Towards creating complete proteomic structural databases of whole organisms

B. Jayaram, Priyanka Dhingra

Department of Chemistry,

Supercomputing Facility for Bioinformatics & Computational Biology,

School of Biological Sciences,

Indian Institute of Technology, Hauz Khas, New Delhi-110016, India

Email: bjayaram@chemistry.iitd.ac.in, priyanka@scfbio-iitd.res.in,

Tel No.: +91-11-2659 1505; +91-11-2659 6786; Fax: +91-11-2658 2037

Website: www.scfbio-iitd.res.in

Abstract

If structures of proteins of whole organisms were available, metabolomic models could be developed, drug targets could be identified, issues of affinity versus specificity could be sorted out and side effects and toxicity brought under control etc. all with greater levels of reliability. Advances in whole genome sequencing projects, annotation algorithms, growing protein sequence information with over half a million entries in the UniProtKB/Swiss-Prot database, progresses in structure based lead molecule design methodologies do uphold this optimism. However, x-ray and NMR structures of less than15% of the protein sequences are available in RCSB protein data bank. The diverging gap between sequence and structure calls for immediate in silico solutions. The biennial community wide structure prediction (CASP) experiments have considerably catalyzed structure prediction attempts world-wide and accuracies of computational models are continually increasing. While ab initio models have crossed the 100amino acid limit, it is still some way from the average sized human protein (~ 350 residues). Homology models which rely on the RCSB structures and the axiom that similar sequences adopt similar structures have been extremely powerful in providing high resolution structures limited only by sequence similarities. With dwindling similarities of query sequences with knowledge bases, newer ab intio /homology hybrid approaches are being explored to bring the structure prediction problem within the realm of feasibility in near future particularly for soluble proteins. The case of membrane bound protein sis still refractory. This review takes a stock of current protein tertiary structure prediction algorithms highlighting the problem areas to overcome and promises there of.