JPP JOURNAL OF PROTEINS AND PROTEOMICS
4(2), December 2013, pp. 65-66

From the Guest Editor

SOME RECENT PROGRESSES AND PERSPECTIVES IN PREDICTIONS OF PROTEIN STRUCTURE, DYNAMICS AND FUNCTION

B. Jayaram

Department of Chemistry, Kusuma School of Biological Sciences and Supercomputing Facility for Bioinformatics and Computational Biology, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016, India.

Specifying the three dimensional structure of a protein given its amino acid sequence continues to be a grand challenge problem. One of the major areas of application includes acceleration of structure based drug discovery. Biomolecular crystallography and NMR are two flourishing areas of research for protein structure determination drawing several talented scientists. Although more than half a million protein sequences are already known (*http://www.uniprot.org*) from diverse organisms, the massive world-wide structural (X-ray/NMR) efforts spread over the last 60 years have led to structures for only ~ 96,000 proteins (*http://www.rcsb.org/*) and a number of these are repetitions. The sequence to structure gap is continually diverging. Equally disconcerting is the fact that new folds emerging from X-ray and NMR have saturated at ~ 1400 since the year 2008. In view of the cost and time and other difficulties involved with these techniques, alternative accelerated methodologies are imperative for determining the three dimensional structures of biomolecules in general and proteins in particular to tackle the increasing burden of disease. Computational methods do offer this promise at least for soluble proteins in the near future.

Protein tertiary structure prediction field owes its success among others to three brilliant initiatives. The first is the PDB by Berman and coworkers [1]. The second is the BLAST by Altschul and coworkers [2]. The third is the CASP (*www.predictioncenter.org*) by Moult and coworkers [3]. Building on these foundations and harnessing advances in molecular modeling and simulation methods, the task now is to predict reliable three dimensional structures of proteins to within 2-3 Å root mean square deviation (RMSD) from the native ahead of experiment [4-6] in realistic time scales.

In CASP10, structures of 55% of the targets (proteins) were predicted accurately to within 5 Å RMSD (root mean square deviation) from the natives (crystal structures) by the best server (QUARK). Recently, *Bhageerath-H* [7-9] (*http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp*) has reached accuracies ≥ 60% for medium resolution (≤ 5 Å RMSD) structures. Further desired improvements include ensuring structure prediction under 5 Å RMSD of soluble monomeric proteins with accuracies closer to 100% and refining the predicted structures to bring down RMSDs to below 2 Å, both of which are non-trivial tasks. There have been sporadic reports in the recent literature of methods yielding > 80% Q3 accuracies in secondary structure prediction [10] which helps in improving tertiary structure prediction accuracies. However much remains to be done for predicting high resolution (< 2 Å RMSD from the native) structures with the current accuracies hovering around 15%. A less well appreciated problem, particularly in refinements, is the force field noise [11] which needs a quantum

Corresponding Author: **B. Jayaram**
E-mail: bjayaram@chemistry.iitd.ac.in; bjayaram@scfbio-iitd.res.in

mechanical resolution. The wish list further includes cutting down the computational time to less than an hour per protein, constructing quaternary structures where applicable, extending the methodology to membrane-bound proteins, predicting structures of all proteins of entire organisms in a high throughput mode etc.. If protein structure prediction accuracies improve, one can conceive of automating "Genome to Hit" molecule pathway [12], mapping protein-protein interactions at genomic scale [13] etc.. Further, with improved accuracies in protein tertiary structure prediction, one can envision creating a computational protein data bank (CoPDB) of modeled structures, which *inter alia* can be utilized for predicting off-target binding of lead compounds. To achieve the objectives, the homology and *de novo* methods will have to be augmented by additional structure generation methods such as chemical logic based alignments [14] and methods based on universalities in protein structures [15]. In matters of alignment and possibly in unraveling folding mechanisms, it pays to think differently about amino acids. Protein modelers are like doubting thomases. They are never sure of their model till the x-ray or NMR structure is released. The quality assessment methods cannot guarantee that the modeled structure is close to the native. A server which can estimate distance to the native in the absence of experimental structure (*http://www.scfbio-iitd.res.in/software/d2n.jsp*) fills this void. In the context of deciphering functional implications of SNPs, particularly those distal to functional/active site, some recent studies [16] suggest that it is not the structure of protein alone but its dynamics which fully explains its function and its alteration. The implications to allosteric site detection and activator design are obvious. Thus one envisages viable prediction methodologies emerging out of a confluence of new solutions to several small issues each of which can thwart a solution. For now, protein structure, dynamics and function prediction continues to be enchanting promising new therapeutics. Conceptual advancements beyond Pauling, Kauzman, Anfinsen and Ramachandran is a whole new exciting dimension all together to the student and researcher.

References

[1] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucl. Acids Res.,* **2000**, *28*, 235-42. *www.pdb.org*

[2] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucl. Acid Res.*, **1997**, *25*, 3389-3402.

[3] (a) J. Moult, K. Fidelis, A. Kryshtafovych, C. Venclovas, *Proteins Struct. Func. Bioinform.,* **2005**, *7*, 225-236. (b) J. Moult, K. Fidelis, A. Kryshtafovych, A. Tramontano, *Proteins,* **2011**, *79*(Suppl. 10), 1-5.

[4] K. A. Dill, J. A. MacCallum, *Science*, **2012**, *338,* 1042-1046.

[5] A. Tramontano, *Protein Structure Prediction: Concepts & Applications*, **2006**, Wiley-VCH, Weinham.

[6] D. E. Shaw, P. Maragakis, K. L. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W, Wriggers, *Science*, **2010**, *15,* 341-346.

[7] B. Jayaram, K. Bhushan, *et al.*, *Nucl. Acids Res.*, **2006**, *34,* 6195-6204.

[8] P. Dhingra and B. Jayaram, *J. Comput. Chem.*, **2013**, 15, 1925-1936.

[9] A. Mishra, S. Rao, A. Mittal, B. Jayaram, *BBA- Proteins Proteom*, **2013**, 8, 1520–1531.

[10] P. Ghanty P, N. R. Pal, R. K. Mudi, *J Bioinform Comput Biol*. 2013, Oct;11(5): 1350012. doi: 10.1142/S0219720013500121.

[11] P. L. Freddolino, C. B. Harrison, Y. Liu, K. Schulten, *Nature Phys.*, **2010**, *6*, 751-759.

[12] Anjali Soni, K. M. Pandey, P. Ray, B. Jayaram, *Curr. Pharm.Design*, **2013**, *19*, 4687-4700; DOI: 10.2174/1381612813199990379.

[13] Q. C. Zhang *et al.*, *Nature*, **2012**, *490*, 556–560. doi:10.1038/nature11503.

[14] B. Jayaram, *Nature Precedings*. *http://hdl.handle.net/10101/npre.2008.2135.1 2008.*

[15] A. Mittal, B. Jayaram, S. R. Shenoy, T. S. Bawa, *J. Biomol. Struc. Dyn.*, **2010**, *28 (2)*: 133-142.

[16] A. K. Padhi, B. Jayaram, J. Gomes, *Scientific Reports,* **2013**, *3*:1225, DOI: 10.1038/srep01225.