2012-13

Funded by Department of Biotechnology
Government of India

Center for Computational Biology and Bioinformatics

School of Computational and Integrative Sciences

Jawaharlal Nehru University

New Delhi

Progress Report of Center of Excellence

School of Computational and Integrative Sciences(SC&IS) presently consists of three centers, namely, Centre for Computational Biology and Bioinformatics (CCBB), High Performance Computing Center (HPCF) and Centre for Complex Systems Study. Our main objective is development of human resource through teaching and research in the frontier areas of Computational biology, Bioinformatics and Systems biology.

At present our strength is ten faculty members in our school and 8 faculty positions are advertised for recruitment in 2012-2013. One G.N.Ramachandran Fellow and one technical person are assisting to keep the teaching and facilities working. Dr. Jean-Numa Gillet from Belgium was appointed as a visiting faculty at our school from May 2012 to July 2013. His research interest is Theoretical Biophysics and Molecular Dynamic Simulations.

A novel characteristic of SC&IS is its emphasis on inter-disciplinary research like Computational Genomics, Development of Tools for Bioinformatics, Structure based Bioinformatics and *in silico* Drug Discovery, Chemoinformatics, Database Management, Microarray Analysis and Datamining, Systems biology and Evolutionary biology.

I. Research Projects :

Major research projects and field of research in CCBB presently, is focused towards the application of computational biology in infectious diseases like *Mycobactera, Leishmani, Amoeba, Malaria*, etc. and development of analytical tools to study biology.

A. *Mycobacterium tuberculosis:* (TB consortium project supported by DBT)

Genomic Comparison tools for Strain identification :

Mycobacterium tuberculosis is a major cause of morbidity and mortality throughout the world. Genomic variations in this organism have been used to type pathogenic strains in a limited scale [Journal of Bacteriology (2002) 184(19):5479-90]. There is not yet any attempt to generate a comprehensive database of all the genomic variations of *M. tuberculosis* though some attempts have been made in this direction. For example, MTBreg (http://www.doe-mbi.ucla.edu/Services/MTBreg/) covers variations that are detected using spoligotyping) and MycoDB (http://xbase.bham.ac.uk/mycodb/about.pl) has some features that allows comparison between two genomes in a limited manner. A comprehensive database has been generated containing genomic difference of different strain and species of *Mycobacteria* belonging to *M. tuberculosis* complex. The variations have been identified using ABWGAT (http://ccbb.jnu.ac.in/tools.html), a comparative genomic tool developed in the project (*see paper #5*)

B. Metabolic pathway analysis for validating drug targets in *M.Tb* :

As metabolism is a fundamental determinant of physiology, metabolic model can predict the phenotypic changes. There is one type of model that has shown a surprising ability to simulate the growth of living cells — namely stoichiometric Flux balance models. Therefore, the important steps in developing and disseminating the metabolic model are (i) Physiological and Phenotypic characterization of the bacteria; (ii) genome annotations; (iii)

metabolic reconstruction; (iv) Flux balance analysis. Use of Publicly accessible software tools has made it possible to identify and analyze the important metabolite with their relevant reactions & involved enzymes which could possibly are used as drug targets. Such applications have been conducted which resulted in identification of a set of proteins which are to be included as important drug targets. Last year M.Tech project entitled *"Studying the role of selective gene set in Mycobacterium tuberculosis metabolic pathway using flux balance analysis"* identified a set of genes for double knockout using new protocol of FBA .

C. Knowledge based compound design against *M.Tb*:

A novel method has been developed in-house to identify a set of chemical compounds with combinatorial motif by the study of known database, having minimum common bioactive substructure (MCBS). Application of data mining procedure to identify the 'relaxed' common pattern (pharmacophore) present in the diverse set of drug like/lead compounds is proposed. Use of graph theoretical approach to identify the Maximal Common Substructure (MCS) and Fuzzy logic techniques to identify the 'relaxed' chemical motifs so that broader pattern can be found is proposed.

D. Development of integrated database of drug targets and antimalarial compounds of *Plasmodium falciparum* (http://pfal.scisjnu.ernet.in).

This database contains most of the information about possible drug targets of malarial parasite. Information regarding structure of protein, their function, domain, sequences, expression stages etc is available in this database. Uniqueness of this database is the active information and homology modeled structures of malaria targets whose structures has not been solved by X-ray techniques.. It also contains detailed description of chemical compounds showing inhibitory assay against Plasmodium falciparum organism or their known drug targets. This project was supported by Ministry of Communication and Information Technology (MCIT).

E. Identification of new drug targets in *Leishmenia*

Leishmania major, a protozoan parasite, is the causative agent of cutaneous leishmaniasis. Due to the development of resistance against the currently available anti-leishmanial drugs, there is a growing need for specific inhibitors and novel drug targets. In this regards, aminoacyl tRNA synthetases, the linchpins of protein synthesis, have received recent attention among the kinetoplastid research community. This is the first comprehensive survey of the aminoacyl tRNA synthetases, their paralogs and other associated proteins from L. major. A total of 26 aminoacyl tRNA synthetases were identified using various computational and bioinformatics tools. Phylogenetic analysis and domain architectures of the L. major aminoacyl tRNA synthetases suggest a probable archeal/eukaryotic origin. Presence of additional domains or N- or C-terminal extensions in 11 aminoacyl tRNA synthetases from L .major suggests possibilities such as additional tRNA binding or oligomerization or editing activity. Four freestanding editing domains were identified in L. major. Domain assignment revealed a novel asparagine tRNA Synthetase paralog, asparagine synthetase A which has been so far reported from prokaryotes and archea.(*see paper #19)*

F. Computational study of Tumor Suppressor protein

Studying p53-MDM2 interactions and also exploring the possibility of identifying suitable drug targets to break this interaction. Well compiled drug databases and screening algorithms are used to select suitable targets and their energetic are also studied in depth. We are also interested in looking at p53-DNA interactions through response elements and we are making efforts to understand the role of p63 and p73 in p53 related pathway.

G. A neurocomputational approach to automaticity in motor skill learning

The first is in the area of multiple robot coalition formation, to find novel algorithms for multiple robot coordination/cooperation. This is being done in collaboration with colleagues at Delhi University. The second is in the area of cognitive modeling for robotic applications. In collaboration with faculty from Vanderbilt University, a model had been developed for mimicking the automaticity that humans exhibit during learning. The groups at JNU pursuing models that better explain various phenomena associated with acquisition and extinction.

II. New Tools Development

A. Computational study of replication origins:

DNA replication is a complex process and crucial for the life cycle of any organism. Protein-DNA interactions play a major role and understanding this complex process is a tremendous challenge to biologists and computational biologists. Modes of replication are not same in prokaryotes and in eukaryotes. We are making an effort to study the origin of replication sequences, its distribution, its prediction and functioning using computational approaches. A method to predict OriC sites in bacteria using a correlation measures and also explored its applicability to higher organisms has been published recently in Biosystems(*see paper #17).*

B. Information theoretic analysis:

For analysis purposes DNA is viewed as a long one dimensional symbolic sequences and the information embedded is important for understanding the biological functions. Information theoretic measures such as Shannon entropy, Renyi entropy and Tsallis entropy are increasingly used to identify variable segments in the DNA sequences. In addition, divergence measures are also applied to DNA for getting evolutionary insights. We employ these IT measures and other techniques to individual genes and genomes sequences as well as multiple aligned sequences to identify biologically interesting motifs *(see paper #18).*

Current Academic Activities(2012-2013).

See Website for more details http://ccbb.jnu.ac.in

- 82 Post Graduate Diploma M.Phil (equivalent)in Bioinformatics during 2001-2006
- During 2006-2012 Forty five M.Tech students passed.
- Fourteen research fellows have been awarded PhD in Computational Biology & Bioinformatics.
- Ten students have done their 6 month projects as long term trainees and five short

term trainees including summer trainees at our school during 2012.

Training & Visitors: http://ccbb.jnu.ac.in/training.html
Seven visitors had given lectures at our school during 2012. http://ccbb.jnu.ac.in/events.html

Supported Fellows & Associates: http://ccbb.jnu.ac.in/fellows_associates.html

- Mr Surajit Mishra was supported to develop CTEP software in collaboration with DBT.
- Mr Adarsh Tayal is working COE project as System Analyst
- Dr V. Gowri as Research Associate 2011-2012
- Kushal Shaw as GN Ramachandran Fellow 2012
- Dr Rashi Gupta as Teaching Associate for course teaching (2012 Winter)
- Dr V.K.Jayram for lectures in Machine Learning (during Feb-March 2012)

BINC Examination :

SC&IS, JNU successfully conducted BINC Exam 2011-2012 during February 26th and 27th 2012. Out of 800 appeared, 38 were awarded BINC fellowships and top ten students were given Cash awards of Rs 10000 each. Bioinformatics National certificate award ceremony was organized in May 2012. DBT secretary Prof M.K.Bhan attended as Chief Guest.

BINC 2013 examination will be held from **Feb 23th -24th 2013.** Now expanded to nine centers namely, Jawaharlal Nehru University, New Delhi; Pune University, Pune; Anna University, Chennai; University of Kerla, Thiruvanthapuram; Calcutta University, Kolkata; Institute of Bioinformatics & Applied Biotechnology, Bangalore; University of Hyderabad, Hyderabad. North Eastern Hill University, Shillong and Assam Agricultural University, Gauhati

BINC-2012 DBT was supported by ERNET connectivity as the major host of online system of registration and communication. In addition to that PfalDB, OpenTox and other online servers are hosted via ERNET connectivity.

Please visit :

http://binc.scisjnu.ernet.in
http://pfal.scisjnu.ernet.in/Malaria/HomeinLink
http://www.opentox.org/

Current Research Projects http://ccbb.jnu.ac.in/research.html

1. MCIT project on “Designing novel antimalarials using target based pharmacophore approach” Ministry of Information Technology, Govt of India, Aug 2009-2012 Extended upto May 2013. PI: Professor Indira Ghosh and Co-PI : N.Subba Rao.

2. DST project on “Development and application of a Biased Monte Carlo simulation

technique: Exploring and characterizing potential energy surfaces of large molecules". 2010-2015.PI: Pradipta Bandhyopadhyay

3. Bioinformatics National Certificate Exam 2011-2013 PI: Prof Indira Ghosh

Projects in collaboration with academia

4. DBT Project on "Establishment of National Database on Tuberculosis(TB) Phase II TB Consortium": Prof. Alok Bhattacharya and Prof Indira Ghosh 2011-2013
5. DBT project on "Computational Core for Plant Metabolomics" Prof Indira Ghosh with IIIT, Hyderabad, 2011-2015
6. DBT Builder Project, Coordinator: Prof R.Bhat, SBT, JNU (SLS, SCMM and SBT and SCIS faculty of JNU) 2012-2017

Conference and Workshop during 2012 http://ccbb.jnu.ac.in/events.html

1. Biomolecules in Motion: Theory and Simulations January 4-6 2013, 16 speakers and Participants:55
2. One Day Meeting of Synthetic Biology, December 21st 2012 : participated by DBT and NIH representative along with 8 speakers and 25 participants.
3. Latest Advances in computational Drug Discovery 29th October 2012 Participants:35
4. National Conference-cum-workshop on "Search for Antimalarials: Mechanism based Approach" 27-29th April 2012. 20 speakers & Participants: 40
4. Annual Open Day on 21st January 2012

Publications 2012 (20 +2) ***http://ccbb.jnu.ac.in/publications.html***

1. Kamaldeep Gill, Abhay K. Singh, Vaishali Kapoor, **Lokesh Nigam**, Rahul Kumar, Prasida Holla, Satya N. Das, Savita Yadav, **Naidu Subbarao**, Bidhu Kalyan Mohanti, Sharmistha Dey, Development of peptide inhibitor as a therapeutic agent against head and neck squamous cell carcinoma (HNSCC) targeting p38α MAP kinase, Accepted for Publication in Biochimica et Biophysica Acta(2012)
2. Ashish Gupta, **Lovekesh Vig** and David C. Noelle "A Cognitive Model for Automaticity in Motor Skill Learning", Journal of Robotic and Intelligent Systems, 2012 (In Press)
3. Shikha Pal, Madhur Mishra, D. **Raja Sudhakar,** Mohammed Haris Siddiqui, In silico designing of a potent analogue against HIV-1 Nef protein and protease by predicting its interaction network with host cell proteins, Journal of Pharmacy and Bioallied Sciences January-March 2013 Vol 5 Issue 1
4. **Singh P and Sengupta S,** Phylogenetic analysis and comparative genomics of purine riboswitch distribution in prokaryotes.Evol Bioinform Online. 2012;8:589-609. doi: 10.4137/EBO.S10048. Epub 2012 Nov 6.
5. **Sarbashis Das**, Priyanka Duggal, Rahul Roy, Vithal P. Myneedu, Digamber Behera, Hanumanthappa K. Prasad & **Alok Bhattacharya** Identification of Hot and Cold spots in genome of Mycobacterium tuberculosis using Shewhart Control Charts Nature Scientific Reports - http://dx.doi.org/10.1038/srep00297
6. Chung-Chau Hon, Christian Weber. Odile Sismeiro, Caroline Proux, Mikael Koutero, Marc Deloger, **Sarbais Das**, Mridula Agrahari, Marie-Agnes Dillies, Bernd Jagla,Jean-Yves-Coppee,

Alok Bhattacharya and Nancy Guillen Quantification of stochastic noise of splicing and polyadenylation in Entamoeba histolytic, Nucleic Acid Research (2012) 1-17, doi: 10.1093/nar/gks1271

7. Jonathan P. Furado, Anuja P Rahalkar, **Sudhanshu Shankar, Pradipa Bandhyopadhyay** and Sridhar R Godre, Facilitating minima search for large water clusters at MP2 level via Molecular tailoring, J.Phy.Chem. Letters(2012)
8. Biswaranjan Meher, Mattaparthi Venkata Satish Kumar, **Smriti Sharma** and **Pradipta Bandhyopadhyay**, Conformational Dynamics of HIV-1 protease: a comparative dynamic simulation study with multiple amber force fields, Journal of Bioinformatics and computational biology(2012)
9. **Sudhanshu Shankar** and **Pradipta Bandhyopadhyay** Determination of low energy structures of a small RNA hairpin using Monte Carlo based techniques, J.Biosciences 37(3) 533-538
10. Uddipan Sarma, Archana Sareen, Moitrayee Maiti, Vanita Kamat, Raki Sudan, Sushmita Pahari,Neetu Srivastava, Somenath Roy, Sitabhra Sinha, **Indira Ghosh**, Ajit G. Chande, Robin Mukhopadhyaya, Bhaskar Saha, "Modeling and Experimental Analyses Reveals Signaling Plasticity in a Bi-Modular Assembly of CD40 Receptor Activated Kinases" PLoS ONE (2012),7: July 2012 e39898 doi:10.1371/journal.pone.0039898
11. Uddipan Sarma & **Indira Ghosh**, "Oscillations in MAPK cascade triggered by two distinct designs of coupled positive and negative feedback loops" BMC Research Notes. (2012) 5:287 DOI:10.1186/1756-0500-5-2872012.
12. Uddipan Sarma & **Indira Ghosh** , "Different designs of kinase-phosphatase interactions and phosphatase sequestration shapes the robustness and signal flow in the MAPK cascade" BMC Systems Biology. (2012) 6:82 DOI:10.1186/1752-0509-6-82.
13. **Rithvik Vinekar**, Chandra Verma, **Indira Ghosh** , "Functional relevance of dynamic properties of dimeric NADP-dependent isocitrate dehydrogenases" BMC Bioinformatics (2012) Volume 13 Supplement 17, S 2.
14. Mobashar Hussain Urf Turabe Fazil, Sunil Kumar, **Naidu Subba Rao** , Chandrabose Selvaraj, Sanjeev Kumar Singh, Haushila Prasad Pandey & Durg Vijai Singh, "Comparative structural analysis of two proteins belonging to quorum sensing system in Vibrio cholerae , Journal of Biomolecular Structure and Dynamics", DOI:10.1080/07391102.2012.687523
15. **Smriti Sharma and Pradipta Bandyopadhyay,** Investigation of the acylation mechanism of class C beta-lactamase: pKa calculation, molecular dynamics simulation and quantum mechanical calculation, Journal of molecular modeling(2012), v18, 481-492
16. **Rupesh Kumar**, Amitabha Bose, Birendra Nath Mallick A Mathematical Model towards Understanding the Mechanism of Neuronal Regulation of Wake-NREMS-REMS States, PLOSOne(2012), 7(8), : e42059. doi:10.1371/journal.pone.0042059
17. **Kushal Shah** and **A. Krishnamachari.** "Nucleotide correlation based measure for identifying origin of replication in genomic sequences", BioSystems vol. 107 (1), 52-55,2012
18. **Kushal Shah** and **A. Krishnamachari.** "On the origin of three base periodicity in genomes", BioSystems, vol.103(3) 142-144,2012.
19. Rentala Madhubala, **V S Gowri, Indira Ghosh,** Amit Sharma. “Unusual domain architecture of aminoacyl tRNA synthetases and their paralogs from Leishmania major” in BMC Genomics (2013).
20. **Jean-Numa Gillet and Indira Ghosh.** “Concepts on the protein folding problem” in Journal of Biomolecular Structure and Dynamics, in press (2013).

Book Chapters:

Kushwaha Hemant Ritturaj and **Ghosh Indira** , “Bioinformatics Approach for Finding Target

Protein in Infectious Disease" In Bioinformatics of Human Proteomics,Translational Bioinformatics, DOI 10.1007/978-94-007-5811-7_10, X. Wang (ed.), © Springer Science+Business Media Dordrecht 2013.

Kushwaha Hemant Ritturaj and Ghosh Indira. "Target Identification using Systems Approach in Infectious Diseases" . In TRANSLATIONAL RESEARCH IN NEW DRUG DEVELOPMENT. Ed. Ray Arunabha & Gulati Kavita (V.P, India), 2012.

Future Plans

- Relocation of School of Computational and Integrative Science to New Building by end of 2013
- One Professors, five Associate Professors and two Assistant Professors are to be recruited under UGC IX Plan during 2013. Two GN Ramachandran fellowships funded by COE will be appointed very soon
- Building up of the research activities in the Center for Complex Systems Studies (CCSS).
- Proposal for Two year M.Sc course(computational and Integrative sciences)
- Research focus on Infectious Diseases and understanding of large systems in biology.

Contact Address

Prof Indira Ghosh

Coorindator, Centre of Excellence in Bioinformatics

Centre for Computational Biology and Bioinformatics

School of Computational and Integrative Sciences

Jawaharlal Nehru University, New Delhi-110067

E-mail: indirag@mail.jnu.ac.in /ighdna@yahoo.co.in

Ph: 011-26738707

Appendix

List of students who have completed their M.Tech and Ph.D successfully & Degree submitted/awarded during 2012(M.Tech and Ph.D. Thesis Titles) Ph.D. Students
http://ccbb.jnu.ac.in/thesis-p-2011-2012.html

M.Tech. Thesis Titles (2010-2012)

Sr.No.	Name of the Students	Thesis Title	Supervisor
1	Sanjeev Kumar	Computational Study of Origin of Replication and ARS Sequences in Few Genomes.	Dr. A Krishnamachari
2	Sahil Moza	Stochastic Modeling of Calcium Initiated Phagocytosis in Entamoeba Histolytica.	Prof. Ram Ramaswamy and Prof. Alok Bhattacharya
3	Tanmaya Meher	Data generation and Data analysis of virtual high throughput screening (vHTS) in Plasmodium falciparum proteins	Prof. Indira Ghosh
4	Neha Aggarwal	Early (Pre-Luca) Evolution of Genetic Code: A finite Population Study	Dr. Supratim Sengupta
5	Sonia Verma	Analysis of Nucleotide Binding Domain Structure and Phylogeny of Fungal Tranaporters.	Dr. Andrew M. Lynn
6	Gulrez Chahal	Structural Identification & comparison of Protein Binding sides using graph Theoretical Methods. (Clique Detection)	Naidu Subbarao
7	Riti Roy	Analysis of transcriptome and retrotransposable elements LINEx/SINEx of Entamoeba histolytica	Prof. Alok Bhattacharya
8	Ashish Kumar Singh	Studying the role of selective gene set in Mycobacterium Tubercuiosis metabolic pathway using flux balance analysis.	Prof. Indira Ghosh
9	Md. Sultan Alam	Analysis of sequence polymorphism of LINEs and SINEx in Entamoeba Histolytica	Prof. Alok Bhattacharya & Prof. Ram Ramaswamy
10	Navneet Chandra Verma	DENSITY OF STATES SIMULATIONS FOR PROTEIN	Dr. Pradipta Bandyopadhyay

		SYSTEM STUDY OF WW-DOMAIN.	
Ph.D – Submitted/awarded during 2012			
11.	Sarbashis Das	Compatative Analysis and Study of Mutations in Bacterial Genomes.	Prof. Alok Bhattacharya
12.	Md.Rehan	Information Theoretic Methods for Identification of Functional Residues in Proteins.	Dr. Andrew M.Lynn