

Evolutionary Analysis of Molecular Movements

by Thomas Darden, Jotun Hein and Mark Sansom.

Background

Comparative Biology is a major contributor to biological understanding and can be applied to any biological objects that are homologous. The large application area presently is sequences, but other rising areas include structures, networks, organs and more. The strength of evolutionary comparison is the ability to detect features of functional importance and thus select properties that demands a functional explanation from a sea of noise. The only requirement for evolutionary comparison is that the objects to be compared are homologous. In this sense movements are perfectly homologous as the movements associated an ancestral molecule would be inherited with modification, when the molecule evolved over evolutionary time. Movements of molecules have a series of special properties that singles them out compared to other comparative traits. If we imagined to compare the movements of globins over time, this would involve about 3000 atom positions, that could be observed in 10^9 time steps of size 10^{-15} seconds. One could imagine 2-10 globins had been observed and the total data set would be of the order 10^{10} spatial positions. Due to the increase in use of molecular dynamics, this kind of analysis has great potential and it is clear that analysis of such data would need entirely new approaches.

Proposed Research

Conformational change appears to be essential to the formation of highly specific but relatively low affinity binding complexes that characterize signal transduction in cells. Hubs, or proteins involved in multiple protein-protein complexes are characterized by having either disordered domains or high surface charge. Proteins containing intrinsically disordered regions have been reviewed recently by Dyson and Wright. They emphasize the importance of dynamics, through coupled folding and binding, to forming short term but highly specific interactions. Thus the study of the dynamics of conformationally variable proteins is an increasingly important topic within the structural biology of signalling processes.

Experimental approaches that have been brought to bear on this problem include X-ray crystallography, NMR and other spectroscopic approaches, calorimetry, and finally various single molecule techniques such as AFM. Of these crystallography provides the highest precision but is limited to a specific trapped conformation. Dynamics is inferred from B-factors and more importantly from comparison of different conformations of the same protein involved in different protein-protein or protein-ligand complexes. The spectroscopic and single molecule approaches provide more direct evidence of dynamics, but at a cost of structural precision. Recent advances in NMR techniques offer a particularly promising avenue of approach to both structure and dynamics.

Dynamics of proteins can be efficiently modeled by coarse grained approaches such as Normal Mode Analysis, that describe the low frequency collective modes of activation. This approach has been used to infer dynamics in protein superfamilies, and thus could be used to discuss the question of what characterizes the types of conformations sampled by homologous proteins. However, since these model the protein as a homogeneous elastic body, the efficiency of the approach comes at the cost of a loss of accuracy and of sequence specific information i.e. dynamics is necessarily determined entirely by the shape of proteins regardless of the identity of their amino-acid sequence. A far more precise description is provided by the technique of molecular dynamics simulations, which provides an atomic level description of protein motions. Currently these simulations are limited to the picosecond to the sub-microsecond time scale, and provide high precision but only semi-quantitative accuracy due to limitations in the force-fields. Comparative molecular dynamics of related proteins increases the reliability of the results and provides an avenue to study the evolution of dynamic modes and allowed conformations.

The study of unfolding and refolding of entire domains is currently not feasible within either of the above two theoretical approaches, and experimental results are few. However the EF-hand domain provides a tractable highly interesting example of a domain that is structurally conserved, but conformationally variable and small enough to study with a comparative molecular dynamics approach. These domains provide a "trigger" for signalling events by undergoing conformational change upon calcium binding. Calmodulin is a primary example of an EF-hand protein that has been studied extensively by experimental approaches, which offers firm ground for validating theoretical results. Many structures of family members are available, and there is extensive genomic information available. The likely conformational changes within the EF-hand domain (a coupled pair of a helix-loop-helix motifs) are limited enough to allow analysis, and the triggering event (calcium binding within a conserved loop) can also be studied by molecular dynamics approaches. Conformational change within the super-family has been studied by comparing conformations of the known structures and hypotheses about the structural cause of the triggering event as well as the importance of specific residues have been proposed. From these it is clear that this domain is capable of very large conformational variability, while retaining the basic helix-loop-helix structural units.

We propose a comparative molecular dynamics study of EF-hand calcium binding proteins. Starting structures for the studies will be selected from the 307 non-redundant PDB structures containing the EF-hand domain. The structural proposals by Gabarek can be addressed. In addition, principal components can be compared to those from a coarse grained normal modes analysis as well as the principal components analysis of structures. These can be compared with the structural classification from CATH or SCOP as well as evolutionary relationships derived from the extensive sequence information, to look for a picture of conserved dynamics from conserved shape, modulated by specific changes introduced by sequence variation to accommodate emerging functional needs in evolution.

Timeline, Supervision and Collaboration

The student will be jointly supervised by Thomas Darden (NIEHS, North Carolina, USA), Jotun Hein (Oxford, UK) and Mark Sansom (Oxford, UK). The scholarship starts October 2007 and is of 4 years duration. The first year will be focused on following courses given by the Systems Biology Doctoral Training Centre (Oxford). In the following 3 years the students will be based in both the US and UK. The project has great potential, but is also demanding as it covers Molecular Dynamics, Statistical Models of Evolution and also Software and Algorithm Design. Applicants must have strong qualifications in the relevant areas.

References

- Alexandrov, V., U.Lehnert, N.Echols, D. Milburn, D.Engelman and M. Gerstein (2006) "Normal modes for predicting protein motions: A comprehensive assessment and associated Web tool" *Protein Science* 14.633643.
- Babini, Bertini, Capozzi, Luchinat, Quattroni and Turano (2005) "Principal Component Analysis of the Conformational Freedom within the EF-Superfamily" *J. Proteome Research* 4.1961-1971
- Bahar, I. and AJ Rader (2005) "Coarse-grained normal mode analysis in structural biology" *Curr.Opin.Struct.Biol.*15.1-7 Gerstein, M. and W. Krebs (1998) "A database of macromolecular motions" *Nuc.Ac.Res.* 26.18.4280-90.
- Grabarek (2006) "Structural Basis for Delivery of the EF-hand Calcium-binding Proteins" *JMB* 359.509-625
- Maguid, S, S. Fernandez-Alberti, Leticia Ferrelli and J. Echave (2000) "Exploring the Common Dynamics of Homologous Proteins. Application to the Globin Family" *Biophysical J.* 89.3-13
- Pang, Arinaminpathy, Sansom and Biggin (2005) "Comparative Molecular Dynamics - Similar Folds and Similar Motions" *Proteins* 61.809-822
- Papaleo et al. (2006) "Flexible and enzymatic cold-adaptation: A comparative molecular dynamics investigation of the elastase family" *Biochemica et Biophysica Acta* 1764.1397-1406.
- Rader,AJ, Chennubhotia, L-W Yang and I. Bahar (2005) "The Gaussian Network Model: Theory and Applications"
- Shrivastava, IH and I. Bahar (2006) "Common Mechanism of Pore Opening Shared by Five Different Potassium Channels" *Biophysical J.* 3929-3940.