The Protein Folding Problem

In theory, all one needs to know in order to fold a protein into its

fully folded proteins and by analyzing the properties of individual amino acids and small peptides (linear chains of amino acids). Fortunately, the architecture of hundreds of native proteins has been determined by such imaging techniques as X-ray crystallography and, more recently, nuclear magnetic resonance (NMR). Both techniques have advanced dramatically in the past decade, as has theoretical work attempting to predict folding mathematically by computer.

Isolated amino acids consist of a central carbon atom—called the alpha car-

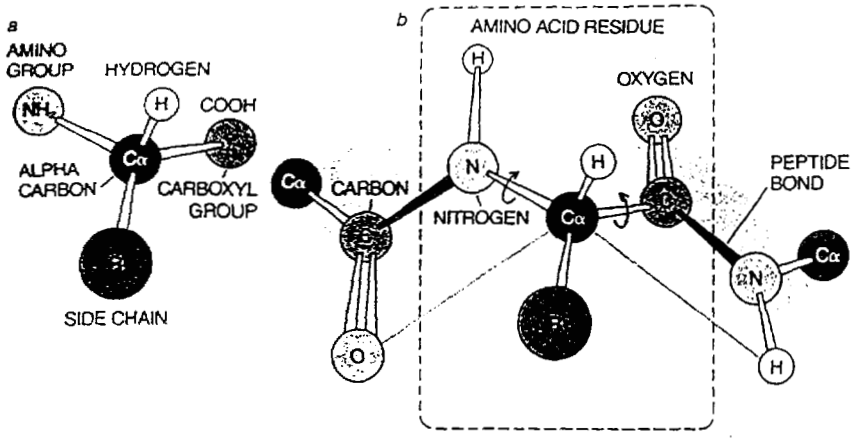
when their oppositely charged regions are close; they are repelled when like charged regions are close.

Nonpolar amino acids can also attract or repel one another, albeit more weakly, because of what are called van der Waals forces. Electrons and protons vibrate constantly, and the vibrations result in attractions between substances that are near one another. The attraction turns into repulsion when the substances are about to touch.

In aqueous solution, polar amino acids tend to be hydrophilic; they attract water molecules, which are quite

peptide bond is severely limited. Indeed, the atoms lying between alpha carbons are held in a single plane, so that they essentially form a stiff plate. Folding of the peptide backbone is therefore accomplished mainly by rotation of the plates around other bonds—namely, those connecting the plates to the alpha carbons.

Examination of the peculiarities of denatured, or unfolded, proteins has added still other hints to how folding is accomplished. Unfolded or newly formed proteins are



AMINO ACIDS (*a*) are linked together in a protein (*b*) by a strong bond that forms between the carboxyl carbon of one amino acid and the amino nitrogen of the next. Because the resulting linkage, which is known as a peptide bond, holds the joined atoms rigidly in a plane, the bond limits the conformational options of the protein. Folding is accomplished mainly by rotation about the axes of the bonds

persecondary structures, and the final assembly of all secondary elements is the tertiary structure. Several tertiary classes have been identified, such as the all alpha-helix class, the all beta-strand class and particular arrangements of combinations of helices and beta strands.

The presence of different secondary elements raises the possibility that certain amino acids favor development of specific secondary arrangements. For example, some amino acid residues are found more often in helices than elsewhere, whereas others tend to be found in beta sheets. On the other hand, none of these or other similar statistical correlations are strong.

Several other discoveries show that, as might be expected from the hydrophobic and hydrophilic properties of the amino acids, the tendency of

Researchers agree on details of the structure of folded proteins, but they diverge on many other points. There is, for instance, little agreement on the nature and number of folding pathways.

At one extreme is the doubtful suggestion that a newly made protein tries out all possible conformations until it finds the unique, stable structure of the native protein. This proposal assumes

the molecule toward greater structural organization by associating with other segments or helping to bring distant segments into contact, or both.

Inherent in this kind of model is the assumption that the hydrophobic effect is large but can be spent incrementally. Some fraction of its energy is expended to influence the formation of secondary elements, and the rest promotes

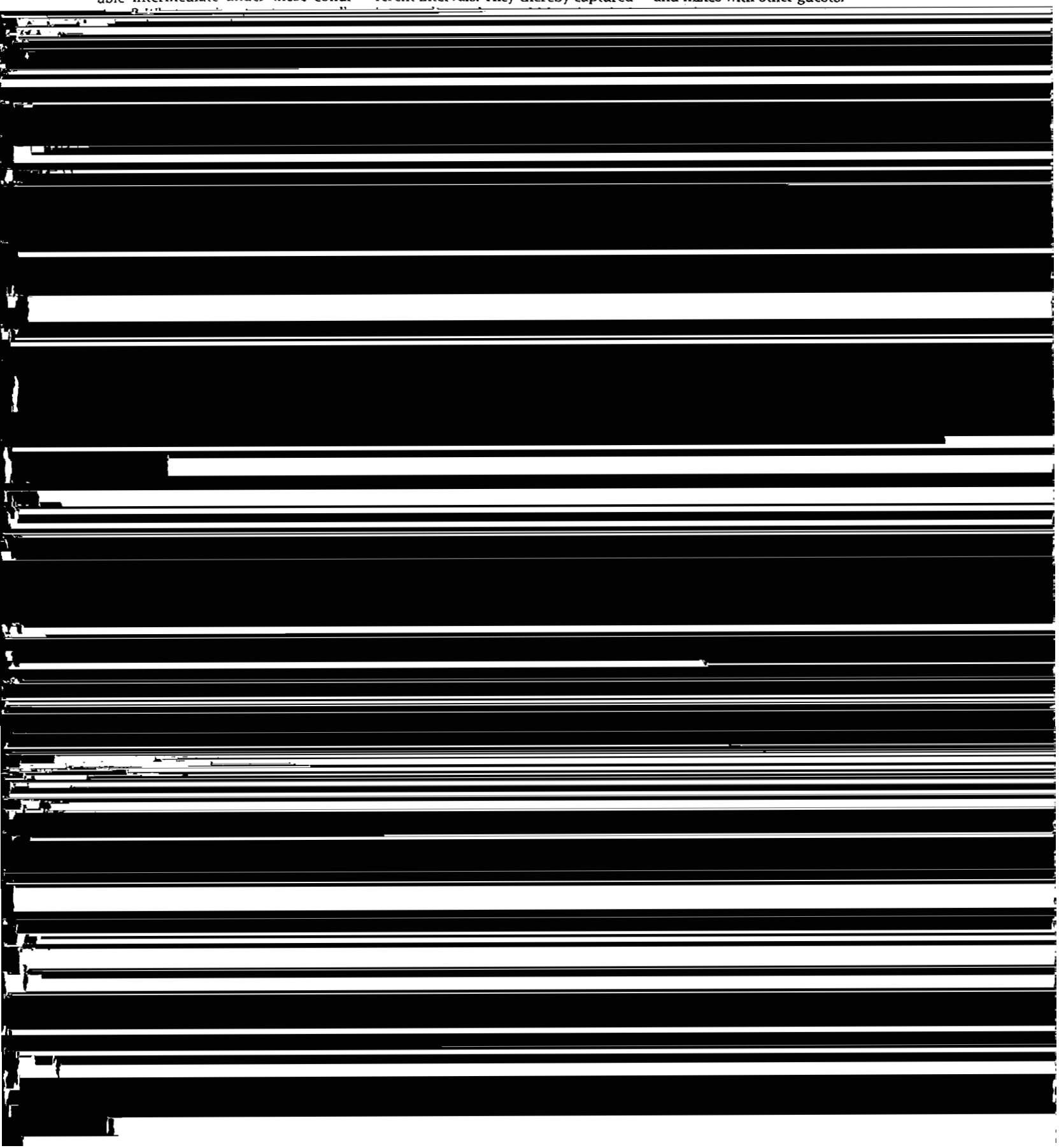
ly cooperative process. Interactions that promote folding by one part of the protein also promote folding elsewhere in the molecule; hence, intermediate shapes do not persist for long. Nevertheless, clever techniques have captured or identified some characteristics of a number of intermediates.

There is now firm evidence, for example, that certain proteins form an in-

ume than the native molecule, it must contain a considerable amount of water, and many of the side chains in the globule do seem to be in contact with water. Yet the force of the hydrophobic effect should be squeezing this water out. How can one have a stable, observable intermediate under these condi-

internal disulfide bonds as it folds. A disulfide bond is a sulfur-sulfur (S-S) linkage between the side chains of two cysteine amino acid residues. Creighton and his co-workers unfolded the native product and then started the folding reaction, interrupting it at different intervals. They thereby captured

that do not exist in the final molecule appear and then disappear. In other words, parts of the molecule apparently act something like a party host who brings two well-matched strangers together and then, when the two are engrossed in conversation, leaves them and mixes with other guests.



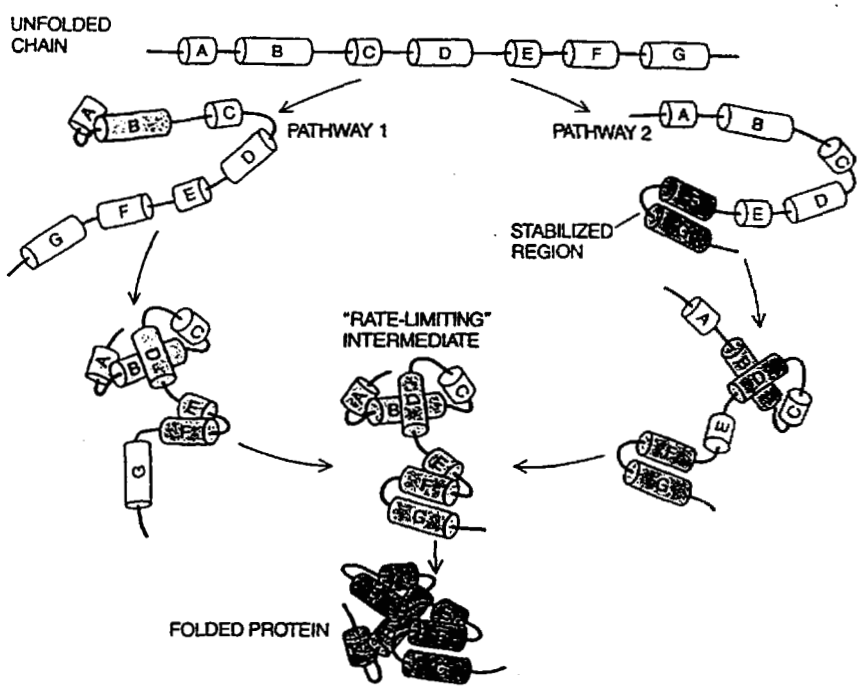
bond early in the folding process and that the bond persists. They wondered whether the supersecondary structure in the region around the bond also formed early and persisted.

To answer the question, they chemically synthesized two separate fragments of the protein, each including one of the two cysteines that participate in the stable disulfide bond. The small peptides had no discernible structure of their own, but when they joined in solution, they adopted a conformation closely resembling that seen in the native chain.

This finding confirms that nativelike structures can indeed form early, and it suggests that certain parts of the molecule may be more important than other parts in initiating folding. The result also indicates that interactions between apparently unstructured segments of a protein may facilitate the development of secondary structure.

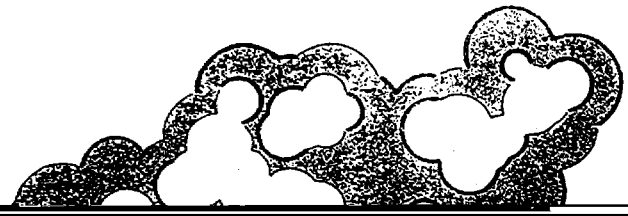
Intermediates are being studied by another ingenious method that capitalizes on the many internal hydrogen bonds found in all native proteins. First normal hydrogen atoms bound to the nitrogen involved in peptide bonds are exchanged with a related atom—the hydrogen isotope deuterium (D)—by placing the chains in heavy water, D_2O. Then folding is initiated.

As folding proceeds, what would have been hydrogen bonds become "deuterium" bonds (N-D-O) instead. At some chosen time, normal water (H_2O) is



PLAUSIBLE MODEL of how proteins fold allows for several energetically favorable pathways, although only two possibilities are shown. First the chain forms regions of unstable structure (*uncolored cylinders*). By associating, certain regions become stabilized (*color*). These stabilized microdomains then facilitate the association of other regions and thus lead the molecule toward increasing structural organization. Eventually, all pathways lead to one or more "rate-limiting" intermediates, which all give rise to the same final conformation for the protein.

Experimentation is not by any means limited to studies of intermediates. A number of scientists are approaching this way, they are testing various hypotheses, such as the proposal that certain sequences of hydrophobic and hy-



WATER-

factors as the influence on energy of the length, stretching and twisting of bonds and the strength of electrostatic interactions, hydrogen bonds and van der Waals forces. The approach has been valuable for confirming or improving models of structures that were

useful ways of thinking about the folding problem have emerged.

Any resolution of that puzzle will have to include a way of defining the force exerted on a protein molecule by water. In principle, estimates of the hydrophobic effect are, or can be, embedded in the potential-energy function, but exactly how best to accomplish that step is far from clear.

One method for analyzing the effect of water has emerged from work done by Byungkook Lee at Yale University in 1971. Lee developed an algorithm to calculate the solvent-accessible area of a protein of known structure—that part of the complex surface in direct contact with water. On the basis of preliminary findings, he and I suggested the algorithm would be useful in studying protein folding.

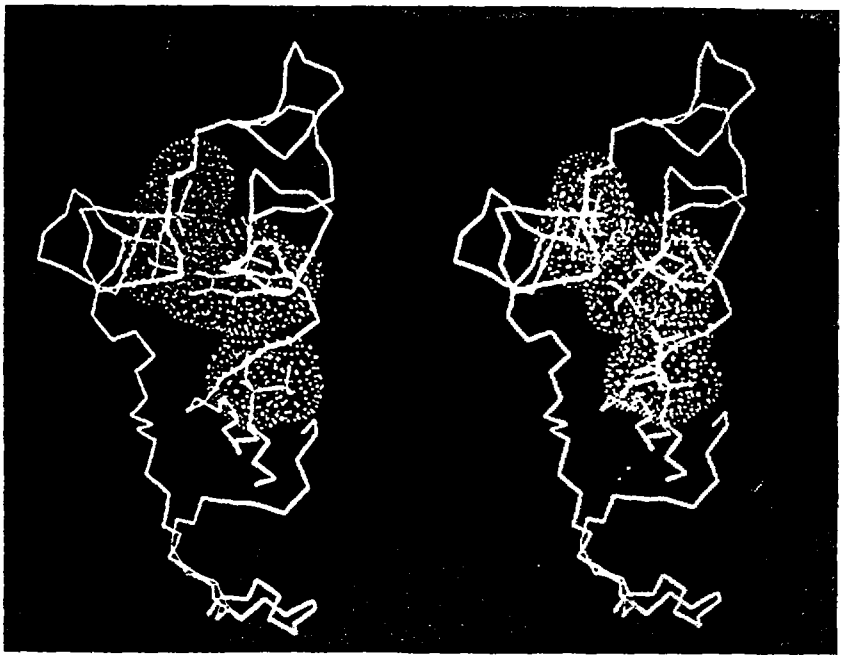
We divide the accessible area of an extended protein chain (or any selected molecule) according to the nature of the atoms that contribute to the area. Are they nonpolar and therefore hydrophobic (mainly carbon and sulfur atoms), or are they polar and therefore hydrophilic (mainly nitrogen and oxygen atoms)?

The surface tension of water in contact with such atoms is known. This tension is, as Cyrus Chothia of the Medical Research Council has pointed out, a direct measure of the force exerted on the molecule by the solvent. Surface tension is high when nonpolar molecules and water are in contact, just as it is when oil is mixed with water—that is, a strong force tends to reduce the area of contact between the water and the oil, and to squeeze a protein chain into a ball. Tension is low when polar atoms and water are in contact, and the hydrophobic effect is not seen.

Summation of the nonpolar accessible areas of an unfolded chain yields a measure of the potential hydrophobic effect. In general, as might be expected from structural analyses, the net force acting on most protein chains is large and positive, tending to reduce contact with the solvent and thus to compact the chain.

Various investigators are also examining the extent to which packing considerations direct folding. In one approach, lists have been made of the amino acid sequences of molecules that adopt essentially the same three-dimensional conformation. On the basis of the steric properties of the amino acids in the molecules—such as shape and volume—Jay W. Ponder of Yale has generated other lists of amino acid sequences that theoretically should adopt the same conformations.

Just how well those sequences actually fit their assigned classes is still being determined experimentally, but many do seem to fit. This finding, together with the profound influence of water, makes it conceivable that the hydrophobic effect and steric considerations by themselves determine how a protein folds.



CONFORMATION of a fold in the interior of the protein crambin (*left*), depicted mainly as a chain of alpha carbons (*orange*), derives from the tight packing of five nonpolar amino acids (*blue spheres*). That conformation is maintained in a computer-generated "mutant" (*right*) even when four of the five amino acids are replaced with others. Indeed, many combinations of amino acids can be accommodated if the substitutes resemble the originals in shape and volume. Knowledge of how amino acids pack may go a long way toward predicting the shape of a protein.

If that is the case, what is the role of long- and short-range electrostatic interactions in protein folding? Undoubtedly, the contribution of such interactions varies from protein to protein. For many proteins, large changes in the formal charges can be made without significantly affecting the final overall structure. Hence, it may be that electrostatic interactions are often more important for stabilizing the final conformation than for forming it in the first place.

Determining whether this possibility is correct requires an ability to gauge the strength of electrostatic interactions. Yet the mathematics is complicated by the fact that atoms in a folding protein are often separated by water, which can mute the long-distance attractions or repulsions in ways difficult to estimate in the absence of detailed structural information. Moreover, as the protein folds, the distances between the atoms constantly change, which adds further complexity.

The precise effects of hydrophobic, steric and electrostatic interactions, then, remain a matter of conjecture. Research into protein folding, however, is proceeding enormously faster today than in the past. Those of us involved in the effort still cannot "play the music," but we are rapidly learning certain of the notes. That progress alone is heartening, as is knowing that a solution to the folding problem will resolve a question of deep scientific interest and, at the same time, have immediate application in biotechnology.

FURTHER READING

PROTEIN FOLDING. Charis Ghélis and Jeannine Yon. Academic Press, 1982.

SPECIFIC INTERMEDIATES IN THE FOLDING REACTIONS OF SMALL PROTEINS AND THE MECHANISM OF PROTEIN FOLDING. Peter S. Kim and Robert L. Baldwin in *Annual Review of Biochemistry*, Vol. 51, pages 459-489; 1982.

PROTEINS, STRUCTURES AND MOLECULAR PROPERTIES. T. E. Creighton. W. H. Freeman and Company, 1983.

FOLDING AND ASSOCIATION OF PROTEINS. Rainer Jaenicke in *Progress in Biophysics & Molecular Biology*, Vol. 49, No. 1, pages 117-237; 1987.

PROTEIN FOLDING: DECIPHERING THE SECOND HALF OF THE GENETIC CODE. L. M. Gierasch and J. King. American Association for the Advancement of Science, 1990.