Synopsis of the Thesis

Title: Computational and experimental studies on protein structure, stability and dynamics

by: Bharat V. Adkar

Thesis Supervisors:
Prof. Raghavan Varadarajan  Prof. Biman Bagchi
Molecular Biophysics Unit  Solid State and Structural Chemistry Unit
Indian Institute of Science
Bangalore-560012, India.

The work in this thesis focuses on the study of three main aspects of proteins, viz, Protein structure, stability, and dynamics. Chapter 1 is a general introduction to the topics studied in this thesis. Chapter 2 deals with the first aspect, i.e., protein structure in which we describe an approach to use saturation mutagenesis phenotypes to guide protein structure prediction. Chapters 3 and 4 discuss how to increase protein stability using surface electrostatics, and Chapter 5 details a method to predict whether a proline substitution in a given protein would be stabilizing or destabilizing. Hence, Chapters 3-5 can be associated with the second aspect, i.e., protein stability. The third aspect, namely protein dynamics, is dealt with in Chapters 6 and 7 which study conformational dynamics of adenylate kinase.

Protein structure prediction is a difficult problem with two major bottlenecks, namely, generation of accurate models and the selection of the most appropriate models from a large pool of decoys. In Chapter 2, the problem of model discrimination is addressed using mutant phenotype information derived from
saturation mutagenesis library. A library of ~1500 single-site mutants of the *E. coli* toxin CcdB (Controller of Cell Division or Death B) has been previously constructed in our lab. The pooled library was characterized in terms of individual mutant phenotypes at various expression levels which were derived from the relative populations of mutants at each expression level. The relative populations of mutants were estimated using deep sequencing. Mutational tolerances were derived from the phenotypic data and were used to define an empirical parameter which correlated with a structural parameter, residue depth. We further studied how this new parameter can be used for model discrimination.

Increasing protein stability in a rational way is a challenging problem and has been addressed by various approaches. One of the most commonly used approaches is optimization of protein core residues. Recently, optimization of protein surface electrostatics has been shown to be a useful approach for increasing stability of proteins. In Chapter 3, from analyses of a dataset of ~1750 non-homologues proteins, we show that proteins having a pI away from physiological pH, possess a significant fraction of unfavorably placed charged amino acids on their surface. One way to increase protein stability in such cases might be to alter these surface charges. This hypothesis was validated experimentally by making charge reversal mutations at putative unfavorable positions on the surface of maltose binding protein, MBP. The observed stabilization can potentially be increased by combining multiple individually stabilizing mutations. Different combinations of such mutations were made and tested in Chapter 4 to decide which mutants can be combined to achieve net stabilization. Ideas were tested through systematic experimentation which involved generation of two-site, three-site, and four-site mutations. A maximum increase in melting temperature (*T_m*) of 3-4 °C over wild-type protein was achieved upon combination of individually stabilizing mutants.

Proline (Pro) has two special stereo-chemical properties when it is a part of a polypeptide chain. First the φ value of Pro has a very constrained distribution and second, Pro lacks an amide hydrogen. Due to these properties, introduction of Pro might perturb stability/activity of the protein. In Chapter 5 we describe a procedure to accurately predict the effects of Pro introduction on protein stability. Pro scanning mutagenesis was carried out on the model protein CcdB and the *in vivo* activity of...
the individual mutants was also examined. A decision tree was constructed, using the special stereo-chemical properties of Pro to maximize correlation of predicted phenotype with the in vivo activity. Binary classification as perturbing or non-perturbing of every Pro substitution was possible using the decision tree. The performance of the decision tree was assessed on various test systems, and the average accuracy was found to be ~75%.

The role of conformational dynamics in enzyme catalysis has been explored in great detail in the literature. In Chapter 6, with the help of very long (350 ns), fully atomistic, explicit solvent molecular dynamics simulations, we studied conformational dynamics of adenylate kinase. We found the existence of a relatively stable state which lies intermediate between the open and closed conformations of the enzyme. The finding was further confirmed by computing a two dimensional configurational free energy surface when motions along each of the two movable domains (LID and NMP) are considered as reaction coordinates. We also discussed possible roles of the intermediate state during enzyme catalysis. The role of water in stabilization of intermediate states was also discussed. In Chapter 7, we studied dynamical coupling between LID and NMP domains of adenylate kinase during domain opening. Our observation suggests that the LID domain should start opening prior to the NMP domain. On the domain opening trajectory, the free energy surface of LID domain was found to be very rugged. We discuss a possible role of water in the ruggedness of the domain motions.

The Appendix contains 3 supplementary parts of the thesis. Appendix I is a mutant dataset obtained from 454 sequencing analysis. It includes the normalized number of reads per mutation at each expression level along with mutational sensitivity score. Appendix II is parameters used for one of the electrostatic calculations. Appendix III contains a list of PDB ids used for database analysis in surface electrostatics work discussed in Chapter 3.