Session I: Structural Biology and Drug Discovery

Chairperson: Professor K. Murlidhar
Department of Zoology
University of Delhi
Delhi, India.

Rapporteurs : Dr. Seema Makhija

Dr. Ravi Toteja

The first session of National symposium Chemistry in Biology: The future of Life Sciences was Structural Biology and Drug Discovery. The session was chaired by Professor K. Muralidhar. Two speakers of the session were:

1. Dr. Amit Sharma, Group leader of Structural and Computational Biology Group in International Centre for Genetic Engineering and Biotechnology.

2. Professor B. Jayaram, professor in School of Biologial Supercomputing Facility for Bioinformatics and Computational Biology IIT, Delhi

The topic of Dr. Sharma talk was "Structural biology of malaria parasite proteins: insight and implications for drug discovery". He talked about the principles that govern biological function of key parasite proteins. For this one has to rely extensively on bioinformatics, biochemistry, cell biology, molecular biology, parasitology and protein crystallography. Using multi-disciplinary approaches, one can elucidate structure-function relationships for several crucial parasite proteins and these analyses will guide therapeutic developments against malaria. Earlier drugs like tetracycline are also effective against *Plasmodium* because it affects epicoplasr. Complex life cycle of the malaria parasite necessitates having an elaborate protein repertoire which is tightly regulated. This requirement highlights two important enzymatic families whose main function is in protein translation, its fidelity and proof reading – (1) aminoacyl-tRNA synthetases (protein translation) and (2) D-amino acid tRNA deacylase (editing or proof reading). There are two classes of aminoacyl t-RNA synthetase which differ in their 3-D structure, t-RNA recognition and ATP binding. During evolution the number of aminoacyl-tRNA synthetases increases and the no. in humans is 40. In *Plasmodium*, the greatest number of aminoacyl-tRNA synthetases is present in apicoplast which is algal in origin. Therefore, tetracycline target against apicoplast which is prokaryotic in origin.

In their lab crystalline structure of one of the aminoacyl-tRNA synthetases, tyrosyl t- RNA synthetase has been deduced. It is a dimer, cytoplasmic protein that has been co-crystallize with tyrosine and ATP. Tyrosyl-t- RNA synthetase in *Plasmodium* could be modulating the immune system of humans when they float in the blood. Differences in the enzyme families of *Plasmodium* and humans may provide target for drug designing.

The editing ability (i.e. correct amino acid get incorporated) is carried by D-amino acid tRNA deacylase. It is a conserved and ubiquitous protein which binds to D-amino acid and t-RNA. It has β1, β2, T1 and T2, A and E sites. T1 for D-amino acid entry, T2 for chirality check , A for ester bond cleavage and E is the exit site. β1 and β2 for adenosine docking and repositioning of ADP.

Structural studies on both these systems would help in rational drug design against these specific sites. The docking of inhibitors to these docking sites may help us to design site specific drugs.

The topic of Professor B. Jayaram talk was "Gene to drug *in silico*: A molecular bioinformatics endeavour." He talked about an ab initio model for gene prediction in prokaryotic genomes which is based on physicochemical characteristics of codons calculated from molecular dynamics simulations. The model requires a specification of three calculated quantities for each codon: the double-helical trinucleotide base pairing energy, the base pair stacking energy, and an index of the propensity of a codon for protein-nucleic acid interactions. The base pairing and stacking energies for each codon are obtained from recently reported MD simulations on all unique tetranucleotide steps, and the third parameter is assigned based on the conjugate rule previously proposed to account for the wobble hypothesis with respect to degeneracies in the genetic code. *Chemgenome* software have been developed based on DNA energetics and using this software one can find out whether the DNA sequence is gene or non gene.

Further he talked about combining bioinformatics tools and *ab initio* methodologies, for expeditious determination of for tertiary structure prediction of small proteins.

The software suite *Bhageerath* have been developed for this. The software comprises of eight modules configured to function independently Starting with amino acid sequence(primary structure) and secondary structure information (helix/sheet/loop) of a protein in the first module, multiple three dimensional atomic level structures are generated sampling the conformational space of the loop dihedrals in the second; in the third module a set of biophysical filters (persistence length, radius of gyration etc.) are applied which are designed to screen the trial structures to reduce the sample size. The resultant structures are refined in the fourth module by a Monte Carlo sampling in dihedral space to remove steric clashes / overlaps in 3-D space. An atomic level energy optimization is carried out in the fifth module and the structures scored based on energy in the sixth. Module seven reduces the probable candidates based on the protein regularity index of the φ (phi) and **ψ (psi)** dihedral values and module eight further reduces the structures selected to 10 using topological equivalence criterion and accessible surface area.

Thus millions of possible structures for a given protein sequence are brought down to 10 candidate structures with the possibility of bracketing the native in these 10. Results on a few small globular helical proteins have shown that native like folds in the root mean square deviation (RMSD) range of 3-6 A are captured in the best 10 structures energy-wise in all the cases without any exception. The 'needle in a hay stack problem' is thus reduced to choosing the best candidate from among the 10 lowest energy structures.

Comparison with the existing bioinformatics tools suggests the performance of the present methodology to be satisfactory and useful particularly when the database is deficient in sequence homologues. Preliminary work on alpha/beta systems has yielded encouraging results. Currently the expected prediction time with *Bhageerath* web server for systems with two secondary structures (with one loop in between) is ~4-5 min; while for systems with three secondary structures (with two loops in

between), it is ~2-3 h on a dedicated 32 processor cluster. Attempts to extend the methodology to larger systems with reduction in computational times are in progress.

Pursuing the dream that once the gene target is identified and validated drug discovery protocols could be automated using Bioinformatics & Computational Biology tools. At IIT Delhi, a computational protocol for active site directed drug design have been developed. The suite of programs (christened "***Sanjeevini***") has the potential to evaluate and /or generate lead-like molecules for any biological target. The various modules of this suite are designed to ensure reliability and generality. Making a drug is more like designing an adaptable key for a dynamic lock. The *Sanjeevini* methodology consists of design of a library of templates, generation of candidate inhibitors, screening candidates via drug-like filters, parameter derivation via quantum mechanical calculations for energy evaluations, Monte Carlo docking and binding affinity estimates based on post facto analyses of all atom molecular dynamics trajectories.