

# nature Structural biology

december 1999  
vol. 6 no. 12

molecular form & function

## Integrative approaches to protein folding

Over the past four decades, much research has been focused on two central questions: what are the determinants of protein structure, and how does a polypeptide fold to its native state? Understanding how a protein folds is more than just a scientific curiosity; it also has practical applications. For example, this information could be helpful in designing proteins that are resistant to denaturation or to proteases, or in engineering proteins with new functions. It may also provide clues to the causes of certain diseases — such as Alzheimer's, which appears to involve misfolding of the amyloid protein. The importance of these widely different applications necessitates an understanding of the folding process both in the test tube and inside the cell.

### State of current research

The finding that protein folding can proceed, in some cases, without the assistance of cellular factors has allowed researchers to study simple folding processes *in vitro* using purified proteins. Refolding *in vitro* is typically initiated by diluting denatured proteins into a buffer that does not contain any denaturant; the dilution ratio is adjusted so that side reactions such as aggregation do not occur. Rapid mixing techniques in conjunction with NMR and fluorescence spectroscopy have facilitated structural characterization of putative intermediates in the folding reactions of several small proteins, including lysozyme, barnase and several SH3 domains. Advances in technology have also made it easier to conduct numerous studies of proteins with single mutations (the 'protein engineering' approach), which have allowed identification of residues that make important interactions at the rate-limiting steps of the folding reactions of these different proteins. It is hoped that the information obtained from examining these simplified systems *in vitro* can be generalized to aid understanding of the folding reactions of more complex proteins, both *in vitro* and *in vivo*, and to facilitate computational analysis and prediction of native state structures and folding reactions.

The success of the protein engineering approach and the advance of molecular biological techniques seem to have driven a shift in *in vitro* protein folding research from 'hypothesis-driven' to 'data-mining' approaches, whereby a large amount of detailed kinetic, thermodynamic, and mutagenesis data are generated for a number of proteins. While these data constitute an important resource for researchers who are developing theoretical and computational tools to predict the characteristics of protein folding reactions and to extract general folding themes, they are not likely to be of immediate interest to a broad audience. To date, most computational analyses have been performed on small proteins (<100 amino acids) with simplified side chain representations, and the results demonstrate that folding *in silico* may qualitatively reproduce some experimental observations. However, the need for even larger data sets, as well as the size and complexity of most proteins, still pose significant challenges for progress in the computational field. Moreover, extrapolating to the *in vivo* situation is difficult, since the *in vivo* folding environment is very different from both computer-simulated and *in vitro* conditions.

Inside the cell, the concentration of protein is much higher, necessitating the function of chaperones — proteins that assist the folding of other proteins and/or the assembly of oligomeric complexes. While the identities of cellular proteins that require one of the chaperones, GroEL,

for proper folding have only recently been investigated<sup>1</sup>, a general role of chaperones seems to be to block unproductive side reactions of the folding process by sequestering the polypeptide chain in an isolation chamber, or by holding the chain in a 'folding competent' state that is committed to fold upon release. However, the extent of folding *in vivo* is typically monitored by indirect methods, such as recovery of enzymatic activity or resistance to protease digestion. Thus, the structural details of the folding mechanism for most proteins and the details of chaperone–substrate interactions *in vivo* are lacking.

### Limitations and new approaches

While significant progress has been made in the protein folding field, each of the areas mentioned above has limitations. *In vitro* studies of small proteins promise 'high resolution' understanding of their folding processes, but the degree to which conclusions from these systems can be extended to more complex proteins is unclear. *In vivo* studies offer a physiologically relevant context for folding reactions, but, in general, only gross features are revealed. Theoretical studies offer the power of prediction, but they are currently limited to simple systems and require experimental data for input and validation. All of these limitations suggest that progress in this field will most likely come from studying many different proteins and from using integrative approaches.

Last month's issue of *Nature Structural Biology* presented three papers<sup>2–4</sup> that argued for the important role of overall topology, as opposed to the formation of individual side chain interactions, in the folding of small proteins. These papers describe experiments with four different proteins (two from each of two different structural families) and come to similar conclusions about the influence of topology. Thus, detailed work on several different systems has highlighted a potentially general feature of protein folding processes. Moreover, one of these papers contains both experimental and theoretical data on the same protein<sup>2</sup>, illustrating the trend toward integrative approaches in this field.

The paper by Andreas Matouschek and coworkers<sup>5</sup> on page 1132 of this issue of *Nature Structural Biology* also provides an example of this trend. They analyze the unfolding mechanism of a test protein, barnase, during protein import into mitochondria, a process with physiological implications, and compare the results to those obtained by studying the *in vitro* reaction with the pure protein. Their results show that the mitochondrial import machinery initiates protein unfolding at the N-terminus instead of within the core of the protein as is observed in *in vitro* experiments. Thus, this work bridges the gap between the *in vitro* and *in vivo* folding studies.

### The challenges ahead

Although there has been much progress in the past four decades, there are still many challenges ahead. The rapid accumulation of data on small proteins provides hope that researchers may be close to understanding the folding codes present in the amino acid sequences of some of these proteins. The next challenge will be to determine how a large multidomain protein folds — that is, do the principles obtained from studies on small proteins apply to more complex proteins? An even greater challenge is to successfully apply these principles to practical issues, such as those mentioned at the beginning of this editorial.

There are also numerous differences of opinion in the protein folding field on a number of issues, fueling debate and inspiring new approaches. For example, the finding that some small proteins fold without populating intermediates has led to questions about whether the intermediates observed in experiments with other proteins are productive or unproductive — that is, do these intermediates truly represent a step along an obligate folding pathway or do they represent 'dead-end' conformations that cannot continue to fold to the native state? This debate has sparked many investigations that aim to characterize structural elements occurring very early (on the order of submilliseconds) in the folding reaction and has highlighted the importance of characterizing the structures of unfolded states. Clearly the outcome of such debates in this field will have implications for general themes of protein folding.

While there has been an increase in the number of protein folding projects funded by the NIH in the last five years<sup>6</sup>, it is interesting to note that in an informal survey, many researchers who study *in vitro* folding reactions feel that it is becoming more difficult to obtain grants for their research. If this is indeed true, it may in part reflect a commonly cited frustration among scientists outside of the folding field that it is difficult to assess the success of such experiments in terms of the overall goal of understanding protein folding in general. Thus, as more detailed information is amassed on individual proteins, it will become even more imperative for those in the folding field to relate their work to other proteins and to the larger picture, through a more integrative approach that includes *in vivo*, *in vitro*, and *in silico* experiments.

1. Houry, W.A., Frishman, D., Eckerskorn, C., Lottspeich, F. & Hartl, F.U. *Nature* **402**, 147–154 (1999)
2. Riddle, D.S., Grantcharova, V.P., Santiago, J.V., Alm, E., Ruczinski, I. & Baker, D. *Nature Struct. Biol.* **6**, 1016–1024 (1999).
3. Chiti, F. *et al.* *Nature Struct. Biol.* **6**, 1005–1009 (1999).
4. Martínez, J.C. & Serrano, L. *Nature Struct. Biol.* **6**, 1010–1015 (1999).
5. Huang, S., Ratliff, K.S., Schwartz, M.P., Spenser J.M. & Matouschek, A. *Nature Struct. Biol.* **6**, 1126–1132 (1999).
6. Computer retrieval of information on scientific projects. <http://www-commons.cit.nih.gov/crisp/>