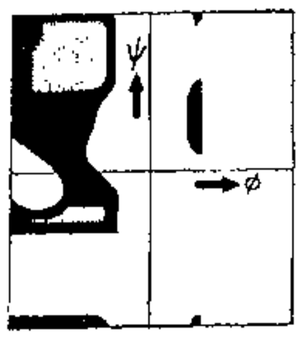


BOOK REVIEW

Protein Folding: Deciphering the Second Half of the Genetic Code. L. M. Gierasch and J. King, eds. American Association for the Advancement of Science, Washington, DC. 1990.

An outcome of a symposium held under the auspices of the American Association for the Advancement of Science, the book has seven sections — I. Structural themes in native proteins, II. Interactions and conformations of amino acids and peptides, III. Recovering active proteins, IV. Intermediates in protein folding, V. Protein folding within the cell, VI. Modelling protein folding structure, VII. Protein design: What can we get away with? These chapters traverse a wide terrain, enabling the conception of an exquisite mural wherein the aesthetics and mechanics of protein folding are magically blended. The reviewers were overawed not only by the coverage in the book but also by the myriads of subtle factors that play a role in protein folding.

The importance of protein folding cannot be overemphasized. The catalytic activity of enzymes and the unique properties of structural and membrane-channel proteins depend on the precise folding of their polypeptide chains. With all the armoury currently available in experimental and theoretical methods, we still cannot predict reliably or even comprehend rationally how the sequence of amino acids in a given polypeptide chain would determine and control the ultimate spatial arrangement of the protein. Even a peripheral comprehension of this domain will have far-reaching implications. Protein folding is at the fulcrum of current endeavours in recombinant-DNA technology and protein engineering.

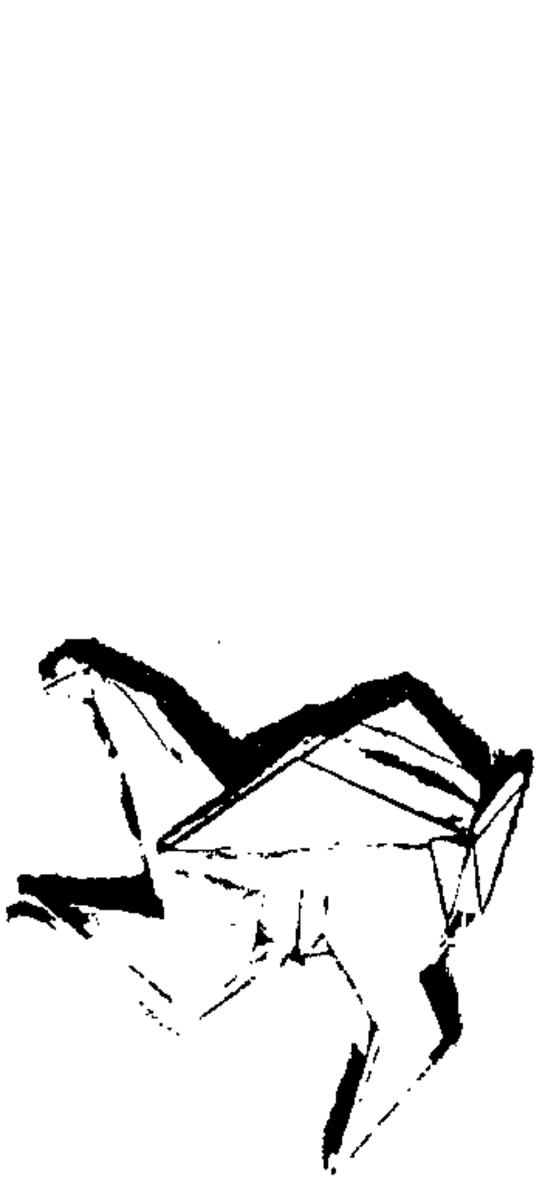
The splendour of protein folding and its relation to another creative arena, namely art, has been effectively brought out by Richardson and Richardson. By an innovative and fascinating analysis, they have compared the folding of proteins to origami, the Japanese art of folding paper that originated around the sixth century. Although largely intuitive, almost all origami creations can be analysed in terms of surfaces generated from straight lines, the relationship between the surfaces so created, and the way arrangement of surfaces can create three-dimensional forms. The magic of origami lies in the fact that, from a two-dimensional sheet of plain paper, a cornucopia of three-dimensional structures can be created with precise straight-line folding. This art has been mathematically analysed and one of the pioneers in this area is T. Sundara Row, whose treatise *Geometric Exercises in Paper Folding* was published in 1893! A reprint of this work is currently available from Dover Press, New York. Interestingly, the relationship between the famous Karmarkar algorithms and *origami* has been analysed in popular science journals.

Richardson and Richardson have, in a highly evocative fashion, compared the protocols in origami with pivotal facets of protein folding, such as folding transitions, convergent evolution, the co-operative behaviour of intermediates, domains, packing arrangements between units of secondary structure, and protein engineering and protein design. The starting point in origami is a plain square sheet of paper and that in protein folding a one-dimensional primary sequence, both devoid of relief. The metamorphosis of both these forms to intricate three-dimensional objects has latent aspects whose comprehension would lead to the understanding of the principles that govern the process. Convincing comparison has been drawn between a pair of origami flowers, patterned after traditional folding, to the two similar domains of the eye lens protein γ-crystallin. The convergent aspect of these two operations has been highlighted with a pair of swans and a pair of proteins, namely trypsin and subtilisin. Both proteins have similar functions and their active sites, comprising serine, histidine and aspartic acid, perfectly superimpose; however, nothing else is common in the rest of their folds. This would suggest that these two proteins must have evolved the same mechanism for proteolysis quite independently. In a similar manner a pair of swans can be created from different folding patterns, except for the formation of the head. Recent results have shown that folding of proteins may proceed through intermediates, some of which may turn inside out and rearrange. Many origami creations follow a similar protocol. Yet another aspect is that, both in origami and protein folding, the final object is seldom recog-

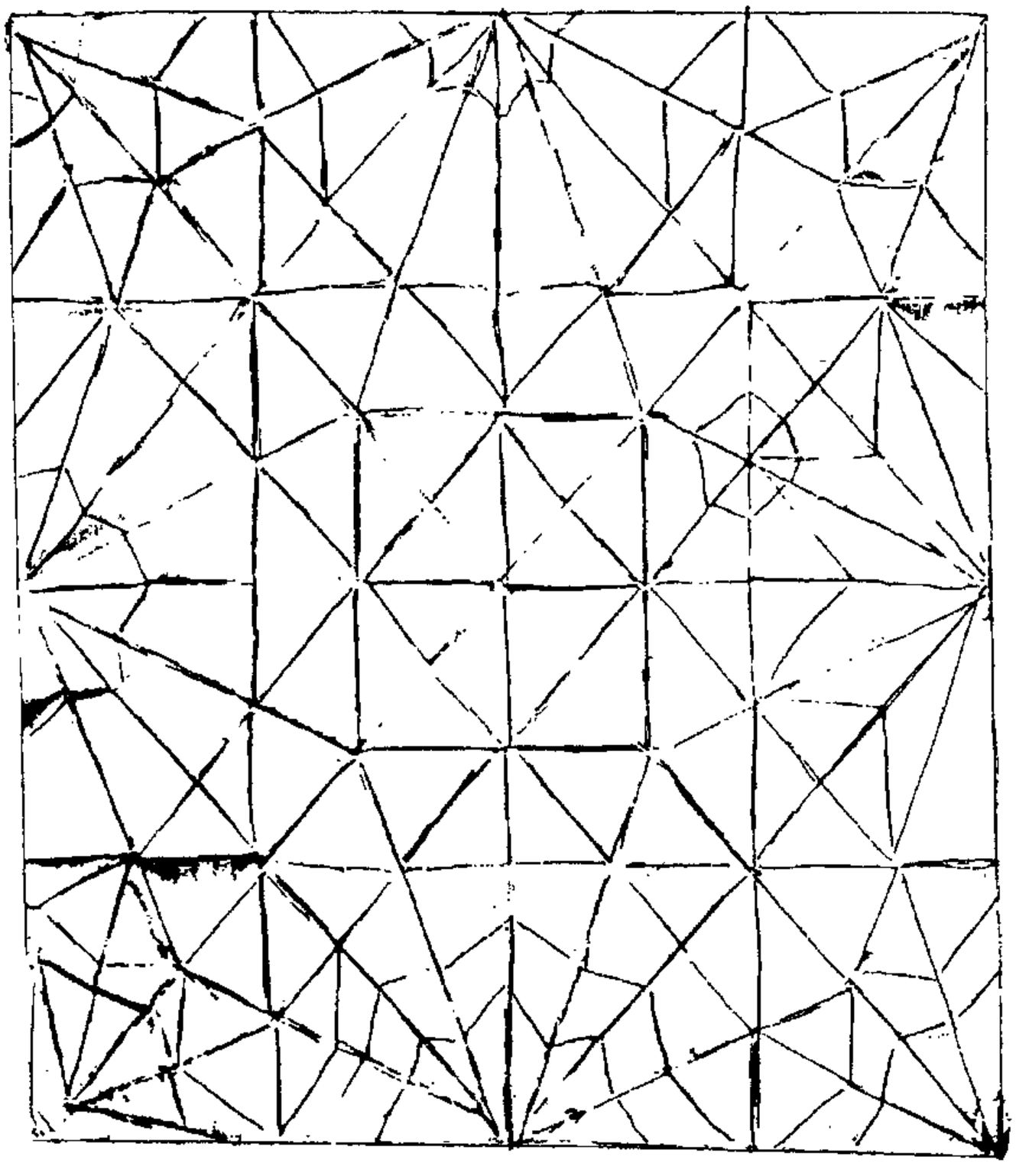
nizable till the very end. It would indeed require a very astute mind to discern what the final product would look like by either looking at a primary sequence or a square sheet of paper where the foldings that would lead to the product are marked. One can illustrate this with the origami form of Pegasus, the flying horse of greek mythology. The form shown in A in Figure 1 is functional in the sense that pulling the legs outward would lead to flapping of the wings! In B, the square paper that would result on opening the Pegasus is shown. The lines in B are akin to the primary sequence in proteins, whose further transformations, although rational, could be extremely difficult to put it in a formalism. The lines, such as depicted in B, and the primary sequence of a protein could be folded and unfolded in a reversible manner. Of course, just as the protein could be denatured in an irreversible manner, the origami design can be destroyed by tearing up the paper! Richardson and Richardson conclude that, both in origami and protein folding, it would be advantageous to focus on the processes involved rather than make deductions from examination of the final forms.

Omega loops, so named because of their resemblance to the Greek letter, are recently recognized structural motifs and are prevalent particularly in globular proteins. An incisive analysis of this motif has been made by Petrow and Rose. Omega loops are situated at the molecular surface and can be 6 to 16 residues in size. Interestingly, although omega loops do not possess regular secondary structures, they are considered important in macromolecular recognition, such as those involved in glycosylation, phosphorylation, supramolecular assembly and transport, since these phenomena are associated with the periphery of the protein molecule. Omega loops are appealing candidates for protein engineering studies. Thus a loop may be swapped or excised and the properties of the modified protein could be studied. From an evolutionary viewpoint, omega loops are often found in homologous locations, as can be exemplified with cytochrome *c* and serine proteases. The molecular biology of the loop motif promise exciting developments.

One can drown a hydrophobic protein! This is possible because of the highly ordered water clusters present at the protein surface. The role of these



A



B

clusters in the folding of hydrophobic proteins have been analysed by Teeter.

The importance of structural features present in the coiled-coil helix and the triple helix—the two superhelical motifs—in the understanding of folding in fibrous proteins has been highlighted by Brodsky.

Urry has analysed the energetics pertaining to protein folding leading to the development of a polymer system capable of amplifying small changes to such a degree that interaction energies can be translated to mechanical work.

Several facets having a bearing on helical peptides are lucidly analysed by Karle and Balaram. The importance of the α-aminoisobutyryl residue in controlling helix formation and in the promotion of 3_{10}- to α-helix transition has been illustrated. The crystal structure analysis of peptides with well-resolved scattering data has provided detailed information pertaining to a number of factors that can control the eventual spatial configuration.

Factors that initiate, stabilize, localize and terminate α-helical structures in peptides have been analysed by Marqusee and Baldwin, with focus on contributions from charged side-chains capable of forming salt bridges. These authors have designed a monomeric 17-residue model peptide containing only alanine, lysine and glutamic acid. This peptide adopts a stable α-helical conformation in water at 1°C.

A novel combination of immunological screening and 2D NMR has been used by Wright, Dyson, Waltho and Lerner for studies of protein folding of peptide fragments of proteins in aqueous solution present in very small amounts.

The studies by Sparrow and Goto on lipid interactions using model peptides has resulted in the suggestion that amphipathic structures can be an alternative to α-helical lipid-binding motif.

Recovery of active proteins from inclusion bodies—proteins with misfolded states—is a challenging task. Preferential degradation of these inclusion bodies would be an attractive approach. This aspect has been examined by Nilsson, Kuntz and Anderson, using bovine pancreatic trypsin inhibitor (BPTI)—perhaps the best understood globular protein—as a model to generate specific mutants that have perturbed folding behaviour. The study of folding mutants suggests a new pathway of protein turnover in *E. coli* in which rapidly folding domains in secreted molecules appear to be preferentially degraded.

Early events in protein folding are difficult to discern because the intrinsic co-operative behaviour makes these intermediates hard to isolate. A novel solution to this problem has been developed by Oas and Kim. Their synthetic strategy comprised preparation of two short peptide segments around the critical 30–51 disulfide bond in BPTI, their oxidative fusion and comparison of the folding of the fused peptides with that of the native protein. Such a comparison has shown that the synthetic compound is a good model for an intermediate state in the folding of BPTI. The advantage here is that further folding is not possible, thus permitting studies on the intermediate state.

The folding efficiency of proteins to native forms *in vivo* is seriously affected by secondary reactions that are mostly destructive. For example, the folding of bovine growth hormone (bGH) is usually connected with aggregation and precipitation of an intermediate. In view of the clinical importance of bGH, preventing this aggregation, which will enhance the folding efficiency, is of interest. Studies by Brems with mutants have shown that modulation of the aggregation behaviour is feasible.

In an evolutionary sense, selective pressure for efficient folding may be one of the most important determinants of protein primary structure. Studies by Krueger, Stock, Scutt and Stock have shown that the requirement for efficient protein folding may be as important a constraint on the primary sequence as the need to correctly position residues involved in, say, the active site of enzymes.

Goldberg has made the pertinent observation that, hardly a decade ago, few accepted the notion that the subject of protein folding was worth expenditure of time and money. Today everybody agrees that much effort has to be devoted here. This change in perception arises from many factors, the important ones being the importance of the biological profile of folding intermediates and the dire need to improve the efficiency of renaturation pathways of proteins produced by genetic engineering techniques.

Insight into the early and late events in protein folding is possible arising from studies by Creighton as a result of extensive investigation of the folding behaviour of BPTI.

Prediction of 3D structures on the basis of primary sequence is not yet fully feasible. A converse approach would be to design an amino acid sequence that will fold to give a predictable 3D structure. Regan, Ho, Wasserman and DeGrado have succeeded in designing and synthesizing a novel protein that folds into a remarkable stable bundle of four α-helices.

As a result of satisfactory fit between theory and experiments, Holtzer, Holtzer and Skolmik have been able to suggest that the unfolding equilibria in two-chain coiled-coil structures, such as is present in tropomyosin, are best represented by a continuum of partly folded intermediates.

Matthews has examined the effect of amino-acid replacements in the folding of the α-subunit of tryptophan synthase. Biophysical and mutagenic studies, in conjunction with X-ray structure studies, have

provided insight into the folding mechanisms. It has been shown that the folding of a single structural domain can occur in stages that correspond to recognizable elements of secondary structure. The rate-limiting step in folding is the association of the amino and carbonyl folding units.

The rate-limiting steps in the folding of yeast cytochrome *c* have been probed by Nall. These studies have brought out the pivotal role played by the isomerization of X–proline linkages in the formation of facets of tertiary structure (burying tyrosines!)

A study of the conformation and interactions of signal peptides by Gierasch has shown that these have a high tendency to form stable α-helical configuration in interfacial environments and, in addition, a high affinity for lipid membranes.

Leader sequences retard folding rates. The importance of this in the export of proteins has been lucidly analysed by Park, Liu, Topping, Cover and Randall.

Studies by King, Fane, Pettingwell, Mitraki, Villfane and Yu have demonstrated again the presence of sequences within polypeptide chains that function primarily in determining the conformation of folding intermediates. Denaturation of these intermediates results in loss of native structure.

The folding of the triple-helical domain of collagen has been examined by Byers using mutants ranging from single amino-acid substitutions to large deletions. These studies have shown that folding alterations have adverse biological effects that could range from mild bone fragility to death in the prenatal period.

Theoretical aspects of protein folding have been examined by Cohen, Gregoret, Presnell and Kunz; Bruccoleri, Haber and Novotny; Smarr; and Bashford, Karplus and Weaver. The folding problem, most fundamentally stated, seeks to unravel the formal code that relates the amino-acid sequence of the protein to its 3D structure. These authors deal with space- and time-related problems of assembly, sampling of conformational space, developments in computing techniques related to this topic, and a novel theoretical frame that has implications for protein folding.

Fascinating approaches to protein design are dealt with in the final section. Kemp and Bowen have crafted a synthetic model using a conformationally constrained analogue of a segment of β-structure as the basic template to form the central two strands of a four-strand β-sheet. It was envisaged that this rigid segment would provide two sets of correctly oriented H-bonding groups and thus act as a nucleation centre for β-sheet formation. Preliminary results indicate that this could be an appropriate test system in β-sheet formation studies.

Carter and Wells have shown that in subtilisin, when the histidine residue at the active site was replaced by alanine, proteolytic activity was only detected against peptides with histidine in the −2 position, as the model predicted. Although activity was low, this opens up possibilities for other such 'substrate-assisted' processes.

Boyd, Manoil, Froshauer, Millan, Green, McGovern, Lee and Beckwith have developed a double marker system to enable the determination of the topography of membrane proteins of unknown structure. Alkaline phosphatase is unable to assemble in the cytoplasm, but shows high activity in the periplasm. Conversely β-galactosidase cannot traverse the membrane, but is active in the cytoplasm. Thus high alkaline phosphatase activity indicates fusion to a periplasmic domain and high β-galactosidase activity implies fusion to a cytoplasmic domain. Thus cytoplasmic and periplasmic domains can be identified along the primary structure of the protein.

The reviewers found the book exceptionally interesting. It not only provides pathways and pointers pertaining to future developments but also elegantly blends diverse disciplines of science to provide some understanding of one of the key phenomena associated with life, namely protein folding.

DARSHAN RANGANATHAN
SUBRAMANIA RANGANATHAN
Department of Chemistry
Indian Institute of Technology
Kanpur 208 016

Edited and published by Prof. S. Ramaseshan, Current Science Association, Bangalore 560 080.
Typeset and printed at Book-Makers (I) Pvt. Ltd., Bangalore 560 009.