Perspective

The Protein Folding Problem

THOMAS E. CREIGHTON

UNTIL RECENTLY, THE "PROTEIN FOLDING PROBLEM" APpeared to be a largely esoteric subject. The complexity of protein structures made them difficult to comprehend and impossible to predict. The other aspect of the problem, of how linear polypeptide chains actually folded into these complex three-dimensional structures, was primarily a matter of speculation. Little experimental information was available; unfolded proteins and incomplete structural domains appeared to be random coils. Most folding transitions that were examined were cooperative, and partially folded intermediates that would define the folding pathway were never present in substantial quantities. Kinetic attempts to elucidate folding pathways were bedeviled by the heterogeneity of the unfolded protein, which was caused by slow cis-trans isomerization of peptide bonds adjacent to proline residues. Folding occurred in vivo so efficiently that it was effectively invisible and virtually impossible to study.

That much has changed became clear at a special 3-day seminar at the AAAS Annual Meeting in Boston. A large and enthusiastic audience was treated to a broad survey of the field, ranging from the conformational properties of small peptides to in vivo folding and assembly of complex viral proteins, from the quest for a better detergent (in this case, an improved subtilisin) to the medical consequences of defects in collagen folding.

It is not yet possible to predict a three-dimensional structure from just the amino acid sequence, except by homology with a protein of known structure. Nevertheless, understanding the basic rules of protein architecture is now well advanced, and it is becoming possible to design folded structures de novo. The group of W. DeGrado (Du Pont) has designed a polypeptide of relatively simple sequence that appears to adopt, at least approximately, the intended "four-helix bundle" conformation. Most impressive was the stability of this conformation. Under optimal conditions, it has a net stability of about 15 kcal/mol, which is greater than that of most natural proteins of similar size. Once its folded conformation is determined in detail, this simple protein should be ideal for examining systematically the basis for its stability; it might even be modified to incorporate biological functions.

The new technology of protein engineering has made it possible to produce and study experimentally any protein (whether natural, modified, or totally synthetic) and has given a major experimental impetus that was evident in many of the presentations. Often, the greatest difficulty is deciding which of the vast number of possible alterations to make. The prudent investigator, if in doubt, devises elegant genetic procedures to select for those replacements that give the desired effect. Besides amino acid changes that alter protein stabilities, J. King (Massachusetts Institute of Technology) showed how mutations that block folding and assembly pathways can be isolated and characterized.

Unfolded proteins are statistical ensembles of many different, interconverting conformations, but just how random they are is difficult to assess. Nevertheless, local nonrandom conformation has now been demonstrated unambiguously in some peptides. Some short α helices are more stable in isolation than would be predicted by the classical helix-coil parameters; the parameters have been measured using the host-guest technique, however, and their relevance to short peptides of mixed sequences is uncertain. Extensive work on the relatively stable α helix of the ribonuclease S peptide has enabled S. Marqusee and R. L. Baldwin (Stanford University) to design peptides 16 residues long that are up to 80 percent helical in water. Furthermore, two examples of β turns in penta- and hexapeptides that are significantly populated in water at low temperatures were demonstrated by P. Wright (Scripps Clinic) using nuclear magnetic resonance (NMR).

Whereas some feel that the presence of such small elements of local structure in an unfolded protein must be significant for initiating folding, others point out that a sequence of five residues has no substantial tendency to adopt the same conformation in unrelated protein structures; also, the β turns observed in the small peptides are not present in the known folded conformations of proteins that contain these sequences. Experimental studies of protein folding have almost invariably demonstrated that, in the absence of aggregation or covalent modification, the rate of folding is independent of which type of unfolded protein is used. Except for the cis-trans isomerization of proline, unfolded proteins appear to equilibrate rapidly under refolding conditions, often through compact but nonnative conformations that may approximate the recently recognized "molten globule" state. Subsequent folding occurs by a very limited number of pathways, and the rate-limiting step usually appears to occur very late, upon approaching the final folded conformation. The popular theoretical models of folding that involve local nucleation events followed by rapid completion of folding appear to be incompatible with most experimental observations.

Partially folded intermediates that would define a folding pathway are usually unstable and not populated at equilibrium; exceptions to two-state behavior are generally either due to a stable molten globule state or to multiple, independently folding structural domains that are apparent in the folded state. One remarkable exception was reported by C. R. Matthews (Pennsylvania State University). The tryptophan synthetase alpha subunit can be cleaved proteolytically and separated into two folded fragments that appear to unfold independently within the intact protein. Nevertheless, only a single domain is apparent in the crystal structure of this protein (*1*), which comprises an eight-stranded α/β barrel of alternating α helices and β strands, of the type first found in triosephosphate isomerase. The two proteolytic fragments correspond to α_5/β_6 and α_3/β_2 elements of secondary structure of the α_8/β_8 barrel. The most likely conclusion from this is that partial β barrels are much more stable than would have been expected. It is tempting to suggest that this indicates that α_8/β_8 barrels and other α/β structures may be constructed from relatively stable substructures. This may be why α_8/β_8 barrels have been found at least a dozen times in apparently unrelated proteins; they would then be the most probable candidates for convergent evolution to the same protein conformation. Much remains to be determined about the conformational properties of protein fragments, which are not always unfolded.

Pathways of folding, defined by the intermediates and the transition states by which the intermediates are formed and disappear, can be determined when the folded conformation requires disulfide bonds, since disulfides can be trapped. Understanding the most

The author is at the MRC Laboratory of Molecular Biology, Cambridge, England.

(Continued on page 344)

Books Received

Advances in Age Pigment Research. E. Aloj Totaro, P. Glees, and F. A. Pisanti, Eds. Pergamon, Elmsford, NY, 1987. x, 427 pp., illus. $94. Advances in the Biosciences, vol. 64. From a workshop, Vico Equense, Italy, May–June 1985.

Alkaline Rocks and Carbonatites of the World. Part 1, North and South America. Alan R. Woolley. British Museum (Natural History), London, and University of Texas Press, Austin, 1987. vi, 216 pp., illus. $65.

Applications of Control Theory in Ecology. Y. Cohen, Ed. Springer-Verlag, New York, 1987. viii, 101 pp., illus. Paper, $16.30. Lecture Notes in Biomathematics, vol. 73. From a symposium, Syracuse, NY, Aug. 1986.

Applied Geodesy. Stuart Turner, Ed. Springer-Verlag, New York, 1987. viii, 393 pp., illus. Paper, $44.60. Lecture Notes in Earth Sciences, vol. 12. From a course, Geneva, Switzerland, April 1986.

Atlas on the Biology of Soil Arthropods. Gerhard Eisenbeis and Wilfried Wichard. Springer-Verlag, New York, 1987. xiv, 437 pp., illus. $120.

The Biology of Dinoflagellates. F. J. R. Taylor, Ed. Blackwell Scientific, Palo Alto, CA, 1987. xii, 785 pp., illus. $180. Botanical Monographs, vol. 21.

Biosphere. The Realm of Life. Robert A. Wallace, Jack L. King, and Gerald P. Sanders. Scott, Foresman, Glenview, IL, 1987. xxviii, 780 pp., illus., + appendices and index. $33.

Cellulose Hydrolysis. L. T. Fan, M. M. Gharpuray, and Y.-H. Lee. Springer-Verlag, New York, 1987. xiv, 198 pp., illus. $101.50. Biotechnology Monographs, vol. 3.

Cerebral Cortex. Vol. 6, Further Aspects of Cortical Function, Including Hippocampus. Edward G. Jones and Alan Peters, Eds. Plenum, New York, 1987. xiv, 464 pp., illus. $69.50.

Cerebrovascular Diseases. William J. Powers and Marcus E. Raichle, Eds. Raven, New York, 1987. xxii, 374 pp., illus. $98. From a conference, St. Louis, MO, March 1986.

Chaos in Biological Systems. H. Degn, A. V. Holden, and L. F. Olsen, Eds. Plenum, New York, 1987. xii, 323 pp., illus. $62.50. NATO Advanced Science Institutes Series A, vol. 138. From a workshop, Cardiff, Wales, Dec. 1986.

Clouds in a Glass of Beer. Simple Experiments in Atmospheric Physics. Craig F. Bohren. Wiley, New York, 1987. xx, 195 pp., illus. Paper, $12.95. Wiley Science Editions.

Dipeptides as New Substrates in Nutrition Therapy. Peptide als neue Substrate in der Ernährungstherapie. S. A. Adibi *et al.*, Eds. Karger, Basel, 1987. viii, 196 pp., illus. $42. Contributions to Infusion Therapy and Clinical Nutrition, vol. 17. From a symposium, Erlangen, F.R.G., July 1986.

Elites and the Idea of Equality. A Comparison of Japan, Sweden, and the United States. Sidney Verba *et al.* Harvard University Press, Cambridge, MA, 1987. xii, 331 pp., illus. $30.

Energy Nutrition of Women. Geoffrey H. Bourne, Ed. Karger, Basel, 1987. x, 265 pp., illus. $162. World Review of Nutrition and Dietetics, vol. 52.

Forestry and the Forest Industries. Past and Future. E. G. Richards, Ed. Nijhoff, Dordrecht, 1987 (U.S. distributor, Kluwer, Norwell, MA). xii, 428 pp., illus. $92.50. Forestry Sciences.

From Hamiltonians to Phase Diagrams. The Electronic and Statistical-Mechanical Theory of sp-Bonded Metals and Alloys. Jürgen Hafner. Springer-Verlag, New York, 1987. xii, 404 pp., illus. $69.50. Springer Series in Solid-State Sciences, vol. 70.

A Handbook of Mexican Roadside Flora. Charles T. Mason, Jr., and Patricia B. Mason. University of Arizona Press, Tucson, AZ, 1987. iv, 380 pp., illus. Paper, $19.95.

The Health of Nations. True Causes of Sickness and Well-Being. Leonard A. Sagan. Basic Books, New York, 1987. vi, 233 pp., illus. $19.95.

Host-Parasite Cellular and Molecular Interactions to Protozoal Infections. K.-P. Chang and David Snary, Eds. Springer-Verlag, New York, 1987. xx, 425 pp., illus. $115. NATO Advanced Science Institutes Series H, vol. 11. From a workshop, Acquafredda di Maratea, Italy, Sept.–Oct. 1986.

How Experiments End. Peter Galison. University of Chicago Press, Chicago, 1987. xii, 330 pp., illus. Paper, $14.95.

Immunology. A Short Course. Eli Benjamini and Sidney Leskowitz. Liss, New York, 1987. xxii, 390 pp., illus. Paper, $22.95.

Keeping the Dream Alive. Putting NASA and America Back in Space. Michael C. Simon. Earth Space Operations, San Diego, CA, 1987. vi, 218 pp. + plates. Paper, $11.95.

Mathematical Aspects of Hodgkin-Huxley Neural Theory. Jane Cronin. Cambridge University Press, New York, 1987. xii, 261 pp., illus. $49.50. Cambridge Studies in Mathematical Biology, vol. 7.

The Measure of Times Past. Pre-Newtonian Chronologies and the Rhetoric of Relative Time. University of Chicago Press, Chicago, 1987. x, 292 pp. $27.50.

The Measurement of Moral Judgment. Anne Colby and Lawrence Kohlberg. Cambridge University Press, New York, 1987. Two vols. Vol. 1, Theoretical Foundations and Research Validation. x, 397 pp., illus. Vol. 2, Standard Issue Scoring Manual. x, 977 pp. Spiral bound. $165.

Mechanistic Approaches to Interactions of Electric and Electromagnetic Fields with Living Systems. Martin Blank and Eugene Findl, Eds. Plenum, New York, 1987. x, 443 pp., illus. $85.

Meteorite Craters. Kathleen Mark. University of Arizona Press, Tucson, AZ, 1987. xvi, 288 pp., illus. $29.95.

Search for a Supertheory. From Atoms to Superstrings. Barry Parker. Plenum, New York 1987. x, 292 pp., illus. $21.95.

The Second Creation. Makers of the Revolution in Twentieth-Century Physics. Robert P. Crease and Charles C. Mann. Collier (Macmillan), New York, 1987. xii, 480 pp., illus. Paper, $11.95. Reprint, 1986 edition.

Second International Conference on Chemical Engineering Education. (Cambridge, U.K., April 1987.) D. C. Freshwater, Ed. Institution of Chemical Engineers, Rugby, U.K., 1987 (U.S. distributor, Pergamon, Elmsford, NY). viii, 398 pp., illus. $51.50. Institution of Chemical Engineers Symposium Series, no. 101.

Solvation Thermodynamics. Arieh Ben-Naim. Plenum, New York, 1987. xiv, 246 pp., illus. $49.50.

Survival Strategies for New Scientists. Carl J. Sindermann. Plenum, New York, 1987. 264 pp. $17.95.

System Design for Human Interaction. Andrew P. Sage, Ed. IEEE Press, New York, 1987. viii, 488 pp., illus. $69.95; to IEEE members, $52.50. IEEE Press Selected Reprint Series.

Systems and Control Encyclopedia. Theory, Technology, Applications. Madan G. Singh, Ed. Pergamon, Elmsford, NY, 1987. Eight volumes. lxxiv, 5611 pp., illus. $2300.

Taking Japan Seriously. A Confucian Perspective on Leading Economic Issues. Ronald Dore. Stanford University Press, Stanford, CA, 1987. x, 264 pp. $35; paper, $11.95.

Taxonomy of Porifera. From the N. E. Atlantic and Mediterranean Sea. Jean Vacelet and Nicole Boury-Esnault, Eds. Springer-Verlag, New York, 1987. viii, 332 pp., illus. $89. NATO Advanced Science Institutes Series G, vol. 13. From a workshop, Marseille, France, Sept. 1986.

Technetium in Chemistry and Nuclear Medicine, 2. Marino Nicolini, Giuliano Bandoli, and Ulderico Mazzi, Eds. Cortina, Verona, 1987 (U.S. distributor, Raven, New York). xxii, 354 pp., illus., + appendix. $111.50. From a symposium, Padua, Italy, Sept. 1985.

Thermal Flow in Porous Media. Horia I. Ene and Dan Poliševski. Reidel, Dordrecht, 1987 (U.S. distributor, Kluwer, Norwell, MA). x, 194 pp., illus. $64. Theory and Applications of Transport in Porous Media.

Thin-Layer Chromatographic R_f Values of Toxicologically Relevant Substances on Standardized Systems. Deutsche Forschungsgemeinschaft. VCH, New York, 1987. 223 pp. $49.50. Report of the DFG Commission for Clinical-Toxocilogical Analysis, no. 7.

The Unfulfilled Promise of Synthetic Fuels. Technological Failure, Policy Immobilism, or Commercial Illusion. Ernest J. Yanarella and William C. Green, Eds. Greenwood, Westport, CT, 1987. xxiv, 225 pp. $39.95. Contributions in Political Science, no. 179.

Viable Populations for Conservation. Michael E. Soulé, Ed. Cambridge University Press, New York, 1987. xii, 189 pp., illus. $44.50; paper, $16.95.

The Viroids. T. O. Diener, Ed. Plenum, New York, 1987. xxii, 344 pp., illus. $59.50. The Viruses.

(Continued from page 267)

detailed known pathway, that of bovine pancreatic trypsin inhibitor (BPTI), requires further information about the conformations of the disulfide intermediates. The major one-disulfide intermediate, with the only disulfide bond between residues 30 and 51, was shown by T. G. Oas and P. S. Kim (Whitehead Institute) to be approximated remarkably closely by two synthetic peptides, corresponding to residues 20 to 33 and 43 to 58, linked by that single disulfide bond. Using genetic engineering techniques with intact BPTI, D. P. Goldenberg (University of Utah) demonstrated that drastic single amino acid replacements of internal residues made varying, but often substantial, alterations of up to 6 kcal/mol to the relative free energies of the various intermediates, the transition states, and the final folded conformation. The conformational properties of such partially folded intermediates may now be elucidated using NMR techniques; when combined with genetic engineering technology, this should permit the conformational forces that direct folding pathways to be elucidated experimentally.

Protein folding in vivo as a biological phenomenon is becoming increasingly amenable to study. Investigations of in vivo gene expression are uncovering many phenomena where folding (or the absence of folding) of the polypeptide gene product is apparent. These include (i) the need to prevent or reverse folding prior to translocating a polypeptide chain through a membrane, (ii) folding and assembling it once across the membrane, and (iii) targeting it to its appropriate cellular location. The involvement of other proteins in many of these steps suggests that it is dangerous to extrapolate protein folding from in vitro to in vivo circumstances. In any case, folding in vivo is certain to come under increasing experimental scrutiny. Finally, the dramatic medical consequences of single site mutations that alter procollagen assembly and folding in humans, described by P. Byers (University of Washington), left no doubt that the problem of protein folding is of more than academic interest.

REFERENCES

1. C. C. Hyde, E. A. Padlan, S. A. Ahmed, E. W. Miles, D. R. Davies, *Fed. Proc.* **46**, 2215 (1987); personal communication.