by the simple statistical fact that two pieces of backbone can come together more rapidly if they are neighbors in the sequence than if they are widely separated.

The second simplest organization of antiparallel helices has at least one non-nearest-neighbor connection in which the chain goes across either the top or bottom of the cylinder of helices. This does not happen in any of the 4-helix bundles, but it may with 5 or more helices as shown for the hemoglobin beta chain and the second domain of thermolysin in Figure 5. In analogy to the beta barrels described below, this second pattern of antiparallel alpha structure is called a Greek key helix bundle. They are much less common than up-and-down helix bundles, they have no characteristic handedness, and the examples so far each include only one non-nearest-neighbor connection. The distribution of topologies found in helix bundle proteins is consistent with the hypothesis that alpha helices associate as near-neighbor pairs independently of what pairs have already formed.

In each of the major structural categories there is a

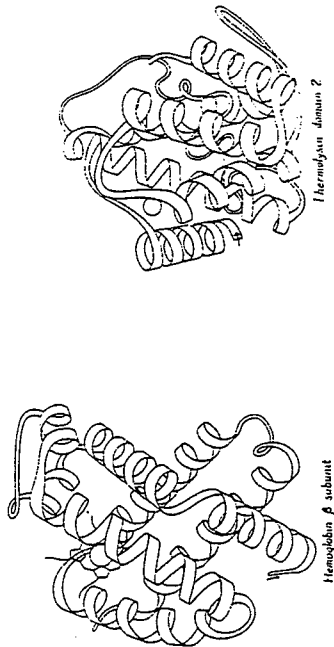


Figure 5. Greek key helix bundle structures (antiparallel alpha category): hemoglobin (24) and thermolysin domain 2 (25).

249

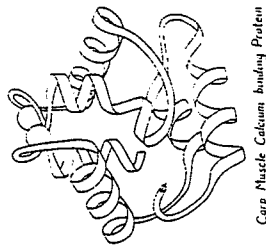


Figure 6. A miscellaneous antiparallel alpha structure: carp muscle calcium-binding protein (26).

small group of miscellaneous proteins that do not fit into a recognized subgroup. Within the antiparallel alpha category these include the carp-muscle Ca-binding protein (Figure 6), whose helices are arranged in almost perpendicular pairs with a Ca-binding loop in between them. There is evidence (3) that this protein is the first representative from a family of structurally similar Ca-regulated proteins. Citrate synthase (4) which includes a long pair of buried helices down the middle, may be the first example of a larger, 3-layer helical domain. In general it may be true that the miscellaneous structures are simply the ones for which our sample is still too small to have seen and recognized their relatives.

Parallel Alpha/Beta Proteins

The second major group of structures are the parallel alpha/beta proteins. The first major subdivision is made up of large but remarkably simple domains with a central, eight-stranded, parallel beta barrel and a concentric outer cylinder of 8 alpha helices. Each helix connects the top of one beta strand to the bottom of the next strand over, so that the entire structure winds continuously around the barrel in one direction, advancing by one beta strand at a time. This structure is called the singly-wound parallel beta barrel. Figure 7 shows end-on and side views of triose phosphate isomerase as an example of a singly-wound beta barrel.

The parallel alpha/beta structures in general, and especially the singly-wound barrels, are dominated by the need to satisfy an extremely strong empirical handedness requirement. The chain that connects two parallel strands in a beta sheet must get from one end of the sheet to the other end (as opposed to a hairpin connection between antiparallel strands, which stays at one end of the sheet); such a crossover loop is a handed structure (see Figure 8). It turns out that 99% of the crossover connections in the known protein structures are righthanded, regardless of the number of strands between them in the sheet or of the length and conformation of the connecting loop (5,6). The connections in a parallel beta barrel must all be on the outside of the barrel, since its interior is filled with hydrophobic side chains and has no room for another piece of backbone. Therefore the need to make righthanded crossover connections forces the chain to move in one continuous direction around the barrel. It would theoretically be possible to skip a strand in this progression and fill it in after going once around the barrel, but it turns out that that would make a knot in the chain as well as a less stable structure during folding. Thus a parallel beta barrel needs to have a singly-wound topology, and all of them are very similar to one another.

The second major division within the parallel alpha/beta

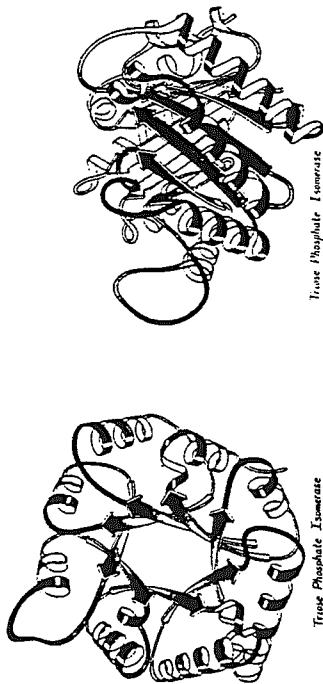


Figure 7. A singly-wound parallel beta barrel (parallel alpha/beta category): triose phosphate isomerase (27), shown in end view (left) and side view (right).

250

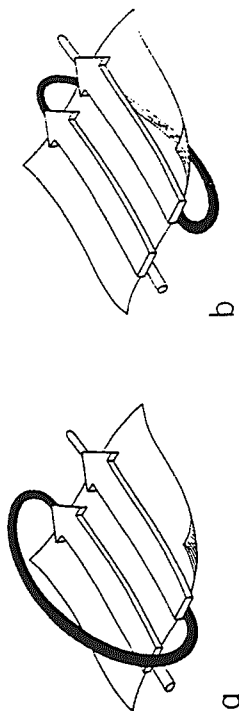


Figure 8. Parallel beta structures, emphasizing the handedness of crossover connections: a) Right-handed crossover connection; b) Left-handed crossover connection (extremely rare).

structures are the doubly-wound parallel beta sheets, or "nucleotide-binding domains" (7). These domains have a central layer of twisted beta sheet with 4 to 9 strands of all parallel or mostly parallel beta structure. The central sheet is flanked on each side by a layer of alpha helices. The beta sheet is buried on both sides and therefore presumably hydrophobic on both sides, which is characteristic of parallel beta structure in general. The flanking helix layers are formed by the crossover connections between beta strands; in order for them to cover both sides of the sheet using only righthanded connections, it is necessary for the sequential progression of strands along the sheet to change direction at least once. In practice the way this is accomplished is remarkably consistent among these proteins: the chain starts near the middle of the beta sheet and winds toward one edge using nearest-neighbor righthanded crossover connections and covering one side of the sheet with helices; it then skips back to the center and winds toward the opposite edge of the sheet, covering the other side with helices. Figure 9 shows a side view and an edge view of lactate dehydrogenase domain 1, the original and classic example of a doubly-wound parallel beta sheet. Figure 10 shows additional doubly-wound sheet domains with varying degrees of similarity to lactate dehydrogenase. This is the commonest type of protein structure, with over 30 examples now known.

Figure 11b shows in highly simplified form the direction change from which the doubly-wound beta sheets get their name, contrasted with the topology of the singly-wound barrel shown in Figure 11a. Figure 11c illustrates the reverse doubly-wound topology which winds in from the edge of the beta sheet toward the center, again with a single direction change to permit covering both sides of the sheet. A priori this seems like an equally reasonable arrangement, but in fact it does not occur in the known protein structures except as a strictly local arrangement within a large sheet dominated by the normal doubly-wound organization. This preference is not understood. It could be produced by a strict requirement for sequential folding from the N terminus plus restriction to only one major nucleation site, but such restrictions are not consistent with the organization found in other major categories of protein structures. Alternatively, it is possible that the doubly-wound sheet proteins have all diverged from a common ancestor. Many of them have a mono- or dinucleotide binding site at the C-terminal end of the beta strands at the center of the sheet where it changes direction, and many of them are enzymes of the glycolytic pathway. Either of these functional characteristics could help to rationalize evolutionary relatedness, but one must at the same time understand why other functional categories of proteins generally include a wide diversity of structural types.

An interesting case from the viewpoint of folding is gly-

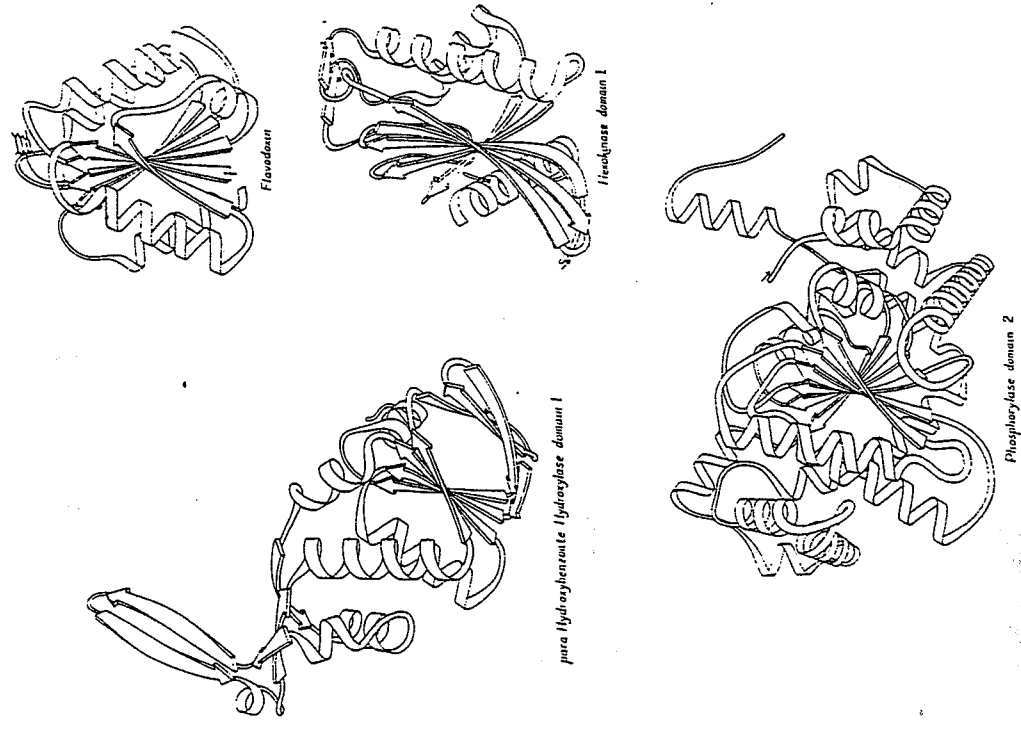


Figure 10. Doubly-wound parallel beta sheet structures (parallel alpha/beta category): p-hydroxybenzoate hydroxylase domain 1 (29), flavodoxin (30), hexokinase domain 1 (32), and glycogen phosphorylase domain 2 (31).

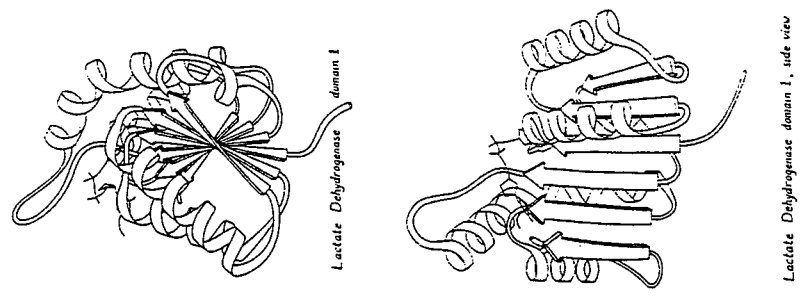


Figure 9. A doubly-wound parallel beta sheet (parallel alpha/beta category): lactate dehydrogenase domain 1 (28), shown in edge view (top) and in side view (bottom).

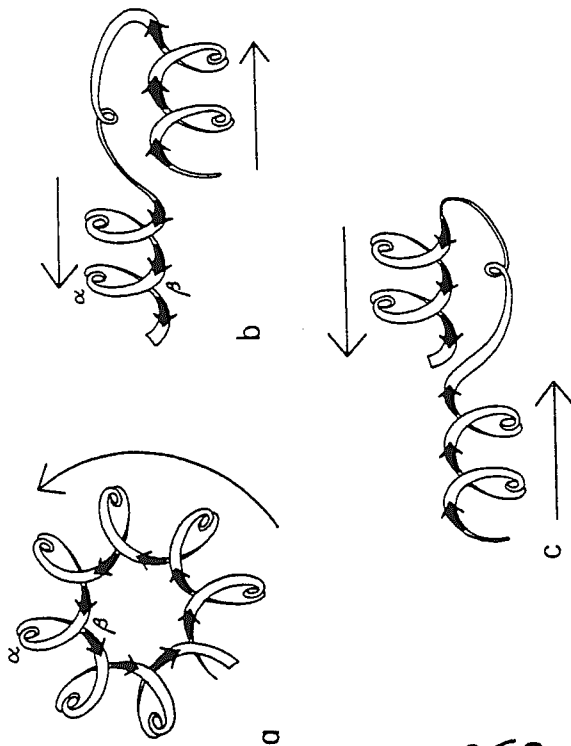


Figure 11. Simplified versions of a) A singly-wound beta barrel; b) A doubly-wound beta sheet; c) A reverse doubly-wound beta sheet.

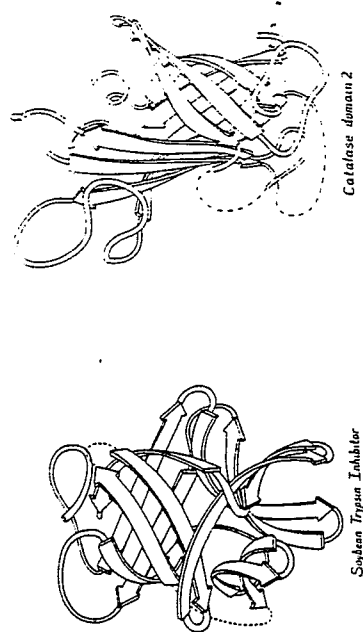


Figure 12. Up-and-down beta barrel structures (antiparallel beta category): soybean trypsin inhibitor (33) and bovine catalase domain 2 (34).

cogen phosphorylase. It has 2 extremely large domains, each with a central parallel beta sheet flanked on each side by 2 layers of helices rather than the usual single layer. Domain 2 of phosphorylase is shown in Figure 10. Its middle 3 layers form a doubly-wound alpha/-beta structure, with the 6 beta strands connected in classic lactate dehydrogenase topology. Those central 3 layers are formed from the middle of the amino acid sequence of the domain, while the outer layers of helices come from the 2 ends of the sequence. Presumably the central part folds up first and the outer layers add on afterward.

Antiparallel Beta Proteins

The third major category is the antiparallel beta structures. As for antiparallel alpha proteins, the simplest type of antiparallel beta domain forms an approximate cylinder, or barrel, with up-and-down nearest-neighbor connectivity between the strands all around. Figure 12 shows soybean trypsin inhibitor and catalase domain 2 as examples. In the beta barrels, however, the simple up-and-down topology is actually rather rare while the Greek key beta barrels are the second commonest domain structure, with about 20 examples known so far. Figure 13 shows increasingly simplified representations of the prealbumin structure to illustrate how the Greek key topology works in 3 dimensions and how it can be rolled out flat to resemble the "Greek key" border motif common on Greek vases. Figure 14 shows several more examples of protein domains that are Greek key beta barrels. The simple feature shared by all the Greek key structures is having at least one (or as many as 4) strand connections that cross an end of the barrel.

The Greek key barrel is a handed structure, which is manifested either in the direction of swirl (counterclockwise as in Figure 13 d and e, rather than clockwise) when viewed from the outside, or else in the righthandedness of the crossover connections between opposite sides of the barrel. All but one of the known examples are counterclockwise or righthanded, the sole exception being Staphylococcal nuclease (8), which is the only beta barrel with as few as 5 strands and is very highly twisted. This handedness is quite a large-scale effect, involving a minimum of about 50 or 60 residues in the sequence. It is very likely that both the common occurrence and the preferred handedness of Greek key beta barrels is a result of the way they fold up (9,10). Figure 15 illustrates how a long 2-stranded ribbon of antiparallel beta structure with the usual righthanded twist (along the strand direction) can automatically produce a Greek key topology with the correct handedness when it loops over onto itself into a compact form. [The connection between strand twist and Greek key handedness is analogous to the relationship between helix pitch and superhelical twist in circular DNA.] Figure 14 includes the larger domain

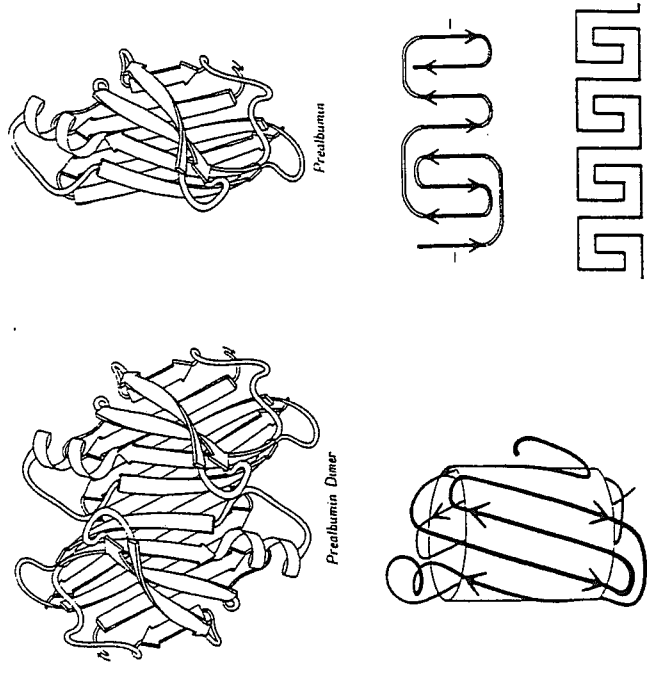


Figure 13. Increasingly simplified representations of a Greek key beta barrel structure (prealbumin (35)). Top: backbone drawings of the dimer and the subunit. Bottom: the backbone shown on the surface of a cylinder; a 2-dimensional topology diagram; and the Greek key motif that topology resembles.

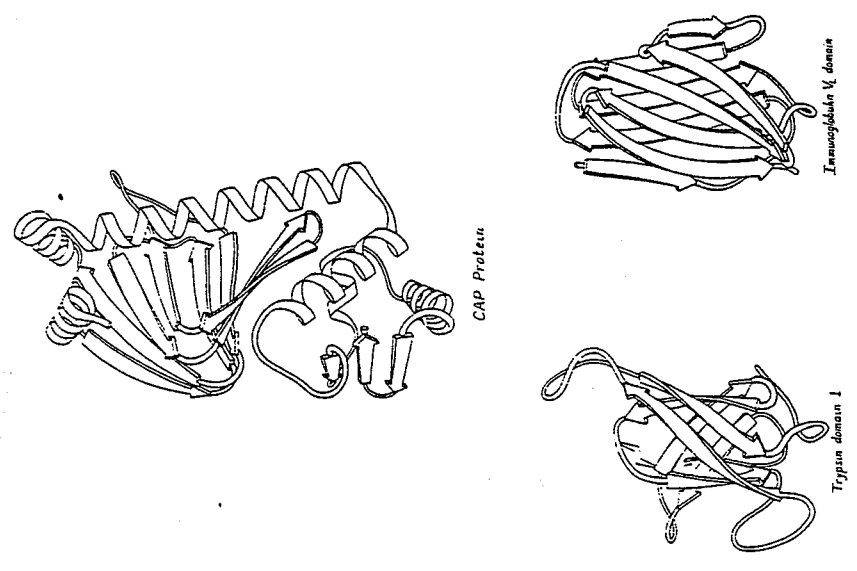


Figure 14. Greek key beta barrel structures (antiparallel beta category): catabolite-activating (CAP) protein (38), trypsin domain 1 (36), and immunoglobulin VL domain (37).

253

of CAP (catabolite-activating protein), which is an 8-stranded beta barrel with "perfect" Greek key topology: starting from a tight turn at the middle of the sequence, the beta strands can be seen to coil around the barrel in pairs all the way to the beginning and ending helices at the top of the drawing. If this structure (or the large domain from influenza virus haemagglutinin (11)) had been the first Greek key barrel found, then the proposal of folding as a double ribbon would probably have been made immediately.

If this double-ribbon winding process actually happens during protein folding, then it has considerable consequences in terms of concerted early folding of quite long segments of the backbone. It is also the clearest available evidence for fundamental differences in early folding processes between the different major classes of protein structures: in the antiparallel beta barrels the statistical distribution of observed topologies seems to argue overwhelmingly for some such handed, concerted early folding mechanism, whereas the formally analogous helix bundle structures prefer up-and-down to Greek key topologies and show no clear handedness preferences. However, even in this case the distinctions are not conclusive, both because the helix topologies are not inconsistent with the folding over of long helix pairs although they certainly do not require it, and also because it is perhaps possible that the resemblances of the Greek key beta barrels could be the result of descent from a common ancestor.

The next subgroup within the antiparallel beta category is the "open-face sandwich", with a single antiparallel beta sheet as the slice of bread and a layer of helices and loops on only one side of the sheet as the sandwich spread. These domains are quite common, but vary considerably in size and in topology (see Figure 16) from only 3 beta strands in bacteriophage protein all the way to 15 strands in bacteriophage T4 protein, the most extravagant example of this class. In addition to the usual hydrophobic side chains between the beta sheet and the helix layer, bacteriophage T4 protein has a "mayonnaise" consisting of 7 bacteriophage T4 molecules bound asymmetrically but specifically in its large hydrophobic interior. Obviously this polypeptide chain could not assume its final folded form without the bacteriophage T4 molecules.

Then there is an interesting set of miscellaneous structures in the antiparallel beta category (see Figure 17). They include sheets that have folded around partway toward making a beta barrel such as pancreatic ribonuclease, structures with several 2-stranded beta ribbons such as lactate dehydrogenase domain 2, and unusual structures such as the "beta annulus" formed by association of the N-terminal tails of 3 subunits of tomato bushy stunt virus protein around a 3-fold axis in the spherical virus capsule (12). Although not a domain in the usual sense, the beta annulus raises interesting folding ques-

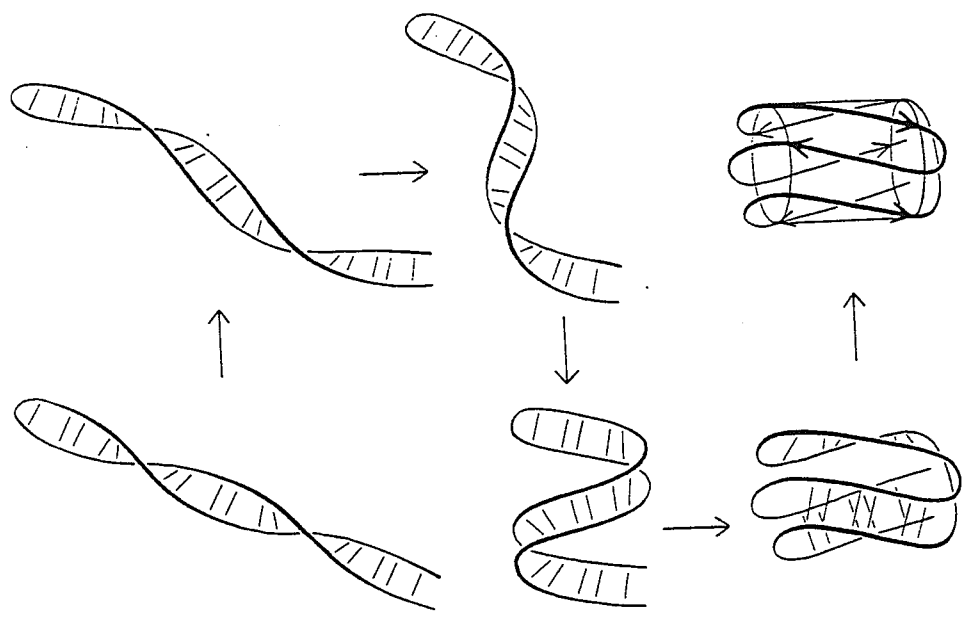


Figure 15. Hypothetical folding of a long, 2-stranded beta ribbon to form a Greek key beta barrel.

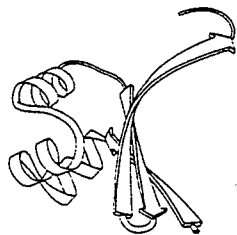
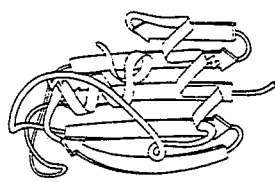
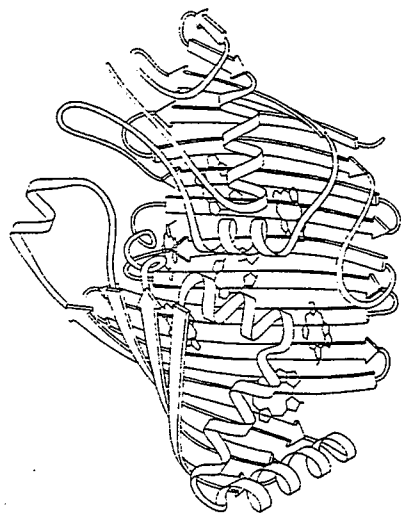


Figure 16. Open-face sandwich structures (antiparallel beta category): bacteriochlorophyll protein (41), cro repressor (39), and subtilisin inhibitor (40).

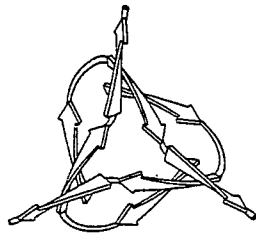
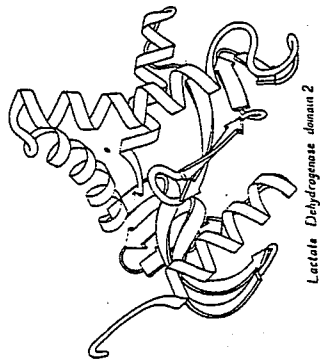
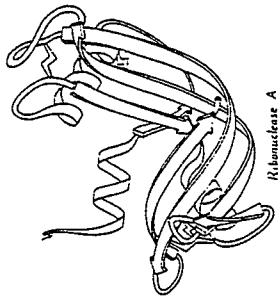


Figure 17. Miscellaneous antiparallel beta structures: pancreatic ribonuclease A (42), lactate dehydrogenase domain 2 (28), and tomato bushy stunt virus "domain" 1 (12).

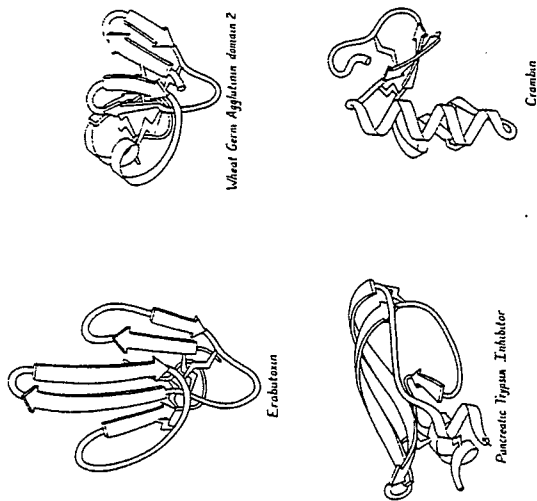


Figure 18. Small disulfide-rich structures (small irregular category): erabutoxin (43), wheat germ agglutinin (44), pancreatic trypsin inhibitor (17), and crambin (45).

256

tions because it provides one of the major contacts that holds the viral subunits together and yet it consists of entirely non-local interactions between 3 separate polypeptide chains so that it seems unlikely to form before at least some primitive association of the subunits into icosahedral geometry.

Small Irregular Proteins

The fourth major category of protein structure is different logically from the other three, which are dominated by patterns of secondary structure. The small irregular proteins have relatively little conventional secondary structure, and what they do have is generally short and irregular. Instead, their organization is apparently dominated either by a high content of disulfides (more than 10% half-cystine) or else by bound metals or prosthetic groups. Figure 18 shows several small, disulfide-rich proteins and Figure 19 shows several small, metal-rich proteins. Almost all of the small irregular domains occur by themselves as single-domain, single-subunit proteins, and the few exceptions (insulin (13), wheat germ agglutinin, and the multi-domain protease inhibitors (14)) are self-associating rather than occurring in combination with one of the other major categories. Disulfide-rich proteins are typically extracellular, and most function as toxins, lectins, enzyme inhibitors, or hormones. The only two which share a specific topology are the snake neurotoxins and the domains of wheat germ agglutinin (15).

Most of the small metal-rich proteins form approximately cylindrical structures with either a simple up-and-down (e.g., rubredoxin, cytochrome c) or a Greek key (e.g., ferredoxin) topology, but where the elements forming the cylinder are a mixture of helices and extended strands (which may or may not form any beta structure). The ultimate example of a small irregular protein is cytochrome c₃, which has 4 hemes enclosed by just over 100 amino acids, and which has almost no secondary structure (see Figure 19).

In general, most of the small irregular proteins give the impression of being small, distorted versions of one of the other structural categories. As shown in Figure 20, they may be distorted helix bundles, distorted beta barrels, or distorted open-face sandwiches. However, there are no small, distorted versions of alpha/beta structures, perhaps because they seem to be dependent on large, buried, and quite regular beta structure for their stability. Although almost all alpha/beta domains contain either nucleotide prosthetic groups or metals, there are no heme groups in spite of all the alpha helices, and no disulfides except the single active-site disulfides in thioredoxin, glutathione peroxidase and glutathione reductase. As a general rule, disulfides appear to be incompatible with either metals or prosthetic groups, a dichotomy which allows

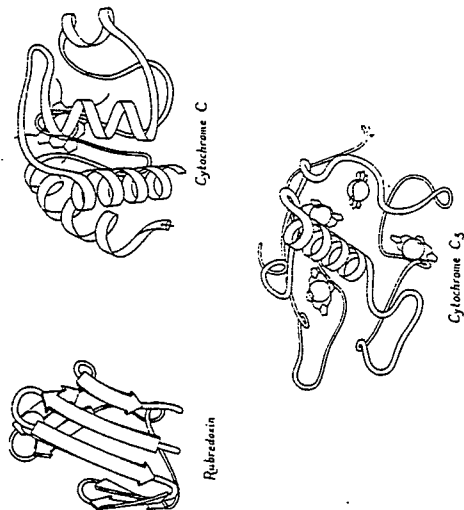


Figure 19. Small metal-rich structures (small irregular category): rubredoxin (46), cytochrome c (47), and cytochrome C₃ (48).

the separation of the small irregular proteins into disulfide-rich and metal-rich. This restriction may possibly have to do with the differences in how disulfides and prosthetic groups are incorporated into proteins during or after folding.

Since so many of the small irregular proteins resemble distorted versions of more regular structures, it is tempting to think that they first fold with a higher content of secondary structure and then rearrange considerably at the stage when the disulfides are formed or the metals incorporated. In fact pancreatic trypsin inhibitor, the protein whose folding intermediates have been most thoroughly characterized (16), is a small disulfide-rich protein and is found to undergo remarkably thorough reorganization during folding. Presumably all proteins rearrange to some extent during the final stages of folding in order to settle into an optimal balance of all energetic factors. However, it is hard to see how so many proteins could have such simple, regular structural organization if they all went through the kind of contortions displayed by trypsin inhibitor. This may then represent another of the differences in importance of various folding processes among the major categories of protein structure.

Discussion

There are three major implications for protein folding which one can take away from this sort of survey of the known native protein structures.

The first suggested conclusion is that most proteins do in fact fit the concept of division into domains that could potentially make plausible folding units, since they can be classified as similar to other intact, single-domain proteins which are competent folding units by definition. However, there are occasional exceptions. In some cases (e.g., para-hydroxybenzoate hydroxylase, pyruvate kinase, phosphofructokinase, etc.) the obvious compact domains in 3 dimensions are formed from two sequentially distant chain segments. Short N- or C-terminal segments of chain fairly often are tightly associated with a domain distant rather than adjacent in sequence (e.g., papain) or even with another subunit (e.g., lactate dehydrogenase). And sometimes a "core" domain (such as the central 3 layers of phosphorylase domain 2) is flanked by a small compact region of structure (in phosphorylase, the outer helix layers) too small, presumably, to fold stably by itself. In all of the cases found so far, one can identify at least one normal, "core" domain which satisfies all criteria of the definition (compactness, chain contiguity, presumed stability as a separate structure, etc.); therefore, one assumes that such core domains fold first, and that the small or non-contiguous regions can fold stably only afterward. The relative importance for folding of chain contiguity versus 3-dimensional

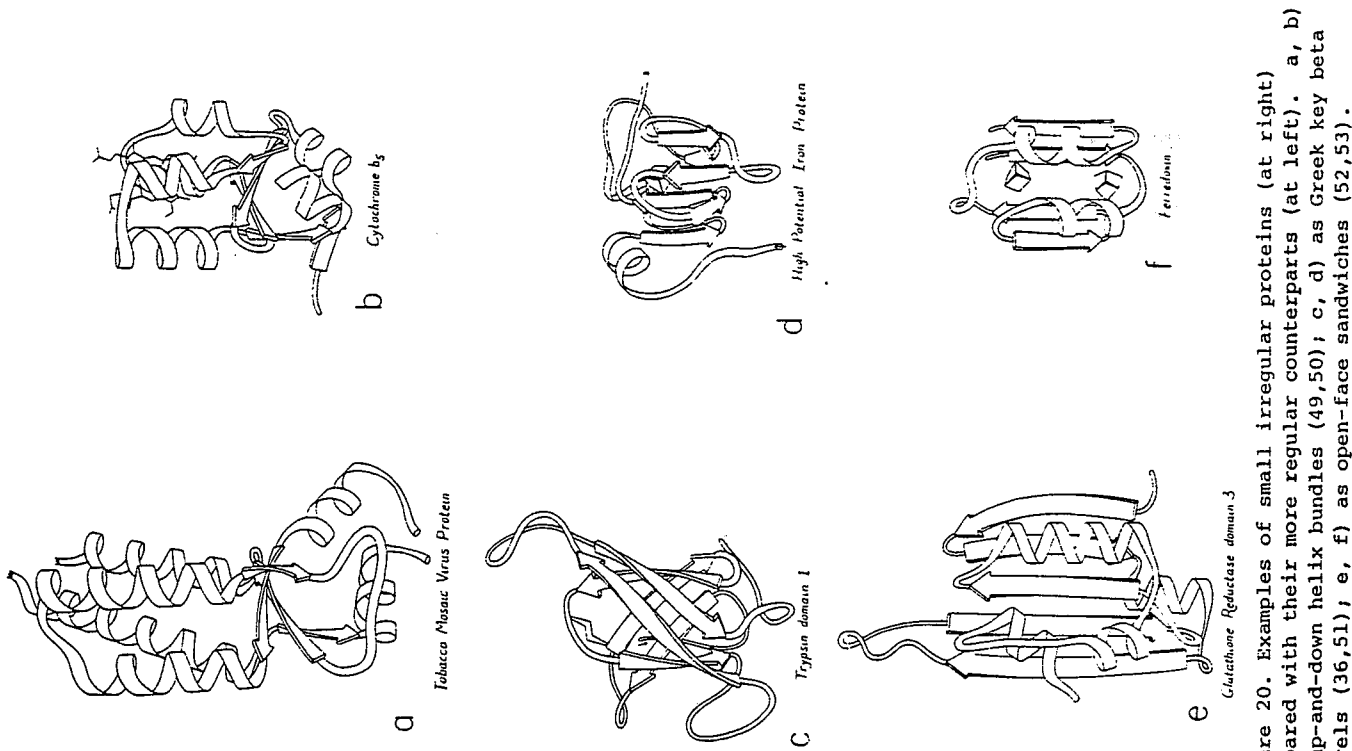


Figure 20. Examples of small irregular proteins (at right) compared with their more regular counterparts (at left). a, b as up-and-down helix bundles (49,50); c, d as Greek key beta barrels (36,51); e, f as open-face sandwiches (52,53).

25.

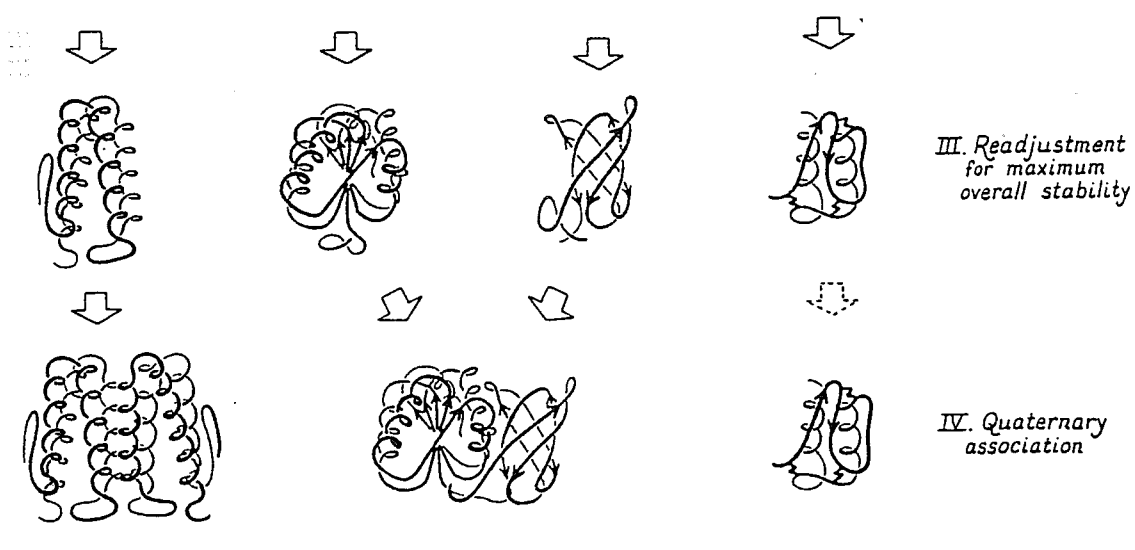
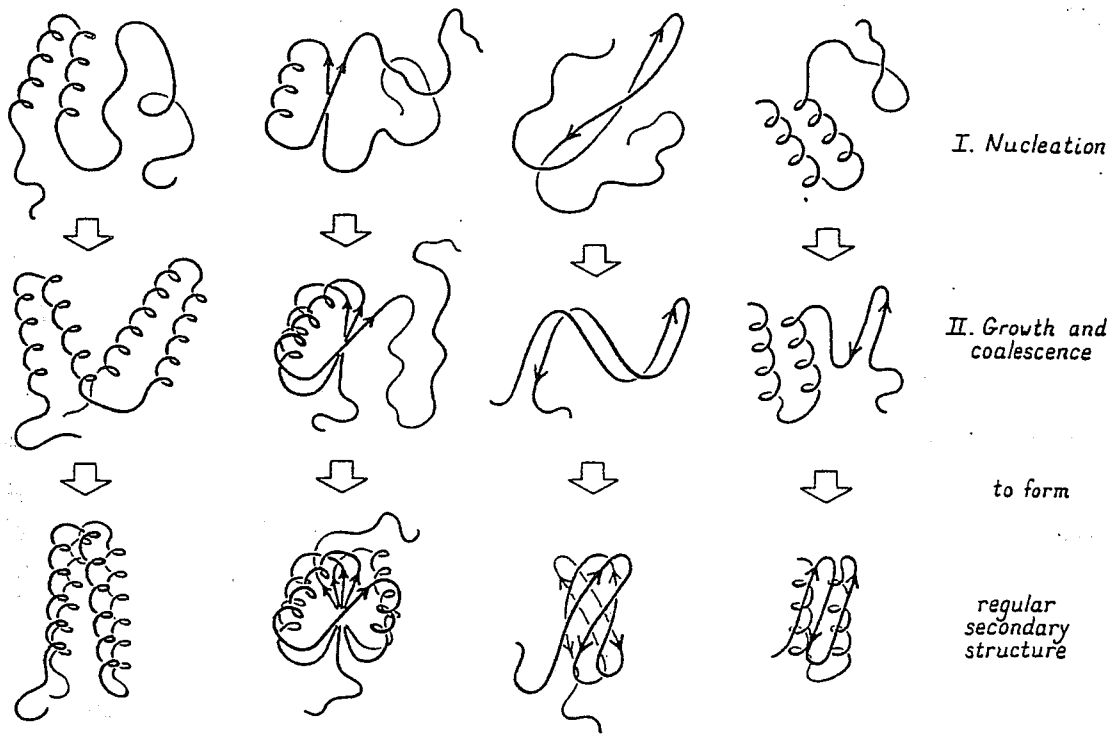


Figure 21. Suggested steps of folding, for proteins in each of the different categories of tertiary structure.

4.

compactness is an unanswered empirical question, however, and it is also possible that a region like the first domain of p-hydroxybenzoate hydroxylase (see Figure 10) can fold early but with a large loop which remains disordered until the domains associate with one another.

The second implication of this survey is that there may be quite long segments of the polypeptide chain that fold as concerted units rather than by strictly local conformational searches or local nucleation-plus-growth. Possible candidates include the crossover connections of the parallel alpha/beta structures, and the double beta ribbons that may be involved in forming the Greek key beta barrels (see Figure 15).

The third major implication is that one can expect important differences in the order and relative importance of various folding processes between the major categories of protein structure and even in some cases for unusual individual structures within those categories (e.g., phosphorylase, bacteriochlorophyll protein, tomato bushy stunt virus domain 1, etc.). Figure 21 summarizes the principal steps that can be surmised to occur during folding, with indication of possible differences to be expected from one category to another. It follows, then, that one important job for experimental folding studies is to examine and compare typical examples of the different major categories, when such choices are possible.

References

1. J.S. Richardson, *Adv. Prot. Chem.* **34**, 167-339 (1981).
2. D.B. Wetlaufer, *Proc. Nat. Acad. Sci. USA* **70**, 697-701 (1973).
3. R.H. Kretsinger, *Ann. Rev. Biochem.* **45**, 239-266 (1976).
4. S. Remington, G. Weigand, R. Huber, *J. Mol. Biol.* **158**, 111-152 (1982).
5. J.S. Richardson, *Proc. Nat. Acad. Sci. USA* **73**, 2619-2623 (1976).
6. M.J.E. Sternberg, J.M. Thornton, *J. Mol. Biol.* **110**, 269-283 (1977).
7. M.G. Rossmann, D., Moras, K.W. Olsen, *Nature (London)* **250**, 194-199 (1974).
8. A. Arnone, C.J. Bier, F.A. Cotton, V.W. Day, E.E. Hazen, Jr., D.C. Richardson, J.S. Richardson, A. Yonath, *J. Biol. Chem.* **246**, 2303-2316 (1971).
9. J.S. Richardson, *Nature (London)* **268**, 495-500 (1977).
10. O.B. Ptitsyn, A.V. Finkelstein, in *Protein Folding* (R. Jaenicke, ed.), Elsevier/North Holland, pp.101-115 (1980).
11. I.A. Wilson, J.J. Skehel, D.C. Wiley, *Nature (London)* **289**, 366-373 (1981).
12. S.C. Harrison, A.J. Olson, C.E. Schutt, F.K. Winkler, G. Bricogne, *Nature (London)* **276**, 368-373 (1978).
13. T.L. Blundell, G.G. Dodson, D.C. Hodgkin, D.A. Mercola,

14. W. Bode, P. Schwager, R. Huber, *J. Mol. Biol.* **118**, 99-112 (1978).
15. J. Drenth, B. Low, J.S. Richardson, C.S. Wright, *J. Biol. Chem.* **255**, 2652-2655 (1980).
16. T.E. Creighton, *J. Mol. Biol.* **113**, 275-341 (1977).
17. J. Dalsenhofer, W. Steigemann, *Acta Crystallogr.* **B31**, 238-250 (1975).
18. F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535-542 (1977).
19. L. Sawyer, D.M. Shotton, J.W. Campbell, P.L. Wendell, H. Muirhead, H.C. Watson, R. Diamond, R.C. Ladner, *J. Mol. Biol.* **118**, 137-208 (1978).
20. T. Blundell, P. Ländley, L. Miller, D. Moss, C. Slingsby, I. Flocke, B. Turnell, G. Wistow, *Nature (London)* **289**, 771-777 (1981).
21. J. Drenth, J.N. Jansonius, R. Koekoek, B.G. Woithers, *Adv. Prot. Chem.* **25**, 79-115 (1971).
22. W.A. Hendrickson, K.B. Ward, *J. Biol. Chem.* **252**, 3012-3018 (1977).
23. F.S. Mathews, P.H. Bethge, E.W. Czerwinski, *J. Biol. Chem.* **254**, 1699-1706 (1979).
24. R.C. Ladner, E.J. Heidner, M.F. Perutz, *J. Mol. Biol.* **114**, 385-414 (1977).
25. P.M. Colman, J.N. Jansonius, B.W. Matthews, *J. Mol. Biol.* **70**, 701-724 (1972).
26. R.H. Kretsinger, C.E. Nockolds, *J. Biol. Chem.* **248**, 3313-3326 (1973).
27. D.W. Banner, A.C. Bloomer, G.A. Petsako, D.C. Phillips, C.I. Pogson, I.A. Wilson, *Nature (London)* **255**, 609-614 (1975).
28. M.J. Adams, G.C. Ford, R. Koekoek, P.J. Lentz, Jr., A. McPherson, M.G. Rossmann, I.E. Smiley, R.W. Schevitz, A.J. Wonacott, *Nature (London)* **227**, 1098-1103 (1970).
29. R.K. Weirenga, R.J. de Jong, K.H. Kalk, W.J.G. Hol, J. Drenth, *J. Mol. Biol.* **131**, 55-73 (1979).
30. R.M. Burnett, G.D. Darling, D.S. Kendall, M.E. LeQueune, S.G. Mayhew, W.W. Smith, M. Ludwig, *J. Biol. Chem.* **249**, 4383-4392 (1974).
31. S. Sprang, R.J. Fletterick, *J. Mol. Biol.* **131**, 523-551 (1979).
32. T.A. Steitz, R.J. Fletterick, W.A. Anderson, C.M. Anderson, *J. Mol. Biol.* **104**, 197-222 (1976).
33. R.M. Sweet, H.T. Wright, J. Janin, C. Chothia, D.M. Blow, *Biochemistry* **13**, 4212-4228 (1974).
34. M.R.N. Murthy, T.J. Reid, III, A. Sicignano, N. Tanaka, M.G. Rossmann, *J. Mol. Biol.* **152**, 465-499 (1981).
35. C.C.F. Blake, M.J. Geisow, S.J. Oatley, B. Rerat, C. Rerat, *J. Mol. Biol.* **121**, 339-356 (1978).

36. R.M. Stroud, L.M. Kay, R.E. Dickerson, J. Mol. Biol. 83, 185-208 (1974).
37. O. Epp, P. Colman, H. Fehlhammer, W. Bode, M. Schiffer, R. Huber, W. Palm, Eur. J. Biochem. 45, 513-524 (1974).
38. D.B. McKay, I.T. Weber, T.A. Steitz, J. Biol. Chem. 257, 9518-9524 (1982).
39. W.A. Anderson, D.H. Ohlendorf, Y. Takeda, B.W. Matthews, Nature (London) 290, 754-758 (1981).
40. Y. Mitsui, Y. Satow, Y. Watanabe, Y. Iitaka, (1979) J. Mol. Biol. 131, 697-724 (1979).
41. B.W. Matthews, R.E. Fenna, M.C. Bolognesi, M.F. Schmid, J. M. Olson, J. Mol. Biol. 131, 259-285 (1979).
42. H.W. Wyckoff, D. Tsernoglou, J.W. Hanson, J.R. Knox, B.K. Lee, F.M. Richards, J. Biol. Chem. 245, 305-328 (1970).
43. B.W. Low, H.S. Preston, A. Sato, L.S. Rosen, J.E. Searl, A.D. Rudko, J.S. Richardson, Proc. Nat. Acad. Sci. USA 73, 2991-2994 (1976).
44. C.S. Wright, J. Mol. Biol. 111, 439-457 (1977).
45. W.A. Hendrickson, M.M. Teeter, Nature 290, 107-113 (1981).
46. K.D. Watenpaugh, L.C. Sieker, L.H. Jensen, J. Mol. Biol. 131, 509-522 (1979).
47. R. Swanson, B.L. Trus, N. Mandel, G. Mandel, O.B. Kallal, R.E. Dickerson, J. Biol. Chem. 252, 759-775 (1977).
48. M. Pierrot, R. Haser, M. Frey, F. Payan, J.-P. Astier, J. Biol. Chem. 257, 14341-14348 (1982).
49. A.C. Bloomer, J.N. Champness, G. Bricogne, R. Staden, A. Klug, Nature (London) 276, 362-368 (1978).
50. F.S. Mathews, M. Levine, P. Argos, J. Mol. Biol. 64, 449-464 (1972).
51. C.W. Carter Jr., J. Kraut, S.T. Freer, N.H. Xuong, R.A. Alden, R.G. Bartsch, J. Biol. Chem. 249, 4212-4225 (1974).
52. R. Thieme, E.F. Pai, R.H. Schirmer, G.E. Schulz, 152, 763-782 (1981).
53. E.T. Adman, L.C. Sieker, L.H. Jensen, J. Biol. Chem. 248, 3987-3996 (1973).

