Pacific Symposium on Biocomputing 4:517-519 (1999)

DISORDER IN PROTEIN STRUCTURE AND FUNCTION

CHARLES R. KISSINGER
Agouron Pharmaceuticals, Inc., 3565 General Atomics Court, San Diego, CA 92121

A. KEITH DUNKER
Department of Biochemistry, Washington State University Pullman, WA 99164-4660

EUGENE SHAKHNOVICH
Department of Chemistry, Harvard University Cambridge, MA 02138

It is commonly assumed that a protein must attain a stable, folded conformation in order to carry out its specific biological function. Not all proteins conform to this simple view of protein structure and function, however. Certain regions within proteins, and in some cases entire proteins, are not ordered into a unique tertiary structure, but instead appear to exist as ensembles of structures. Protein structures determined by X-ray crystallography and NMR have revealed numerous such disordered regions, some of them quite extensive. Recent progress in predicting regions of disorder from amino acid sequence has provided evidence that these regions occur in nature with an unexpectedly high frequency[1]. There is now a growing awareness of the fundamental importance of disordered protein sequences in many biological processes.

Disordered protein sequences function in some cases to mechanically uncouple structured domains, making their dynamics less constrained. Linkers of this type are important in a diverse collection of proteins, from viral attachment proteins to transcription factors. Disordered regions also provide access for protease digestion, which is critical for the regulation of many important cellular processes. Disorder-to-order transitions in proteins may be one of the major factors in biomechanics, for example in the development of force by protein assemblies.

Disorder-to-order transitions may have a crucial role to play in macromolecular recognition. There are numerous examples of protein-protein, protein-nucleic acid, and protein-ligand interactions involving disordered protein segments. It has been postulated that disorder-to-order transitions provide a mechanism for uncoupling binding affinity and specificity[2], thereby permitting weak but highly specific interactions, or conversely, strong but relatively nonspecific interactions.

The papers in this session provide examples of the diverse set of biological processes in which disordered protein sequences participate. These papers also illustrate the diversity of techniques that can be productively applied to the study of protein disorder. Two papers demonstrate that the synthesis of experimental and computational results can yield new insights into the nature of disordered regions. Landry *et al.* use NMR data and dynamics simulations to examine the conformational preferences of a mobile loop that modulates the affinity of a chaperonin/cochaperonin complex. The authors also discuss why disordered loops might be essential for modulating the affinity of this complex. Mathieson, Penkett and Smith combine NMR results with theoretical predictions to provide evidence of non-random side chain interactions in a biologically active, yet unfolded, fibronectin binding protein. Their work shows that significant structural information can be obtained from a nearly completely unfolded protein.

Three other papers demonstrate how molecular biology techniques can be used to probe the characteristics of binding interactions that involve disordered regions. Makowski and Rodi present a novel application of phage-display technology in which peptides with an affinity for the anti-cancer drug taxol are shown to have sequence similarities to a disordered loop on Bcl-2, a protein implicated in breast cancer. Their work not only provides potential insights into the mechanism of action of taxol, but also into the nature of recognition processes involving disordered loops. Sessions *et al*. describe a highly unusual protein, CD2, in which there is interconversion between multiple metastable folds that occurs through exchange of β-strands. The authors use site-directed mutagenesis to manipulate the relative stabilities and interconversion rates of the alternative folds, and interpret their results in light of crystallographic and kinetic data. Shaiu, Hu and Hsieh present *in vivo* studies showing that poorly conserved protease-sensitive terminal regions of the eukaryotic topoisomerases I and II have essential functions in the nuclear import and targeting of these enzymes to transcriptionally active regions of the chromosomes. These new insights were reached using a combination of molecular biology approaches, which included the use of a novel fusion of the amino terminus of topoisomerase I to a reporter molecule, β-galactosidase, to determine the effects of this disordered segment on the cellular distribution of the reporter molecule.

The final two papers combine biophysical and biochemical approaches to gain information about disordered proteins. Nieslanik *et al.* use a combination of crystallographic and kinetic studies to probe the nature of the transition state for a ligand-dependent disorder-to-helix transition in wild type and mutant glutathione S-transferases. Yang *et al*. use fragment complementation to examine the stability and folding kinetics of thioredoxin reassembled from different sets of disordered fragments.

The very nature of protein disorder makes it a challenging subject of study. As the papers in this session demonstrate, advances in our understanding will require a multi-disciplinary approach combining experimental, computational and theoretical

results. We hope that this session will serve to increase awareness among the computational biologists assembled here, and also among a much wider audience, of this important yet frequently overlooked category of protein structure.

References

1. P. Romero, Z. Obradovic, C.R. Kissinger, J.E. Villafranca, E. Garner, S. Guilliot and A.K. Dunker, "Thousands of Proteins Likely to Have Long Disordered Regions", *Pac. Symp. Biocomp.* **3**, 437 (1998).
2. A.K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, Z. Obradovic, C. Kissinger and J.E. Villafranca, "Protein Disorder and the Evolution of Molecular Recognition: Theory, Predictions and Observation", *Pac. Symp. Biocomp.* **3**, 473 (1998).