PROTEIN FOLDING

An unacknowledged problem for structural genomics?

Daniel J. Goldstein

It took more than 50 years, but it happened. Molecular biologists—perhaps not so quick after all—finally realized what J.D. Bernal knew all along: that protein structure was the key to biology, and the advent of structural genomics must be celebrated as the real beginning of making sense of the mind-boggling accumulation of sequences. Yet apart from the technical and economic problems that remain to be solved, structural genomists may also have to deal with a more fundamental obstacle to their approach of assigning meaningful structures to raw sequences based on protein folding classes determined from high-throughput crystallography.

Commenting on the role of chaperonins in protein folding, R. John Ellis[1] remembered David Green's remark on the difference between human design and the realities of nature: A clever engineer can make a vacuum cleaner from the wreck of a car, but this does not prove that cars contain vacuum cleaners. This reflection is also appropriate to structural molecular biology. Do the structures determined by X-ray crystallography and nuclear magnetic resonance represent the whole and final story of the actual shape of a protein in its natural milieu?

We have all been educated in the gospel of Anfinsen's proposal that the primary structure of a polypeptide determines unequivocally its tertiary structure. The millions of denaturation-renaturation experiments that have been performed in laboratories all over the world bear witness to the extraordinary importance of this insight. However, its generality was tempered by reality when in 1993 Cy Levinthal and his associates[2] showed that the functional channel-forming domain of colicin E1, which crystallizes in a myoglobin-like fold, undergoes a dramatic change in three-dimensional structure in the context of a lipid bilayer, by inserting two central hydrophobic helices into the membrane interior.

The three-dimensional fold of the colicin membrane spanning domain is present also in several other toxins (e.g., the pore-forming domains of insecticidal delta-endotoxins, exotoxin A, and diphtheria toxin)[3], and, surprisingly enough, in the antiapoptotic and proapoptotic members of the protein Bcl-2 family, such as Bcl-2 and Bax[4]. Both Bcl-2 and Bax form membrane pores, presumably by the same mechanisms as colicin E1, and interact with each other[5]. These results clearly show that the modules of secondary structure that make up a given protein species are not rigidly constrained in a single set of interactions that lead to a unique three-dimensional structure. Most recently, work with prions showed that a single polypeptide chain can not only adopt several different tertiary structures, but can also serve as a template for fixing the configuration of other proteins[6,7].

Not only in prion proteins can a single sequence adopt two folds. The amino-terminal domain of the lymphocyte adhesion molecule CD2, a molecule belonging to the immunoglobulin superfamily, can adopt either a monomeric or a dimeric form, the dimer being formed by the *intercalation* of two polypeptide chains[8].

Since the pioneering experiments with hemagglutinin and the serpins, it has been known that pH changes or proteolytic cleavage induce drastic refolding. The fact that in a homosocial environment (a sample of a single protein species purified to homogeneity) a polypeptide chain adopts a certain secondary structure does not mean that some of its segments cannot adopt different secondary structures when exposed to a different environment. This was experimentally shown by Minor Jr et al.[9] who constructed a chameleon sequence that folds either as an extended β-strand or as an α-helix depending on the position in which it is inserted in a carrier protein. This striking result shows that the conformation of the polypeptide can be critically dependent on the molecular matrix in which it is located[9].

Taken together, these experiments indicate that the spatial arrangement of the modules of secondary structure changes as a function of the biochemical context. In this sense, Oleg Ptitsyn's "molten globule" should not be considered a metastable intermediate between a single native and a single denatured structure, but an intermediate between a single denatured structure and a variety of native forms, determined by the chemical atmosphere in which the polypeptide is immersed and the interactive opportunities that it can provide.

The fact that specific proteins are usually identified while searching for a catalytic or structural element that fulfills a given function does not necessarily mean that the described function is the only one that the protein can accomplish. One strongly suspects that the many crevices, canyons, depressions, and gaps that punctuate any protein surface are places that interact with numerous micro- and macromolecular ligands, inside the cell or in the extracellular spaces, the identity of which is not known. That is, the functional significance of these "pockets" and crevices, which may be unsuspected catalytic, associative, and regulatory sites, is uncertain. The multiple associations of the Bcl-2 protein with an ever increasing number of proteins involved in the regulation of apoptosis and cell proliferation, and the fascinating functional associations among glycogen synthetase kinase, cadherin, and the APC gene product, are just some examples that illustrate the variety of ensembles that make up the protein framework of the biological world. The discovery of the first endogenous anesthetic molecule, oleamide, by Richard A. Lerner[10] and his lucid conception of fluidity transmitters open the possibility that even the Cinderella domain of the membrane integral proteins—the membrane-spanning putative alpha-helices normally dismissed as mere anchors for the lipid bilayer—may be key allosteric sites for the regulation of the function of membrane receptors, channels, and transporters.

Moreover, the biological and chemical reality of the "disordered" segments found so often in crystallographic structures might not be disordered at all, but represent discrete sectors of the polypeptide chain that are only structured when interacting with other molecules. Williams et al.[11] very recently presented crystallographic evidence of massive refolding of the homodimeric cytochrome cd, nitrite reductase during catalysis. Upon substrate binding, the coordination of the iron of one of two hemes changes radically and the "disorder" of the first eight residues of the two domains found in the inactive enzyme is propagated to 35 amino acid long tracts. Therefore, the structure derived from the study of identical polypeptides floating in solution or forming crystals—i.e., the rarified experimental environments of structural molecular biology—may be only one of the possible three-dimensional folds.

Given these considerations, structural genomists might consider assigning a high priority to understanding the extent to which protein–protein and other molecular interactions determine native folding patterns before their databases get too full.

Daniel J. Goldstein is a professor in the Department of Biological Sciences, Faculty of Exact and Natural Sciences, University of Buenos Aires, Argentina. (postmast@djgol.uba.ar).

1. Ellis, R.J. 1996. *Science* **272:**1448–1449.
2. Shin, Y.-K. et al. 1993. *Science* **259:**960–963.
3. Parker, M.W. and Pattus, F. 1993. *Trends in Biochemical Sciences* **18:**391–395.
4. Muchmore S.W. et al. 1996. *Nature* **381:**335–338.
5. Reed, J.C. 1997. *Nature* **387:**773–776.
6. Prusiner, S.B. 1997. *Science* **278:**245–251.
7. Telling, G.C. et al. 1996. *Science* **274:**2079–2082.
8. Murray, A.J. et al. 1995. *PNAS* **92:**7337–7341.
9. Minor, D.L. Jr, and Kim, P.S. 1996. *Nature* **380:**730–734.
10. Lerner, R.A. 1997. *PNAS* **94:**13375–13377.
11. Williams, P.A. et al. 1997. *Nature* **389:**406–412.