# Kolkata International School cum Conference on Systems Biology

# KOLSYSBIO

**December 29'th 2012 to January 3'rd 2013**

Venue: Auditorium, Saha Institute of Nuclear Physics, Kolkata

# Abstract Book


Saha Institute of Nuclear Physics
Kolkata, INDIA

## Abstracts of School Talks

# KOLKATA INTERNATIONAL SCHOOL CUM CONFERENCE ON SYSTEMS BIOLOGY (KOLSYSBIO)

**December 29, 2012 to December 31, 2012**

## CONVENORS:

**Pradeep K Mohanty, SINP**
**Soumen Roy, Bose Institute**

# S01: Introduction to molecular biology: systems (modularity) all the way down

*James Shapiro*
*The University of Chicago, USA*

The lecture will introduce the students to some classical examples of molecular cell biology. The emphasis will be on how the cell functions as a system through molecular interactions. Another main point is that each of the interacting molecules itself functions as a multi-component system.

# S02/S04: Structural bioinformatics: applications of tools and principles

*Prasanna Venkatraman*
*Advanced Centre for Treatment, Research and Education in Cancer, Mumbai*

The tutorial will include an introduction to structural bioinformatics from a biologist's persepctive, importance of protein-protein interactions, rules that govern these interactions, their importance in biological function.

Then there will be a discussion on bioinformatic studies that dwell on these protein-protein interactions.

Subsequently our own thoughts and the use of these prinicples in our research which will be illustrated using various examples.

# S03: Introduction to quantitative analysis for biologists

*Srinivasan Ramachandran*
*Functional Genomics Unit,*
*CSIR-Institute of Genomics and Integrative Biology, Delhi*

These times are the age of big data. The availability of various types of big data is changing the approach to analyze and develop new strategies for drug targets, drugs, vaccine candidates to combat infectious diseases. We present here some approaches to analyze examples of big data: screening of chemical molecules for identifying potential inhibitors, high scoring immune epitopes, single nucleotide variations between strains of pathogen.

**ParallelVSR: a cheminformatics pipeline for parallel virtual screening on R platform**

We are recently receiving access to small molecule databases in the public domain. The chemical search space has therefore widened. New strategies need to be developed for efficient virtual screening (VS). In recent times we also are gifted to receive many screening algorithms these includes Vina, ChemmineR, etc. Computer speed has now become the technical issue and this requirement needs to be addressed without compromising accuracy. Such requirements are now being met through the development of freely available, fast and easy to use virtual screening pipeline, which is based on integrated approach, has potential for use in the field of drug discovery.

We describe here ParallelVSR a cheminformatics pipeline in R platform. We have made efforts to in integrate sequentially the ligand based virtual screening (LBVS) and structure based virtual screening (SBVS). LBVS used when the three dimensional structure of target protein is not available but one or several of its inhibitors are known and in that case structurally similar molecules to the known ones are sought from a chemical library of large size assuming that these structurally similar molecules may have similar activity. SBVS requires the knowledge of three-dimensional structure of the target protein. In this method, a set of small molecules are docked into the protein, and the stability of the protein-ligand complex and the details of ligand interaction with the active site of the target protein are the measures of activity of the docked ligand.

Starting with a small molecule as an input query the pipeline system performs fast LBVS using ChemmineR. Through this we can reduce a large chemical

library of small molecules to a manageable size and then followed through relatively slow SBVS using AutoDock Vina to rank the screened molecules on the basis of docking energy. We used data level parallelization for both LBVS and SBVS using Message Passing Interface (MPI) as communication protocol. We developed ParallelVSR comes with easy to run R *scripts*, which can run on MPI enabled system. This development offers added advantage of simplicity in using this pipeline, Parallel VSR is designed to meet the challenge of speed in VS.

**Immunological Data Modeling, Scripting and Analysis of Pathogen Genomes**

Genome sequence data from various pathogenic species and strains open new opportunities for development of new therapeutics and vaccine candidates through Bioinformatics analysis. Analysis of data is now possible using structured scripting. R is a programming language integrated with an R environment, facilitating easy and rapid data analysis with the help of its integrated suite of software facilities.

In the genomics era, vaccine targets prediction starts by using bioinformatics analysis of microbial genome sequences. This approach is called Reverse Vaccinology (RV). As newer bioinformatics algorithms develop and appear in public domain, the original approach can now be complemented by enhanced RV, which uses additional algorithms for prioritizing the vaccine candidates. We start with collection of known vaccine candidates and a set of predicted vaccine candidates from whole genome sequences of the selected pathogen genomes. These predicted vaccine candidates are adhesins and adhesin-like proteins identified by using various algorithms like MAAP for plasmodium species, FungalRV adhesin predictor for human pathogenic fungi and SPAAN for bacterial species. These sequences are analyzed through publicly available algorithms to obtain additional information on Ortholog, Paralog, BetaWrap Motifs, Transmembrane Domains, Signal Peptides, Conserved Domains, Similarity to human proteins from Human Reference Proteins, T-cell epitopes, B-cell epitopes, Discotopes and Allergens prediction. This combined approach forms the basis for enhanced Reverse Vaccinology. The combined data produced through these integrated analyses can be interogatted through well structured decision trees. We prepare scripts running in object oriented mode. From these results we derive a set of most probable adhesin vaccine candidates with additional qualification. Furthermore, the degree of conservation of the epitopes can also be investigated. These results could enable the development of epitope based vaccines in future.

# S05: Case examples of cheminformatics, integrated immunoinformatics and single nucleotide variations

*Srinivasan Ramachandran*
*Functional Genomics Unit,*
*Institute of Genomics and Integrative Biology, Delhi*

**Analyzing single nucleotide variation in Pathogenic Bacteria**
Single nucleotide variations (SNVs) are variations in the nucleotide at single positions. Single nucleotide variations in the genome can have varying consequences depending on its location in the genme. We developed a pipeline of scripts integrating multiple algorithms through R platform. Here we show one case example for multiple strains of *Mycobacterium tuberculosis* of CAS spoligotype from India. In the present analysis, genome sequences of three strains of *M. tuberculosis*: AHHX, AHHY and AHHZ were analyzed for the single nucleotide variations with reference to the H37Rv, a pathogenic laboratory strain. Total SNVs were 1310, 1286 and 1131 in AHHX, AHHY and AHHZ respectively with reference to H37Rv. These SNVs covered 0.030%, 0.029% and 0.026% of the reference genome, in AHHX, AHHY and AHHZ respectively. Since the protein coding percentage of H37Rv is 91.3%, the SNVs were found more in the coding region than in intergenic region. Transition to transversion ratios (Ts/Tv ratio) were also calculated in each strain and analysis of SNVs supports the transitional bias in all the strains. Further study on the synonymous and non-synonymous SNVs in different functional categories of genes as per classification given by TubercuList suggests that 10-11% in cell wall and cell processes, 7-8% in conserved hypotheticals, 10-11% in information pathways, 7-9% insertion seqs and phages, 9-10% in intermediary metabolism and respiration, 13-15% in lipid metabolism, 13-17% in PE/PPE, 9-10% in regulatory proteins,7-8% in virulence, detoxification, adaptation, 6-7% in unknown and 2% of the total SNVs were not categorized. The length truncation of open reading frames due to SNVs was also detected in the proteins produced by the corresponding coding transcript. The number of SNVs causing decrease in length was 10 in each of the three strains whereas SNVs causing increase in the length of AHHX, AHHY and AHHZ were found to be 8, 8 and 7 respectively. Our findings show that majority of truncation in length occur due to substitution of G/C–>A/T whereas increase in length occur due to substitution of A/T–>G/C. Since the GC content of H37Rv is 65.9%, these single nucleotide variations might have role in increase or decrease of the GC content in the transcripts of the Indian strains of *M. tuberculosis*.

# S07/S11: Synthetic genetic engineering

*Calin Guet*
*Institute of Science and Technology, Vienna, Austria*

Synthetic biology engineering techniques are often used in systems biology experimental approaches. I will discuss the potential and constraints of such engineering approaches. Emphasis will be placed on the power that synthetic biology has to address fundamental questions of living matter and its limitations in engineering biological machines.

# S08: Modelling intrinsic and extrinsic noise in biochemical networks

*Vahid Shahrezaei*
*Department of Mathematics, Imperial College, London*

Gene expression is significantly stochastic making modeling of genetic and biochemical networks challenging. This stochasticity arises because of both inherent stochasticity in biochemistry (intrinsic fluctuations), as well as interactions of the system of interest with other stochastic systems in the cell or its environment (extrinsic fluctuations). In this talk I describe different approaches used to model such stochasticity.

# S09: DNA-based genetics teaches us how cells control genome change

*James Shapiro*
*The University of Chicago, USA*

The lecture will introduce the students to the molecular basis of mutation and genome change. The main point will be that ALL genome changes result from dedicated biochemical activity. These activities are subject to cell regulatory circuits. Thus, the cell has the ability to change its genome when necessary.

# S12: Rule-based modelling in systems biology

*Vahid Shahrezaei*
*Department of Mathematics, Imperial College, London*

Much of the complexity of biochemical networks comes from the information-processing abilities of protein modifications and, and binding to the receptors, ion-channels, signalling molecules or transcription factors. The large number of modifications and binding events produces a combinatorial increase in the size of such reaction networks. Rule-based modelling is a framework that is devised to deal with the combinatorial complexity. An additional problem is combinatorial increase in the number of parameters required to fit experimental data as the number of protein interactions increases. It therefore challenges the creation, updating, and re-use of biochemical models. I discuss a rule-based modelling framework that exploits the intrinsic modularity of protein structure to address regulatory complexity. This methodology provides a basis for scalable, modular and executable modelling of biochemical networks in systems and synthetic biology.

## Abstracts of Conference Talks

# KOLKATA INTERNATIONAL SCHOOL CUM CONFERENCE ON SYSTEMS BIOLOGY (KOLSYSBIO)

**January 1, 2013 to January 3, 2013**

## CONVENORS:

**Pradeep K Mohanty, SINP**
**Soumen Roy, Bose Institute**

# C01: What DNA teaches about evolution

*James Shapiro*
*The University of Chicago, USA*

The lecture will make nine basic points about genome change in evolution:
1.Evolution is complex, not reducible to simple formulae.
2.Evolutionary thinking has a long history full of ongoing discoveries.
3.Cell mergers are an important source of abrupt evolutionary novelty.
4.Horizontal DNA transfer is an important source of rapid evolutionary novelty.
5.Cells actively repair and restructure their genomes (the genome as a RW memory system).
6.Proteins evolve by swapping segments, not simply by changing one amino acid at a time.
7.Mobile genetic elements can rapidly modify genome function at multiple locations and establish genomic networks.
8.Inter-specific hybridization and whole genome duplications are further sources of rapid evolutionary innovation.
9.Genome restructuring (natural genetic engineering) is regulated and activated by stress, cell fusions and inter-specific hybridization.

# C02: Application of structural bioinformatics in the identification of potential therapeutically vulnerable targets and their experimental validation

*Prasanna Venkatraman*
*Advanced Centre for Treatment, Research and Education in Cancer Mumbai*

High throughput genomic and proteomic studies grow at a rapid rate. But decoding information arising out of such studies to discover novel functions, build and understand communication networks and more importantly recognizing therapeutic targets remains a major challenge. One such communication network is created by protein-protein interactions. Since these interactions are rewired and new networks established in aberrant conditions like cancer they become attractive drug candidates. However in order to identify such putative targets, studies aimed at recognizing differential networks need to be coupled with residue level details of interactions. With this aim, we explore the power of structural bioinformatics to identify physiologically relevant direct interacting partners of gankyrin, an oncoprotein over expressed in several epithelial cancers. We validate our prediction by demonstrating endogenous interaction, interactions using purified proteins and confirm that the site of interaction is through predicted residues by site directed mutagenesis. We identify Hot Spot Sites in the gankyrin network and uncover interactions specific and resident in cancer cells that could be potential therapeutic targets!

# C03: Identification of novel adhesins of *M. tuberculosis* H37Rv using integrated approach of multiple algorithms and experimental analysis.

*Srinivasan Ramachandran*
*Functional Genomics Unit,*
*Institute of Genomics and Integrative Biology, Delhi*

Bacterial cell wall biogenesis and its metabolism play crucial role in the growth of bacteria. Pathogenic bacteria interacting with eukaryotic host express adhesins on their surface. These adhesins help in bacterial attachment to the host cell receptors and aids in colonization. Characterization of *M. tuberculosis* attachment with respiratory mucosa led to identification of various adhesins, such as surface exposed heparin-binding hemagglutinin adhesin (HBHA), which is required for bacterial attachment on lung and extrapulmonary dissemination of bacterium and laminin binding protein (LBP) that is involved in cytoadherence through laminin recognition. The complete genome sequences of various species of the genus mycobacteria offer a great opportunity for detailed biological characterization of various genes and their proteins, thereby saving lot of time and resources required for experimental screening. Further, experimental screening involves precise model for setting up receptor binding interactions. This approach also results in identification of limited numbers of adhesins because of the technical setup. In the present work, computational screening of all adhesins in *M. tuberculosis* based on sequence features has been carried out on *M. tuberculosis*. We have used an integrated approach of multiple algorithms to boost the computational power and also carried out experimental testing. Initial computational screening of the whole proteome of *M. tuberculosis* using SPAAN lead to the identification of proteins with high probability of being adhesin or adhesin-like protein. Localization prediction of these proteins were predicted using various subcellular localization prediction algorithms, namely, LOCTree, PSORTb and SubLoc. Further, proteins with low molecular weight were given preferences for the ease of cloning, expression and purification of the proteins. Because our goal was to identify novel adhesins of *M. tuberculosis*, we excluded the proteins which were reported in literature. These selected proteins were cloned, expressed and purified and tested for their ability to bind to the extracellular matrix proteins using a modified ELISA method. Here we report Rv0309, Rv2599 and Rv3717 as novel potential adhesins of *M. tuberculosis* H37Rv. With the analysis of their ability to bind to proteins of extracellular ma-

trix we observed that Rv0309 binds to fibronectin, Rv2599 binds to fibronectin, laminin and collagen, whereas Rv3717 binds to both fibronectin and laminin. Our results expand the number of known adhesins of *M. tuberculosis*. This expanded information will be useful for vaccine candidate predictions.

# C04: Input-output relations in biochemical networks

*Vahid Shahrezaei*
*Department of Mathematics, Imperial College, London*

Biological cells process information and make reliable decisions via biochemical signalling networks. The input-output relations in these networks are evolved to produce reliable function in spite of stochasticity present in their dynamics. Here, I present some results on mathematical modelling of a common class of biochemical signalling motifs, the phosphorylation depohosphorylation cycle. I discuss the role of enzyme saturation, multiple binding sites, diffusion-limited reactions and enzymatic complex formation. I illustrate how these networks can produce linear response, ultrasensitive response or non-monotonic response. I will also discuss the connection to yeast mating and T cell receptor signalling.

# C05: Cytoskeleton dynamics, traction forces and mechanical responses investigating the interface of biochemistry and mechanics in neurons.

*Aurnab Ghose*
*Indian Institute of Science Education and Research, Pune*

Accurate growth cone-mediated axonal pathfinding is the underlying mechanism responsible for establishing the stereotyped neuronal circuitry. Growth cones are specialized tips of neurites that sense guidance cues and mediate directional translocation towards synaptic targets. Both the sensory and the motile activities of the growth cone depends on the generation of protrusive processes viz., filopodia and lamellipodia. We have investigated the role of Fmn2, a non-canonical actin nucleator, in the regulating axon guidance. Fmn2 is enriched in the developing nervous system and knockdown of Fmn2 results in guidance defects in vivo. Mechanistically, Fmn2 affects the translocation of growth cones by affecting the ability to generate traction forces. Investigations into the spatiotemporal regulation of traction forces reveal that growth cone appear to translocate using a central compression type of growth.

# C06: Mitochondrial variability

*Nick Jones*
*Department of Mathematics, Imperial College, London*

A view of mitochondria as a set of static, isolated, genetically homogeneous organelles is now markedly inconsistent with data. They are a networked, fluctuating, ensemble under continuous control. I will discuss our work investigating the sources of mitochondrial variability and touch on implications for cell-to-cell variability, global gene expression, cell cycle duration and development.

# C08: Systems biology approaches to drug discovery

*Nagasuma Chandra*
*Indian Institute of Science, Bangalore*

Systems biology, an emerging discipline seeks to study biochemical and biological systems from a holistic perspective, contrary to the reductionist approach that has dominated biology. Systems biology requires close co-ordination between high throughput experiment, mathematical abstractions and computational analysis. An important outcome promised by the systems level studies is an understanding of the physiology in normal health, as an integrated function of several individual components, interactions among them and their regulation. In particular, an ultimate goal is to study the effect of naturally occurring perturbations that lead to disease and explore possible ways of reversing such pathological effects through therapeutic intervention.

The science of drug discovery itself has been witnessing multiple paradigm shifts in the recent past due to several factors such as availability of genome sequences, significant growth of sequence, structure and function level databases. The omics scale experiments, development of system-level models as well as adaptation and application of computational methods to biological problems provide a further new direction to drug discovery research. Systems level understanding has the potential to address several important issues that arise in drug discovery, such as the choice of an optimal target, causes for failure of existing drugs including drug resistance, adverse effects and causes of drug toxicity. This talk will provide an overview of the principles and practices of systems biology and how it will impact all three major branches of medicine- diagnosis, treatment and prevention.

# C09: Chemical chaperones assist intracellular folding to buffer mutational variations

*Kausik Chakraborty*
*Institute of Genomics and Integrative Biology, New Delhi*

Hidden genetic variations have the potential to lead to the evolution of new traits. Molecular chaperones, which assist protein folding, may conceal genetic variations in protein-coding regions. Here we investigate whether the chemical milieu of cells has the potential to alleviate intracellular protein folding, a possibility that could implicate osmolytes in concealing genetic variations. We found that the model osmolyte trimethylamine *N*-oxide (TMAO) can buffer mutations that impose kinetic traps in the folding pathways of two model proteins. Using this information, we rationally designed TMAO-dependent mutants *in vivo*, starting from a TMAO-independent protein. We show that different osmolytes buffer a unique spectrum of mutations. Consequently, the chemical milieu of cells may alter the folding pathways of unique mutant variants in polymorphic populations and lead to unanticipated spectra of genetic buffering

# C10: Re-mining data to predict the future: our experience with breast cancer

*Kartiki V. Desai*
*National Institute of Biomedical Genomics, Kolkata*

Three important clinical care challenges plague breast cancer therapy, non-respons- iveness, increasing drug resistance and the presence of undesirable side effects. Clearly, better patient stratification, identification of novel prognostic/predictive markers will help devise more personalized treatments and help overcome these problems. Extracellular receptors and/or secreted growth factors that drive cancer metastasis make provocative targets as they are easily accessible on the cell surface and usually show high response rates at lower drug doses, resulting in negligible side effects. To discover novel candidate receptor/secreted oncogenes, we used a whole-genome data-mining approach that takes advantage of large microarray studies of breast tumors annotated for clinical outcomes. Based on analysis of 14 independent breast cancer cohorts (2027 patients), we identified more than 30 cell surface/secreted proteins that when expressed highly, were associated with poor patient survival across multiple cohorts. This enriched gene set offered unexplored opportunities for immunotherapy and drug development. We established a cell-based screening platform to prioritize the most clinically relevant genes, and explored pathways downstream to each candidate in order to develop assay read-outs for future drug/antibody/compound screening. Our efforts identified two genes with great translational potential: Serine protease inhibitor Kazal-type 1 (SPINK1) and Jumonji Domain Containing protein 6 (JMJD6). SPINK1 affected multiple aggressive properties in breast cancer: survival, invasiveness and chemoresistance. Because SPINK1 effects could be abrogated by neutralizing antibodies, we suggest that SPINK1 is a viable potential therapeutic target. On the other hand, JMJD6, a histone arginine demethylase emerged as a marker of tumor aggressiveness, therapy resistance and affected both cell growth as well as cell shape. Our results suggest that the extrapolation of gene-survival associations in primary tumors to phenotypic analysis *in vitro* holds promise as a platform for discovering new therapeutic targets.

## C11: Systems biology at the single cell level

*Calin Guet*
*Institute of Science and Technology, Vienna, Austria*

The level of the single cell is fundamental when extracting relevant parameters for modelling endeavors. Most experimental work at the level of the single cell has been done on classic systems such as the lac operon or bacterial chemotaxis. When is it relevant to study the biology of the single cell? The mar operon will serve as a case study.

# C12: Does a critical peptide fold hold the key to Alzheimers?

*Sudipta Maiti*
*Department of Chemical Sciences,*
*Tata Institute of Fundamental Research, Mumbai*

Aggregation of the Amyloid Beta peptide is a likely cause of Alzheimer's disease. It is believed that the peptide also changes its molecular conformation as it aggregates, though details of this process are not known. We hypothesize that folding is the critical event that turns a benign peptide into a toxic one. Understanding this process would require a simultaneous investigation of aggregation, conformation and bio-activity. We use Fluorescence Correlation Spectroscopy (FCS), Forster Resonance Energy Transfer (FRET), fluorescence quenching, and vesicle binding and cell imaging assays to address these questions in physiological buffer conditions. We show that the major conformational transition (where the peptide end-to-end FRET efficiency changes from <25 % to >50 %) takes place right at the first step of aggregation, viz. with the formation of the small oligomers. This change predominantly involves the core region of the peptide. This monomer to oligomer transition increases the affinity of A$\beta$ for artificial lipid vesicles by at least an order of magnitude. Confocal microscopy shows that the oligomers at physiological concentrations have a strong affinity for the plasma membrane of living HEK 293T cells, while the monomers at the same concentration do not have any detectable affinity. Our results imply that the major change of molecular structure of amyloid beta occurs at the initial step of aggregation, and suggest that this plays a major role in transforming nontoxic monomers into toxic cell-adherent oligomers.

# C13: Kinetics of gene regulatory networks: predicting novel interactions and understanding signal integration

*Partho Sarothi Ray*
*Department of Biological Sciences,*
*Indian Institute of Science Education and Research, Kolkata*

Biological networks provide insightful means to elucidate information transfer in biology and have proved useful in understanding the functioning of living systems at the whole-system level. Most biological networks such as gene regulatory networks, metabolic networks and signaling networks represent molecular interaction networks, and do not contain or represent kinetic information. However, it is crucial to understand the kinetics of the different components of the network, as it is the dynamic changes in the concentrations and the relative affinities of the network components that finally cause changes in cellular function, especially in response to stimuli. Therefore, we have endeavored to develop dynamic model simulations, based on experimental kinetic data, to describe the functioning of gene regulatory networks. We have used this methodology to investigate experimental observations in the IFN-Gamma Activated Inhibitor of Translation (GAIT) system of translation regulation of pro-inflammatory genes. Interestingly, dynamic modeling of the system predicted a novel interaction and led to the discovery of an unprecedented molecular regulatory mechanism. Therefore, incorporation of kinetic information in a gene regulatory network not only leads to better understanding of the dynamics of the network but has important predictive functions that enrich the network with more comprehensive interaction information. We are now trying to apply the methodology to address the question of signal integration at single gene level, which is a fundamental question in biological regulation, using RNA-binding protein (RBP) and microRNA-mediated regulation of translation of p53 as a model system.

# C14: Interacting networks from genome-scale data with applications to complex disease genetics

*Tom Michoel*
*The Roslin Institute, University of Edinburgh, UK*

One of the central hypotheses of systems biology is that molecular networks regulate protein levels which affect physiological states. When perturbed by genetic or environmental factors, these networks may become dysfunctional, and thereby cause disease. In human populations, naturally occuring DNA variants together with environmental perturbations induce changes in mRNA expression levels from which molecular disease networks can be reconstructed. I will present mathematical and computational methods (i) to predict disease networks from integrated genetics, genomics and clinical data and (ii) to study interacting networks across regulatory levels, tissues or organisms. As an application, I will demonstrate how these methods are used to reveal the tissue distribution of inherited risk for coronary artery disease, using a dataset of more than 700 microarrays profiled in different vascular and metabolic tissues from 150 heart disease patients (the Stockholm Atherosclerosis Gene Expression (STAGE) cohort).

## Abstracts of Poster Presentations

# KOLKATA INTERNATIONAL SCHOOL CUM CONFERENCE ON SYSTEMS BIOLOGY (KOLSYSBIO)

## December 29, 2012 to January 3, 2013

*CONVENORS:*

*Pradeep K Mohanty, SINP*
*Soumen Roy, Bose Institute*

# P01: Comparative network analysis

*Sumeet Agarwal*
*Department of Electrical Engineering,*
*Indian Institute of Technology Delhi*
*Hauz Khas, New Delhi - 110016*

Many real-world systems are naturally represented as networks, and a variety of measures exist that characterise their structure. However, studies of networks typically employ only small, partly arbitrarily selected subsets of these, and the lack of a comparison makes it unclear which structural diagnostics are redundant or complementary. We present a highly comparative study of networks and network features, analysing a wide variety of networks, derived both from empirical observation and from mathematical models. We make use of a total of over four hundred network metrics or summary statistics thereof. We demonstrate how our approach can be used to organise and classify networks, as well as to obtain insights into how network structure relates to functionally relevant characteristics in a variety of settings. These include detecting structural features of metabolic networks that correlate with biological evolution, and constructing summary statistics that allow for efficient fitting of evolutionary models to observed protein interaction networks, via a novel Approximate Bayesian algorithm. Our methodology provides a general-purpose data-driven approach to aid in the study and understanding of networked systems.

# P02: Systems approaches for identification of essential genes: *Mycobacterium tuberculosis* as case study

*Priyanka Baloni, Soma Ghosh, Sumanta Mukherjee, Praveen Anand and Nagasuma Chandra*
*Department of Biochemistry, Indian Institute of Science*

A systems approach is necessary to identify essential genes in an organism. Essential genes are defined as the minimal set of genes required for the survival of an organism. In the present work, we report an integrated approach to identify a set of essential genes in the causative agent of tuberculosis, Mycobacterium tuberculosis (M.tb). Different levels of studies have been incorporated to overcome the limitations of individual methods. As a first criterion, genes which show consistently high expression values in the 39 wild-type samples are short-listed as important for the survival of the organism. Next, at the metabolic level, FBA combined with gene expression data is used to identify genes that result in lethal knock-outs. Following that, a weighted and directed protein-protein interaction network is obtained from various sources and important controlling points identified using broken path analysis. The genes are then checked for phyletic retention with the basic assumption that genes that are conserved across the species are essential for the survival of the organism. The genes identified from the different methods are further combined using set-theory to obtain a final list of 229 essential genes. These are further validated using the transposon mutagenesis dataset, showing an overall agreement of 73for the 229 genes conserved amino acid residue of its protein products are identified and mapped on the binding sites of the protein. It is seen that majority of the conserved residues fall on the binding site. Thus, a general and robust methodology has been developed for the identification of essential genes in M.tb. It should be noted that the method can be easily extrapolated to other organisms.

# P03: Computational study of differential gene expression in *Bordetella pertussis*

*Arnab Bandyopadhyay and Suman K Banik*
*Department of Chemistry, Bose Institute*
*93/1 A P C Road, Kolkata 700 009, India*

Bacterial pathogens undergo profound physiological changes when they infect their host and require coordinated regulation of gene expression in response to the stress encountered during infection. *Bordetella pertussis*, a human pathogen and causative agent for the disease whooping cough, expresses a number of virulence factors that enable the bacteria to colonize in the respiratory tract of the host, intoxicate the host and cause disease. A number of environmental factors, such as temperature and chemical agents, are known to control the transcription of the virulence related genes, e.g., filamentous haemagglutinin, pertussis toxin, adenylate cyclase etc., in *B. pertussis*. The expression of virulence factors is coordinately regulated by the *bvg* locus, which encodes the proteins BvgA and BvgS, the sensor and the response regulator domain, respectively, of *B. pertussis*. By sequence homology these fall into the 'two component' family of bacterial signal transduction machinery. In the first step of activation, *bvg* locus activates its own autoregulated promoter and the promoter of the adherence factor filamentous haemagglutinin. The second step occurs several hours later and consist of the activation of adenylate cyclase and pertussis toxin genes.

The aforesaid phenomena can be explained in terms of reaction kinetics, as in a living cell several sub-cellular events occur as a manifestation of biochemical reactions. Considering all the biochemical reactions follow mass action kinetics we have developed a computational model for signal transduction and differential gene regulation in *B. pertussis*, that explains the existing experimental results.

# P04: Key to network controllability

*Soumya Jyoti Banerjee and Soumen roy*
*Bose Institute, 93/1 APC Roy Road,Kolkata 700009 India*

Controlling complex systems has always been a challenge. Of late, there has been significant interest in controllability of complex networks. Recent studies has claimed that a network's controllability is to a great extent encoded by the underlying network's degree distribution $P(k_{in}, k_{out})$ and degree correlations. Is the controllability of a network decided almost completely by the immediate neighborhood of a node, while, even slightly distant nodes play no role at all? Motivated by the above question, we tried to define controllability metric in a more general way, using distance based measures like closeness centrality and betweenness centrality of a network. We want to show that controllability of a network may not only require the purely local connectivity measures like in-degree and out-degree. Degree reflects information about the immediate neighborhood of a node. However, closeness centrality encodes both local and global information and signifies a node's potential to choose good control paths passing through it. Again, a node with high betweenness centrality could be rather distant from the node in question but could connect it to a matched or unmatched node. Thus, we are immediately led to investigate the important role that closeness centrality and betweenness centrality should play in deciding controllability.

# P05: Mutations alter protein interactions leading to loss and gain of biological functions

*Mahashweta Basu, Nitai P. Bhattacharyya and P. K. Mohanty*
*Saha Institute of Nuclear Physics*
*1/AF-Bidhannagar, Kolkata-700064*

We tested the hypothesis that loss/gain of interactions of mutant HTT which causes Huntingtons disease (HD), or mutant TP53 involved in various cancers, result in loss or gain of biological function(s) in respective pathological conditions. We have constructed protein-protein interaction network (PPIN) with the proteins interacting with wild type and mutant proteins and modularized the networks. Functional role of modules are assessed by Gene Ontology (GO) enrichment analysis for the biological processes (BPs). Such analyses reveal the plausible candidate BPs of the modules. The unique GO terms enriched significantly (Hyp ¡ 0.0001) with proteins in modules of mutant PPIN indicate gain of BPs due to mutation whereas the enriched GO terms in wild type PPIN represents the loss. Several other enriched GO terms common in the modules of mutant and wild type networks indicate loss as well as gain of biological functions in pathological conditions. It turns out that most of BPs assigned to the modules of these protein networks are already known to be altered in HD or cancers. We argue that gain of BPs, and thus the biological functions, are due to new interacting partners acquired by mutant proteins. The methodology we adopted here could be applied to genetic diseases where the mutation alters the ability of the protein to interact with other proteins.

# P06: Formal methods for analysis of biological systems

*Sukanya Basu and Supratik Chakraborty*
*Dept of Computer Science and Engineering, IIT Bombay*

Recently, there has been an explosion of interest in applying mathematical techniques to understand system-level dynamics of biological systems, and even to predict outcomes of virtual biological experiments.There are several open questions that need be answered in order to develop a meaningful abstraction-refinement framework for biological systems. The first question concerns the discovery of an appropriate abstraction function and the corresponding concretization function. In the context of static analysis and formal verification of computer systems, abstraction and concretization functions have been traditionally been defined as pairs in a Galois connection between a concrete state space and an abstract state space. One needs to re-examine these notions in the context of biological networks. In addition, one needs to devise techniques for automatic discovery of abstractions. As an initial proposal, we wish to use ideas from predicate abstraction of programs to abstract the information represented in a biological network using a set of carefully chosen and biologically relevant predicates. An n-dimensional predicate is a mapping from an n-dimensional domain to the range True, False. Refinement of an abstraction obtained using a set of predicates has traditionally been done (in the context of program verification) by making available additional predicates for use in abstracting the concrete behaviour. In the context of biological networks, this approach can be used as well. However, the nature of biological systems offers additional opportunities for refinement. Yet another important aspect to consider is the choice of the underlying reasoning engine. How does one represent the abstraction, and analyze it to obtain meaningful sub-structures? In this context, the use of satisfiability and/or constraint solvers appears to be a promising direction to explore. We have some very preliminary evidence that this might be a fruitful line to pursue. Needless to say, this requires a significant amount of additional investigation.

We propose to develop an iterative abstraction-refinement framework for analyzing large biological networks. The goal would be to identify all sub-structures/ information content from the network that are relevant to a accuracy by a biologist. We should be able to provide guarantees that no relevant sub-structure has been missed by our analysis, and strive to ensure that as few non-relevant sub-structures as possible are returned by the analysis.

# P07: Construction of a colon cancer differentiation model to identify new targets for colorectal cancer

*Nirmalya Dasgupta, Bhupesh Kumar Thakur, Atri Ta, Santasabuj Das*
*National Institute of Cholera and Enteric Diseases,*
*P-33 C.I.T. Rd., Scheme XM, Beliaghata, Kolkata 700010, India.*

Colorectal cancer ($CRC$) is the third most common form of cancer and the second leading cause of death among cancers worldwide. Sequential accumulation of mutations in specific genes, such as adenomatous polyposis coli (APC), Kirsten-ras ($K-ras$) and p53 drives the transition from healthy colonic epithelia through increasingly dysplastic adenoma to colorectal cancer. Human colon cancer-derived cell line HT-29 expresses mutant APC and p53 proteins, the commonest combination of mutations (27.1%) in colon cancers. However, HT-29 cells can be differentiated, which ultimately leads to apoptosis of the cells by butyrate, a bacterial metabolite present in the intestinal lumen. To achieve differentiation, butyrate regulates approximately 1500 genes at the transcriptional level. We constructed a Butyrate Network with these 1500 genes. Analysis of the network divided the hubs into two categories with respect to their presence in the p53-APC interactome. Substantial differences in the biochemical properties and degree of the hubs were observed between the two categories. Butyrate network was more vulnerable to the removal of the hub proteins, which were shared by the p53-APC interactome. We identified Syk, a non-receptor tyrosine kinase as a butyrate regulated hub protein present in the p53-APC interactome. Expression of Syk was associated with Breast and head and neck carcinomas. However, its role in colon carcinomas is yet to be investigated. We found that Syk may play an important role in colon cancer growth and metastasis. This association needs to be further investigated through clinical studies.

# P08: Rule-based modeling of host-pathogen interactions: a focus on tuberculosis

*Soma Ghosh, Prasad KVS, Saraswathi Vishveshwara, Nagasuma Chandra*
*Molecular Biophysics Unit, Indian Institute of Science;*
*Computer Science and Engineering, Chalmers University of Technology;*
*Department of Biochemistry, Indian Institute of Science*

Iron is an essential element in most biological cells, required for the function of many proteins in diverse processes such as oxygen transport, repression, detoxification and DNA synthesis. Intricate mechanisms have evolved to tightly regulate iron metabolism both at the cellular and organismal levels. Iron overload as well as iron insufficiency are both known to lead to pathological conditions such as hemochromatosis and anemia respectively. *Mycobacterium tuberculosis (M.tb)* has formulated various means of attacking the host system to establish itself within the host system. One such crucial strategy is the exploitation of the iron resources of the host system. Both host and pathogen compete with each other to obtain and maintain the required concentration of iron through complex molecular interactions. The extent of complexity, makes it important to obtain a systems perspective of the interplay between host and pathogen with respect to iron homeostasis. We have reconstructed a systems model comprising of 66 components and 78 protein-protein or protein-metabolite interactions, captured as a set of 194 rules. A rule-based modelling approach, Kappa is used to simulate the system separately under infection and non-infection conditions. Various perturbations including knock-outs and dual perturbation are carried out to identify bottle necks in the network that can lead to decrease in the pathogenesis. The model is able to re-establish the importance of iron dependent regulator (ideR) in M.tb and transferrin (Tf) in the host. Perturbations, where iron storage is increased appear to enhance nutritional immunity and the analysis indicates its adverse effects on the host, while decreasing the rate of iron uptake by Tf is shown to be beneficial for the host. Simulation and perturbation studies help in identifying Tf as a possible drug target. Regulating mycobactin concentration is also identified as a possible strategy to control bacterial growth. The model reported in this study presents a comprehensive framework to study iron homeostasis and provide significant insights into iron homeostasis and potential drug targets for combating TB infection.

# P09: Gene set control analysis (GSCA) predicts haematopoietic control mechanisms from genome-wide transcription factor binding data.

*Dr. Anagha Joshi*
*The Roslin Institute, University of Edinburgh, UK*

Transcription factors are key regulators of both normal and malignant haematopoiesis. Chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-Seq) has become the method of choice to interrogate the genome-wide impact of transcription factors. Here we have collected and integrated 142 publicly available ChIP-Seq datasets for both normal and leukaemic murine blood cell types, and introduce the new bioinformatic tool Gene Set Control Analysis (GSCA). GSCA predicts likely upstream regulators for lists of genes based on statistical significance of binding event enrichment within the gene loci of a user-supplied gene set. We show that GSCA analysis of lineage-restricted gene sets reveals expected as well as previously unrecognised candidate upstream regulators. Moreover, application of GSCA to leukaemic gene sets allowed us to predict the reactivation of blood stem cell control mechanisms as a likely contributor to LMO2 driven leukaemia as well as to clarify the recent debate on the role of Myc in leukaemia stem cell transcriptional programmes. GSCA therefore provides a valuable new addition to analysing gene sets of interest, complementary to Gene Ontology and Gene Set Enrichment analyses. To facilitate access to the wider research community, we have implemented GSCA as freely accessible web tool (http://bioinformatics.cscr.cam.ac.uk/GSCA/GSCA.html).

# P10: Structural and functional analysis of hypothetical proteins in Mycobacterium tuberculosis using bioinformatics tools

*Lakshmi Maganti, Nanda Ghoshal*
*Structural Biology and Bioinformatics Division,*
*CSIR - Indian Institute of Chemical Biology, Kolkata- 700032.*

Genome sequencing projects have led to an explosion of large amount of gene products of which many are hypothetical proteins with unknown function. Analyzing and annotating the functions of hypothetical proteins are important in Mycobacterium tuberculosis (Mtb), the causative agent of tuberculosis (TB), which remains as a serious public health threat. Genome sequencing of Mycobacterium tuberculosis holds promise for the development of new drug candidates and search for protein coding regions in the genes. However, up to 50% of genes within a genome are often labeled unknown, uncharacterized or hypothetical. The biological functions of proteins encoded by these genes are not known limiting our understanding of virulence and pathogenicity of these organisms. In this study, six hypothetical proteins of Mycobacterium tuberculosis were retrieved from swissprot and analyzed for their structural and functional characteristics by using various bioinformatics tools like CDD-BLAST, INTERPROSCAN, PFAM, STRING. The analyses revealed that some of them possessed functionally important domains, protein-protein interacting partners and belong to families, which are Ppx-GPPA phosphate, Exonuclease, CblQ-cobalt transport protein, peptidase S13, Glycerate kinase, DNA primase S. This suggests that those hypothetical proteins may have the functions of the respective families to which they belong. In this way, searching functional conserved domains and structure prediction of uncharacterized (hypothetical) proteins will make them the future drug targets by identifying their existence in the metabolic pathways of Mycobacterium tuberculosis life cycle. The structure prediction of those proteins along with identification of the binding sites have been done, which may be useful for identifying novel drug candidates against these hypothetical proteins to assist in the control of tuberculosis.

# P11: Histone acetyl transferase-1 from Leishmania donovani (LdHAT1) is regulated by S-phase kinase LdCyc1-CRK3

*Anup Kumar Maity and Partha Saha*
*Crystallography and Molecular Biology Division,*
*Saha Institute of Nuclear Physics*
*1/AF-Bidhannagar, Kolkata-700064*

Protein phosphorylation is a key regulatory post-translational modification and cyclin dependent protein kinases (Cdks) play a central role in the regulation of cell cycle progression and cell division by reversibly phosphorylating different proteins in eukaryotes. Several Cdk related kinases (CRKs) and cyclins have been identified in the early branching pathogenic kinetoplastida parasites Leishmania and an S-phase cell cycle kinase LdCyc1-CRK3 from Leishmania donovani was characterized from our laboratory. The cyclin subunit LdCyc1 contains a conserved MRAIL motif, which is responsible for interaction with proteins harboring RXL type cyclin binding (Cy) motif. Hence, a screening for substrates was carried out based on the presence of Cy motif as well as Cdk phosphorylation site (S/T-P-X-R/K) in the targets resulting in the identification of three substrates of LdCyc1-CRK3. All the three substrates are phosphorylated by LdCyc1-CRK3 in a Cy-motif dependent manner. Among the three identified substrates, one is a unique protein with no known homologues. Another one contains Ku-70 related conserved domain and the third one, which we term as LdHAT1, is similar to MYST family histone acetyl transferase (HAT). In eukaryotes, histone acetylation is one of the major posttranslational modifications that remodels chromatin to regulate DNA replication and repair, gene expression, cell proliferation and terminal differentiation. We have shown that LdHAT1 acetylates histone H4 at K10 residue, and interestingly, the HAT activity is down-regulated after its phosphorylation by LdCyc1-CRK3. Therefore, the S-phase specific kinase LdCyc1-CRK3 may play an important role in the regulation of LdHAT1 activity through phosphorylation to control the cell cycle and life cycle specific chromatin dynamics in L. donovani.

# P12: Characterization of functional domain in natively unfolded human proteins

*Mritunjoy Maity and Nakul C. Maiti*
*Structural Biology and Bioinformatics Division,*
*Indian Institute of Chemical Biology*
*Council of Scientific and Industrial Research,*
*4, Raja S.C. Mullick Road, Kolkata 700032*

Studying unstructured proteins are of growing importance as it was recently discovered that many proteins without any globular fold involve in cell signalling and pathology of neurological disorder and other diseases. We computationally defined the compositional aspects of the binding/functional region of a particular class of human proteins which poses little or no globular fold. Many of these proteins possessed multiple binding regions (BR) and the sequence length of the BRs varied from region to region. Number of residues in binding region found to increase with the protein sequence length and distributed throughout protein sequence. The regions were distributed throughout the protein sequence and not localized. Grand average of hydropathy of most of the protein was negative, however many binding regions were with + ve values. Majority of residues in the binding region showed extended/random coil conformation possibly making them adaptable to different target molecules.

# P13: Interplay between Fur and HNS in controlling virulence gene expression in Salmonella typhimurium

*Mahendra Kumar Prajapat and Supreet Saini*

*Chemical Engineering, Indian Institute of Technology Gandhinagar, VGEC Campus, Chandkheda, Ahmdeabad, Gujarat 382424, India; Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra - 400067, India*

*Salmonella enterica* is a pathogen responsible for a large number of enteric diseases in a wide-range of hosts. Fur (Ferric uptake regulator) and HNS (histone-like small nucleoid-associated protein) are two of the global regulators involved in controlling gene expression during the infection cycle of the *Salmonella*. Fur is a $Fe^{2+}$ dependent transcriptional activator or repressor of gene expression. HNS is a DNA binding transcriptional regulator which negatively regulates gene expression by shielding DNA from transcription factors and RNA polymerase. In this work, we demonstrate computationally that Fur and HNS have disproportionately high density of binding sites in the Pathogenicity Islands on the *Salmonella* chromosome. Moreover, the frequency of binding sites for the two proteins is correlated throughout the genome of the organism. These results indicate a complex interplay between Fur and HNS in regulating cellular global behavior.

# P14: Potent anticancer activity of small peptides and their physical properties

*Biswadip Banerji, Sumit Kumar Pramanik,*
*Uttam Pal, Nakul Chandra Maiti*
*Department of Chemistry and*
*Department of Structural Biology and Bioinformatics;*
*CSIR-Indian Institute of Chemical Biology*
*4, Raja S.C. Mullick Road, Kolkata - 700032, India*

Cancer is a severe threat to human society. In the scientific community worldwide cancer remains a big challenge as there are no remedies as of now. Cancer is quite complicated as it involves multiple signaling pathways and it may be caused by genetic disorders. Various natural products and synthetic molecules have been designed to prevent cell proliferation. Peptide-based anticancer drugs, however, are not explored properly. Though peptides have their inherent proteolytic instability; they could still act as anticancer agents. Here we wish to report small peptide which can act as anticancer molecule. Potent anticancer activities were confirmed by MTT assay (a laboratory test and a standard colorimetric assay, which measures changes in colour, for measuring cellular proliferation) and phase contrast images. The IC50 value of these compounds ranges in the low-micromolar level.

# P15: Investigating gene co-expression modules in optimally growing Mycobacterium tuberculosis

*Bhanwar Lal Puniya, Srikant Prasad Verma and Srinivasan Ramachandran*
*CSIR-Institute of Genomics and Integrative Biology*

Biological systems are robust, which allow them to survive in changing environmental conditions. These robust traits have tendency to be selected in evolution. Modularity in the cellular organization of dynamics of expression is a one of the property of robust biological system. In this work we have constructed weighted gene co-expression network from gene expression data of mid log phase of optimally growing different strains of Mycobacterium tuberculosis. We have used 1068 genes, which have statistically significant variation in gene expression at least between one pair of strains. We have identified 7 co-expressed modules out of which 4 modules are in good agreement with biological enrichment analysis. Turquoise module was enriched with fatty acid biosynthesis, DNA replication and non membrane bound intracellular organelles. The blue module was enriched with two component system. Blue module has NADH dehydrogenase and Cell membrane. The Black module was containing mostly Transposons and transposition related genes. These modules were used for investigating highly weighted connections. These highly weighted connections were filtered by using weight threshold of 0.08. We have compared these connected gene pairs with other sources such as STRING database to identify the genomic neighborhood, gene fusion, conserved co-expression and phylogenetic co-occurrences. To identify the common transcriptional regulation between genes we have compared these with recently published Transcriptional Regulatory Network. We have found 202 connections out of total 1642 connections were in agreement. We have used these filtered network modules for identification of coordinated biological functions and hubs which could be attractive candidates for drug target identification. We identified Rv1611 (trpC), Rv0545c (pitA) and Rv3404c as top hubs. We found 39 high confidence drug targets mapped on these modules. Top 3 hubs were already predicted as high confidence drug targets. We have identified 7 novel candidates for drug targets which have degree greater than average degree of mapped high confidence drug targets and lacks homology with human proteins. Using this information we have proposed Rv0996, pknA, fgd2, lpqL, murC and phos1 as novel candidates for drug targets.

# P16: Use of structural bioinformatics to deduce novel functions of proteins

*Amit Kumar Singh Gautam, Prasanna lab, ACTREC*

The primary structure of protein defines its structure, function and localization. Short amino acid sequences play an important role in signaling and carrying localization or retention signals. We hypothesize that solvent exposed short sequences may play central role in protein-protein interaction and can bring together proteins of very different functions in a system wide network. We have tested this principle under different biological conditions-
1) In order to globally profile natural substrate of human endo-proteases, we have designed a novel method. We used minimal protease recognition sequence (4 residues) to fish out the putative substrates and filter them based on whether the recognition sequence is exposed to solvent (and hence accessible for protease) and whether protease and substrate are in the same cellular compartment. We have developed an online server to predict natural endo-protease substrate with high confidence (PNSAS). As a proof of principle, we identified several novel substrates for a serine protease matriptase, elevated level of which has been associated with several diseases including cancer. One of the novel putative matriptase substrate was desmogleine-2 (Dsg-2).
2) We designed tetra amino acid peptide from C-terminus of proteins and screened for binding with PSMD9. PSMD9 is a non-ATPases shuttling proteasome subunit containing PDZ like domain. The function of PSMD9 is not well understood; with the help of Cterminal short peptides we identified some novel binding partners of PSMD9. By pull down and immunoprecipitation experiments we showed that protein containing short sequence also interact with PSMD9 and mutation or deletion of these residues abrogate binding. Using classical binding studies, homology based modeling, structure based analysis of binding sites and site directed mutagenesis, we have provided the first glimpse of the probable structure of PDZ-like domain of PSMD9, its binding pockets and their role in protein-protein recognition.
3) Oncogene PSMD10 interacts with one of ATPases subunit of proteasome. The EEVD sequence was essential for this interaction. Since PSMD10 is an oncogene it will rewire cellular networks. It, therefore, becomes very important to identify the binding partners. We have identified several novel PSMD10 binding partners by searching the EEXD sequences in proteome and filtering them based on solvent exposure. Many of these interactions have been proved by immunoprecipitation and mutagenesis.

# P17: Bringing Down the House: A Non-linear Pay-off Driven Growth Model for Remediation of Antibiotic Resistant Infection by Introduction of Cheats

*Sanhita Ray and Anjan Kr. Dasgupta*
*Department of Biochemistry*
*University of Calcutta*
*Kolkata 700019*

Microbial cheats may serve as a useful weapon in the war against antibiotic resistant infections. Cheaters that do not produce public goods (e.g. beta-lactamase) required for conferring antibiotic resistance may destabilize a resistant population when introduced into the same niche. However, the conditions under which such remediation would occur can only be understood with the help of realistic dynamical models. We propose a model wherein a pay-off term having non-linear dependence on worker frequency decreases the death rate attributable to antibiotics. Simulations and analytical results show: if cheaters have a critical advantage in terms of growth rate, even low antibiotic dosages are sufficient for bringing down a resistant infection. We also deal with the question of how transference of resistance genes to cheaters may affect results.

# P18: Towards cataloguing peptide-mediated protein-protein interactions

*Debasree Sarkar, Tanmoy Jana and Sudipto Saha*
*Bose Institute*
*Kolkata-700054*

Protein Protein Interactions (PPIs) are known to be regulated by linear motifs (LMs), which are short elements embedded within large protein sequences. These short LMs are often present in the disordered regions in the eukaryotic proteins and are responsible for many protein-protein interactions. LMs plays a vital role in many cell signalling pathways including Wnt , MAPK and TGF in forming peptide mediated PPIs and complexes. To better understand the role of the LMs in pathways involved in oncoproteins including p53, RAS, MYC and APC, we focus on compiling and curating LMs using text mining followed by manual curation. We are developing a database with an aim to curate the LMs to avoid false positives by using filtering logic such as surface accessibility, structure, cellular compartment and orthogonal search in different species. Targeting protein networks using LMs have a potential in developing drug targets.

# P19: Stochastic optimization based study of dimerization kinetics

*Srijeeta Talukder, Shrabani Sen, Ralf Metzler, Suman K Banik, Pinaki Chaudhury ; Department of Chemistry, University of Calcutta; Institute for Physics & Astronomy, University of Potsdam; Department of Chemistry, Bose Institute*

The evaluation of correct individual step based pathways (both the nature of the reaction as well as the correct magnitude of the rate constant) present in a multi-step reaction scheme is central in establishing a complete reaction model in chemical and biological reaction networks. The conventional way to study reaction kinetics is to write down the mean field rate equations for the process, integrate them and follow the variation in the concentration of each species involved as a function of time. However this strategy is too simplistic and fails in situations where the number of reacting species is small (e.g. biochemical reaction in living cell), since for small number of particles fluctuations in the species population become relevant. In the conventional rate equations approach, it is assumed that the process is both continuous and deterministic. Reaction kinetics at such low concentrations are intrinsically discrete and stochastic. The stochastic simulation algorithm (SSA) is an elegant formulation to incorporate these effects and predict cor- rect results in a complicated multi-step reaction network. The rate constants associated with each individual step in a multi-step reaction scheme might not always be known a priori, or there might be a range of values of the rate constants, for which predictions for the overall reaction are compatible with experimental data. The correct prediction of all individual rate constants is not always an easy task and involves an optimization process. If an optimization scheme can be linked to SSA, then it should be possible to evaluate a correct set of reaction parameters, quantifying the complete kinetic behavior of a reaction network. We investigate the potential of numerical algorithms to decipher the kinetic parameters involved in a multi-step chemical reaction. To this end we study a dimerization kinetics of protein as a model system. We follow the dimerization kinetics using a stochastic simulation algorithm and combine it with three different optimization techniques (Genetic Algorithm, Simulated Annealing and Parallel Tempering) to obtain the rate constants involved in each reaction step. We find good convergence of the numerical scheme to the rate constants of the process. We also perform a sensitivity test on the reaction kinetic parameters to see the relative effects of the parameters for the associated profile of the monomer/dimer distribution.

# P20: Drug targets in Leishmania spp

*Barnali Waugh and Rahul Banerjee*
*Crystallography and Molecular Biology Division*
*Saha Institute of Nuclear Physics*
*1/AF, BidhanNagar, Kolkata-64*

Leishmania spp. are the causative agents of a wide spectrum of diseases referred to as leishmaniasis. Of these, visceral leishmaniasis can prove to be fatal unless treated properly in time. Several thousands of patients are afflicted by these parasites especially amongst poorer sections of the populace in tropical/subtropical regions of third world countries. Till date, there is no vaccine for this disease and only a few drugs are available which have toxic side effects, in addition strains resistant to these drugs have also appeared.
Experimental approach:
Several proteins from the glycolytic pathway are unique to the parasite with respect to human and thus could qualify as valid drug targets. Specifically the protein phosphofructokinase has been validated as a drug target and its inhibition would be lethal to the parasite. The clones of four enzymes from the glycolytic pathway or related to it have been obtained (phosphofructokinase [PFK], fructose1,6-bisphosphatse[FBP],phosphoglyceratekinase[PGK], phosphoglucomutase[PGM] ) and with the exception of PFK have been successfully expressed. Of these four proteins, PGK has been purified to homogeneity. Biophysical studies on these proteins are envisaged in the near future. In addition, crystallization trials on Aldose reductase(ALR)) from L. donovani , yielded crystals which were found to diffract poorly. Presently the crystallization trials for improving the quality of the crystal are being carried out.
Computational approach:
Since the enzyme PFK is unique to the parasite and is a validated drug target a pharmacophore is being designed based on the structure of its inhibitors as reported in the literature. Initially two models of PFK were constructed using the crystal structures of PFK (in apo-form: 2HIG) and the same protein ligated to ATP (3F5M) from T. Brucei as templates. About thirty confirmed inhibitors with relevant IC50 values were selected from the literature, and their atomic models constructed using locally developed software, and subjected to several cycles of energy minimization. These molecules will now be docked onto the enzyme subsequently leading to the design of an effective pharmacophore.