Genomic Med. (2008) 2:295–302
DOI 10.1007/s11568-009-9088-3

ABSTRACTS

Computational biology and structural proteomics

226: In silico Putative Drug targets in *Leptospira interrogans* and Homology Modeling of UDP-*N*-acetylglucosamine 1-carboxyvinyltransferase MurA

Umamaheswari Amineni, Dibyabhaba Pradhan, Vani I. Priyadrshini, G. Subramnyam

Sri Venkateswara Institute of Medical Sciences University, Alipiri road, SVIMS, Tirupati 517 507, India

Infectious diseases are the second leading causes of death worldwide. There is a need to develop new antimicrobial agents. Traditional method of drug discovery is time consuming, yields few drug targets and little intracellular information guiding target selection and associated with side-effects. To overcome these difficulties, the focus in drug development has been shifted from genetics to genomics. Availability of bacterial and human genome sequence in public domains and the development of bioinformatics tools facilitate to identification of genes, essential for survival of pathogens, which can be used as putative drug targets for subjecting to experimental analysis. In this study by using computational genomic approach, we identified 1395 genes on chromosome 1 and 28 genes on chromosome two as essential in *Leptospira interrogans*. Only 15 genes on chromosome one were non-human homologs, when analyzed using subtractive genomic approach. Pathway analysis on these genes encode their key role in unique essential pathways such as pathogen's metabolism, persistence, virulence, flagellar biosynthesis, amino sugar metabolism and peptidoglycan biosynthesis etc. for the survival of *L. interrogans*. A homology model of one of the potential drug targets UDP-*N*-acetylglucosamine 1-carboxyvinyltransferase MurA was built using the available crystal structure of 1DLG.

227: Comparative Analysis of R and S isoforms of 12-Lipoxygenases: Homology modeling and docking studies

P. Aparoy, T. Leela, P. Reddanna

Dept. of Animal Sciences, University of Hyderabad, Gachi Bowli, Hyderabad, India

12R lipoxygenases (12R-LOX) is a member of the lipoxygenases family, a group of structurally related family of non-heme iron-containing dioxygenases involved in the oxygenation of polyunsaturated fatty acids. Since discovering its role in psoriasis, considerable efforts have been made to gain deeper understanding of 12R-LOX structure and function. The present study is aimed at predicting human 12R-LOX structure by constructing a homology model. Based upon Blast results, rabbit reticulocyte 15-Lipoxygenase 1LOX (protein data bank) was considered as a template for homology modeling. The 3D model was generated with MODELER in InsightII and further refined using CHARMm. The final refined model is further assessed and verified by Profile-3D and PROCHECK. The docking of the substrate, arachidonic acid was also performed and found consistent with results from mutagenesis experiments. Further to understand the relationship of protein structure with stereo specificity, a comparative analysis of 12R-LOX model was done with that of 12S-LOX homology model to identify differences in the binding site topology and interacting residues, which might be utilized to develop selective 12R-LOX inhibitors. In 12R-LOX the amino acids at the active site that can be targeted for strong hydrogen bonding interactions for inhibitor design are Glu394, Glu407, Asn438, Arg442, Ser452, Leu635 and Ile701. The active site of 12R-LOX shows pocket near the Gly441 residue, which may be the potential oxygen-binding pocket.

228: Intrinsic versus induced effects in DNA structure

[1]**Manju Bansal**, [1]Deepti Karandur, [2]Nikhila Jayaprakash, [1]Arvind Marathe

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India, [2]Department of Biotechnology, Indian Institute of Technology, Chennai-600036, India

An important question of biological relevance is the polymorphism of the double-helical DNA structure in its free form, and the changes that it undergoes upon protein-binding. We have analyzed a database of free DNA crystal structures to assess the inherent variability of the free DNA structure and have compared it with a database of protein-bound DNA crystal structures to ascertain the protein-induced variations. Most of the dinucleotide steps in free DNA display high flexibility, while protein binding prefers the duplex in B-DNA conformation and in certain cases, causes the DNA backbone to attain energetically unfavourable conformations. At the gross structural level, several

protein-bound DNA duplexes assume a curved conformation in the absence of any large protein-induced distortions, indicating that a series of normal structural parameters at the dinucleotide and trinucleotide level, similar to the ones in free B-DNA, can give rise to curvature at the overall level. Thus the free DNA molecule, even in the crystalline state, samples a large amount of conformational space without the aid of any large ligands. For most of the bound DNA structures, across a wide variety of protein families, the average parameters are quite close to the free 'B-like' DNA oligomer values, indicating that in a large number of cases, protein binding may be the result of the protein 'finding' a suitable stretch of genomic DNA in the 'right' conformation, at that time point and then locking it in that conformation for a certain time period.

The crystal structure database analysis was complemented by molecular dynamics studies on the quorum sensing transcription factor, traR, bound to the trabox DNA sequence-d($A_1T_2G_3T_4G_5$-$C_6A_7G_8A_9T_{10}C_{11}T_{12}G_{13}C_{14}A_{15}C_{16}A_{17}T_{18}$). Simulations have also been carried out on an unbound trabox and a mutated trabox sequence. The central spacer region d(A_7-T_{12}) in the unbound, unmutated trabox is relatively straight with a characteristic narrow minor groove, a conformation similar to the one observed in the crystal structure complex. The unbound, unmutated trabox samples a wide range of bent conformations, including the one observed in the traR-trabox complex crystal structure. Mutations $G_8 > C$ and $C_{11} > G$ introduce kinks around the CA/TG steps that distort the spacer region, and affect the groove width and the overall conformation of the trabox, possibly making it unfavourable for binding by either or both the monomers of the traR. Further analysis is being carried out.

229: Using support vector machine for the identification of disease-associated missense SNPs

Yasha Bhasin, Vinod Scaria, Bhupesh Taneja, Samir K. Brahmachari

Institute of Genomics and Integrative Biology, Mall Road, Delhi, India

Variations in the genome leading to change in concentration of proteins occur in regulatory regions of genes, whereas structural variations leading to loss or gain of function occur in exonic regions of genes. It is well known that few amino-acid residues play a crucial role in the three-dimensional architecture of proteins. Substitutions in these residues are likely to bring about a change in function of the protein. The quality and completeness of presently available SNP databases allows the application of machine learning techniques to identify key functional residues in proteins. In this work, we have developed an SVM-based method to predict the potential functional consequences of missense SNPs. We introduce a novel sequence-based attribute, 'exon conservation', based on premise that the invariant regions of proteins should be important for function, and any variation in these regions would be more likely to be deleterious (1, 2). The other attributes used by our method utilize individual properties of SNPs, including physico-chemical variation; post-translational modification; secondary structure; accessible surface area; functional domains, repeats and sites; transmembrane topology; and protein stability. The classifier achieved an overall accuracy of 75%. Analysis of the results will be presented.

References

1. Prakash T, Ramakrishnan C, Dash D, Brahmachari SK. Conformational analysis of invariant peptide sequences in bacterial genomes, J Mol Biol. 2005 Feb 4;345(5):937–55.
2. Prakash T, Khandelwal M, Dasgupta D, Dash D, Brahmachari SK. CoPS: Comprehensive Peptide Signature database, Bioinformatics. 2004 Nov 1;20(16):2886–8.

230: Structure-function correlations in LuxS from bacteria: Analysis of protein–protein interface clusters by graph theoretical approach

Moitrayee Bhattacharyya, Saraswathi Vishveshwara

Indian Institute of Science, Molecular Biophysics Unit, Bangalore, India

The genome of a wide variety of prokaryotes contains the luxs gene homologue, which encodes for the protein S-ribosylhomocysteine lyase (LuxS). This protein is responsible for the production of the quorum sensing molecule, AI-2, which coordinates changes in behavior as a function of cell density. Very often quorum sensing has been directly associated to pathogenicity. But in some prokaryotes no pathogenic role has been attributed to quorum sensing and it has been found to have an effect on metabolism, and thus the overall fitness and 'well being' of the organism. In some organisms, the LuxS has a role in controlling flagellar morphogenesis and it was also known to modulate the toxin production in some virulent bacteria. So it is a single protein performing diverse roles. In our study, we have considered LuxS from a variety of bacterial species. Some of the protein structures were modelled due to the lack of crystallographically available structures. The interface which contains the active site was known to be structurally and functionally important. Thus, the protein structure network (PSN) graphs were constructed to characterize the interface of this homodimeric protein. Our analysis revealed potential sites of mutation and geometric patterns on the interface that was not evident from conventional sequence alignment studies. The key features presented by the protein interface is being investigated for the classification of the proteins in relation to their function. This sort of characterization enables a better understanding of the relation between the luxs gene and its functional role in the prokaryotes.

231: A novel mechanism of action of *Mycobacterium tuberculosis* pantothenate kinase involving mobility of ligands

[1]**Bhaskar Chetnani**, [1]Satyabrata Das, [1]Parimal Kumar, [1,2]A. Surolia, [1]M. Vijayan

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore, 560012, India, [2]National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi, 110067, India

As part of a structural genomics programme on TB proteins, the crystal structures of pantothenate kinase from *Mycobacterium tuberculosis* (MtPanK) and several of its complexes have been determined. PanK performs ATP-mediated phosphorylation of pantothenate (vitamin B5) to form phosphopantothenate, which is the first step in the universal coenzyme A (CoA) biosynthesis pathway. The enzyme is feedback regulated by CoA and it is the primary mechanism by which bacteria maintain their intracellular concentration of CoA. In addition to the PanK-CoA complex, the present study primarily revolves around a complex with the ATP analogue AMPPCP and pantothenate (initiation complex), one with ADP and phosphopantothenate (end complex) and another containing ADP and ATP with half occupancies (intermediate complex). Also analyzed are a complex with ADP and another with ADP and pantothenate. The structures reveal a novel mechanism involving different, although overlapping, locations of ATP in the initiation complex and ADP in the end complex. Pantothenate in the initiation complex and phosphopantothenate in the end complex also occupy different locations. This mechanism is different from that proposed for the E.coli enzyme. The differences between the two

mechanisms can be explained in terms of structural alterations caused by differences in amino acid residues at specific locations. Substantial differences in mechanism of action to the extent reported here, between homologous enzymes are unusual. The kind of changes in the location of ligands during enzyme action exhibited by MtPanK is novel.

232: Three-dimensional structure of *Vibrio cholerae* Hemolysin oligomer by Cryoelectron Microscopy

Somnath Dutta, Budhaditya Mazumdar, K. K. Banerjee, A. N. Ghosh

National Institute of Cholera and Enteric Diseases, P-33, C.I.T. Road, Scheme-XM, Beleghata, Kolkata-700010, India

Vibrio cholerae hemolysin (HlyA) is an extra cellular membrane damaging protein. Molecular weight of hemolysin (HlyA) is 65000 Da. This protein exists in two stable states, a water-soluble monomer and an oligomeric integral membrane protein. The protein is synthesized as an 82 kDa preprohemolysin by *Vibrio cholerae* EI Tor 01 and non-Ol strains and exported to the culture medium as the 79 kDa prohemolysin (proHlyA). Proteolytic removal of the 132-residue N-terminal stretch generates the mature 65 kDa HlyA with a specific hemolytic activity. HlyA transforms itself in contact with target biomembranes and synthetic lipid vesicles containing cholesterol into water filled transmembrane heptameric channels of internal diameter 1.5 nm. Recent transmission electron micrographic study has revealed that HlyA binds to cholesterol and forms oligomers at the interface of cholesterol and water. The aim of the present study is to determine the three-dimensional structure of 65 kDa hemolysin using cryoelectron microscopy and single particle methods. Holey carbon grid was glow-discharged and 65 kDa hemolysin sample was placed on the surface of the holey grid. After blotting nearly to dry with a piece of filter paper, the grid was plunged into liquid ethane at −180°C and protein molecules were embedded in vitreous ice. The grid was placed on a GATAN 626DH cryo-holder and inserted into FEI Tecnai 12 BioTwin transmission electron microscope. Electron micrographs of frozen hydrated HlyA were taken at 120 kV. Images were taken at different defocus values using 'low dose' software. These micrographs were digitized using Nikon Coolscan 9000ED film scanner. The three-dimensional structure of Hly A was determined using the EMAN 1.7 software operating on Linux Fedora Core 4 platform. 3200 HlyA particles were selected using 'boxer' program of EMAN software. Contrast transfer function (CTF) correction was performed using 'ctfit' program of EMAN software. Rotational symmetry was assessed initially in EMAN using 'startcsym'. Finally three-dimensional reconstruction of 65 kDa *Vibrio cholerae* hemolysin was performed using 'refine' program of EMAN software. The resolution was determined by using a 0.5 Fourier shell coefficient (FSC) threshold. The 0.5 value of Fourier shell coefficient (FSC) indicates a final resolution was bellow 25 Å. Three-dimensional image of 65 kDa *Vibrio Cholerae* hemolysin was visualized using CHIMERA software.

233: Identification and analysis of novel repeats and domains in human proteome

Lalitha Guruprasad, Rao Satyanarayana, Hema Latha Golaconda

School of Chemistry, University of Hyderabad, GachiBowli Road, Hyderabad, 500 046, India

A predominant part of human genome consists of repetitive sequences, encompassing large segmental duplications, interspersed transposon derived repeats and tandem repeats. Amino acid repeats, known, as homopolymeric tracts are present in nearly one-fifth of human gene products. The uncontrolled expansion of trinucloetide repeats in human coding sequences is associated with several neurodegenerative disorders. Examples are Huntington's disease and dentatorubropallidolusyan atrophy, both associated to abnormally long expansions of CAG runs encoding polyglutamine tracts. Repeat structures in humans have been found to play vital roles in various biological functions such as signal transduction, apoptosis, transcription regulation and several diseases. Realizing the importance of amino acid repeats in proteins, we undertook the study of identifying the novel amino acid sequence repeats and domains in human proteome. We have implemented repeat identification method TRUST and automated the database searching methods that could be applied to a complete proteome. Using these methods for human proteome, we identified 7 domains and 18 repeats that have not been reported so far. Repeats in proteins comprise less than 55 amino acid residues and often present in multiple copy numbers and in tandem. All the repeats are required for correct folding and function of protein. Whereas, domains have greater than 55 amino acid residues and present as single or multiple copy numbers in proteins. Domains are structural and functional units and are independent of the rest of the protein. We have considered only those repeats with length greater than 25 amino acid residues in this work. Our analysis suggests that some of the repeats and domains identified in this work are associated with diseases such as polycystic kidney disease and Williams-Beuren syndrome. In this presentation, we discuss our automated methods for the identification of novel repeats and domains from any proteome. We further report our findings of novel repeats and domains from the human proteome analysis and suggest their functional importance and role in human diseases.

234: Targetability of *Mycobacterium tuberculosis*: Comparison of host and pathogen pocketomes

Yeturu Kalidas, Karthik Raman, Nagasuma Chandra

Bioinformatics Centre, IISc, Indian Institute of Science, Bangalore, India

Target identification is a critical step in modern drug discovery. Identifying the right target, however, is by no means simple, since a variety of factors need to be considered simultaneously. One of the pressing problems with most clinically used drugs are the adverse drug reactions due to interaction with unintended host proteins. No systematic method is available so far to address this issue. We have developed novel algorithms 'Pocket Depth' and 'Pocket Match' to identify putative binding site pockets in protein structures and to compare them with each other. We obtain all possible crystal structures and homology models of proteins and in some cases their domains in both the human and the *Mycobacterium tuberculosis* proteome and identify pockets in them using Pocket Depth, a new depth based method. Pockets are also identified by another algorithm Ligsitecsc that utilizes residue conservation information. Combining prediction from the two methods, we shortlist pockets which have high ranks and contain conserved amino acid residues in that family. We then carry out an all-vs.-all comparison of the shortlisted pockets in the two proteomes using Pocket Match. Proteins containing similar pockets are then clustered, and a targetability index is computed. Besides targetability, biological insights obtained about the similarities in binding pockets in diverse proteins will also be presented. This study provides a basis for a rational and systems-level method to understand drug pharmacodynamics and further to use such knowledge in discovery of new safer drugs for tuberculosis.

235: DNA replication firing efficiency profiling of human chromosomes 21 and 22 by mathematical modeling

R. Krishna Murthy Karuturi, Huaien Luo, Majid Eshaghi, Jianhua Liu

Genome Institute of Singapore, Genome, 60 Biopolis ST, S138672, Singapore

DNA replication is a key event in cell division cycle taking place within a short phase called S-phase. Profiling the DNA replication firing origins and efficiency have been of great interest to biologists. But the current accurate methods such as molecular combing cannot profile efficiency of all origins in the genome in high throughput manner while there are no other accurate high throughput methods for such task. Here we provide a mathematical model for DNA replication in eukaryotes and provide means of computing efficiency of firing using DNA replication pattern in high throughput time-course DNA replication experiments. Our analysis of DNA replication in *S. cerevesiae* and *S. pombe* showed that our model works well. Hence, using our model, we obtain DNA replication firing efficiency profile of Chr21 and Chr22 of Human genome. It reveals similar relationships between firing timing and efficiency as in *S. cerevesiae* and *S. Pombe*.

236: High quality manual genome annotation at WTSI

J. E. Loveland, C. A. Steward, J. P. Almeida, I. Barnes, D. R. Carvalho-Silva, A. Frankish, J. G. R. Gilbert, L. Gordon, E. Hart, J. M. Mudge, C. Snow, M. M. Suner, S. Trevanion, L. Wilming, J. L. Harrow, T. Hubbard

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs, United Kingdom

The HAVANA group from the Wellcome Trust Sanger Institute is responsible for the manual annotation of coding, transcript and pseudogene loci on human, mouse and zebrafish finished genomic sequence. The total number of protein coding genes and the extent of alternative splicing of loci on the human genome is still unclear. In collaboration with Ensembl, RefSeq at NCBI and UCSC, the CCDS project (Consensus CoDing Sequence) is working to define a core set of protein coding transcripts. Initially limited to human, the project was recently extended to include mouse. Any CCDS candidate transcripts where there is disagreement between the collaborators are manually re inspected, discussed and, where possible, an agreement is reached on a structure. The end result is a combined, non-redundant gene set that is an ongoing activity with the resolution of differences and the refinement of the gene set between CCDS update cycles. The HAVANA group is also leading the ENCyclopedia Of DNA Elements (ENCODE) scale-up of the gene annotation project (GENCODE). The project includes seven other partner institutes and will integrate expert manual annotation, computational predictions and targeted experimental analysis to generate a complete reference gene set for the whole genome. This will also include the analysis of pseudogenes, experimental validation of putative/novel genes and examination of the protein-coding potential of genes using comparative and structural analysis. This collaboration will also feed into the CCDS project and provide the genomic community with an accurate gene catalogue for the human genome.

All manual annotation produced by Havana group is displayed on the VErtebrate Genome Annotation (VEGA) database (http://vega.sanger.ac.uk/).

237: Defining expression signatures of known cancer genes using seriation analysis of SAGE libraries from Cancer Genome Anatomy Project (CGAP)

Olena Morozova, Vyacheslav Morozov, Martin Hirst, Marco Marra

Genome Sciences Centre, British Columbia Cancer Agency, Suite 100, 570 West 7th Avenue, Vancouver, BC V5Z 4E6, Canada

Cancer is a genomic disease involving many types of molecular aberrations. Recently, a number of cancer genome re-sequencing studies have identified an abundance of sequence variation in cancer samples. Since most of this variation is neutral, a primary challenge of modern cancer genomics has been to distinguish causative cancer mutations from the abundance of neutral polymorphisms. Evolutionary models wherein causative sequence variants are presumed to be under positive selection in somatic neoplasms have been adopted for this purpose. However, these models do not take into account the complex scope of a cancerous phenotype, including gene expression changes, and thus represent an initial development in the emerging field of cancer genomics. In order to develop more advanced models for predicting causative cancer variants (cancer drivers), a better understanding of the interplay of sequence variation and other phenotypic characteristics of cancer cells is required. To further our understanding of the role of gene expression in defining cancer driver mutations, we examined gene expression patterns of known cancer genes from the Cancer Gene Census (http://www.sanger.ac.uk/genetics/CGP/Census/) in human cancer samples represented in the Cancer Genome Anatomy Project (http://cgap.nci.nih.gov/) resource. The cancer genes in the Census were defined as such based on the clinical implications of mutations in their coding sequence in cancer patients. To identify common expression signatures among cancer genes, we used a novel seriation algorithm that reorders genes based on the similarity of their Serial analysis of gene expression (SAGE) profiles. The seriation of cancer genes expressed in hematopoietic tissues showed that the genes were either consistently up or consistently down regulated in Acute myelogenous leukemia cancer samples versus normal samples. In contrast, seriation analysis of Illumina sequence tag libraries from melanoma and normal skin revealed inconsistencies in the expression patterns of cancer genes in melanoma samples. The finding of characteristic expression signatures of cancer genes in a number of cancer samples of the same tumor type implies that directional selection of gene expression, in addition to that of DNA sequence, occurs in some cancer systems. Therefore, we suggest that gene expression information should be incorporated into the evolutionary models of cancer that currently use sequence data alone.

238: Design of Human non-pancreatic secretory phospholipase A2 (hnps-PLA2) inhibitors: A Qsar based approach and its confirmation with structure-based studies

Amit Nagal, Neha Sharma, Mohan Krishna, Hemedra Yadav, Jindal Pankaj, Pooran Solanki, Katta Hement, Swapnil Jaiswal

Birla Institute of Scientific Research, Statue Circle, Jaipur 302001, India

Phospholipases A2 (PLA2s) are enzymes that catalyze the hydrolysis of the sn-2 acyl ester linkage of phospholipids, producing fatty acids and lysophospholipids. Quantitative structural alignment was performed for a series of 72 indole inhibitors of the human non-pancreatic secretory phospholipase A2 (hnps-PLA2). Three-dimensional QSAR model was then established using the CoMFA method. The set of 72 compounds was divided into two subsets, the first one constituting the

training set (56 compounds), while the second constituted the test set (16 compounds). A good correlation between predicted and experimental activity data allows to validate the 3D QSAR model. A second and global 3D QSAR including all the compounds was established, allowing the creation of the hnps-PLA2 pharmacophore. Three-dimensional quantitative structure-activity relationship (3D-QSAR) models were developed. The studies included molecular field analysis (MFA) and receptor surface analysis (RSA). The cross-validated r2 (r2cv) values are 0.85 and 0.80 for MFA and RSA, respectively. The predictive ability of these models was validated by 16 test set molecules. The results of the best QSAR model were further compared with structure-based investigations using docking studies with the crystal structure of human non-pancreatic secretory phospholipase A2 (hnps-PLA2). Thereby providing new guidelines for the design of novel inhibitors. The QSAR studies made on the Indole-3-acetic acid derivatives employing good regression methods and controlled parameters from descriptor calculation, Molecular-Field Analysis, Receptor-Surface-Analysis; the predicted activities are falling in the same range to the experimental IC50 values for the molecules with good activities. Of the three methods used, RSA has generated most satisfactory results in predicting the activities. Structure-Based Drug-Designing studies were carried out for the analogues. The docking studies made on the Indole-3-acetic acid, which have been evaluated with QSAR proved satisfactory. The experimentally determined activity of Mol No.82 was showing high activity in Auto Dock result based on the docking energy. LigandFit gave the highest dock score of 69.374 for compound 82 with its bioactive conformation. The dock score obtained showed a good correlation with its experimental activity. Finally, these results provide useful information in understanding the structural and chemical features of Phospholipase A2 inhibitors and in designing New-potential compounds.

239: Structural Characterization and Refinement of NorA, a multidrug resistant efflux pump

[1]**Amit Nargotra**, [1]Sujata Sharma, [2]Bhupesh Taneja, [1]Surrinder Koul, [1]G. N. Qazi

[1]Indian Institute of Integrative Medicine, Canal Road Jammu, India, [2]Institute of Genomics and Integrative Biology, Mall Roal New Delhi, India

The rapid spread of bacteria expressing multidrug resistance (MDR) has necessitated the discovery of new antibacterials and resistance-modifying agents. Since the initial discovery of bacterial efflux pumps in the 1980s involving Gram+ve and Gram-ve and of late mycobacterium, the efflux pumps are able to extrude structurally diverse compounds, including antibiotics used and making these therapeutically ineffective. Efflux mechanisms have become broadly recognized as major components of resistance to many classes of antibiotics. It is therefore imperative that new antibiotics, resistance-modifying agents and, more specifically, efflux pumps are characterized. NorA is a Staphylococcus aureus multidrug efflux pump, the activity of which confers decreased susceptibility to many structurally unrelated agents, including fluoroquinolones, resulting in a multidrug resistant (MDR) phenotype. Even though this protein is significantly important in terms of drug resistance, the 3D structure of this protein is still not available. Therefore, efforts are being made by us to arrive at the 3D structure of this important protein, using in silico approach involving threading method (due to poor homology) of this protein with the existing 3D structures in the PDB. One of the approaches adopted recently by us was based on the alignment between the NorA sequence and that of glycerol-3-phosphate transporter (PDBID: 1PW4) of E. coli and then using MODELLER 8v0 [Nargotra et al. (2007) International Conference on Bioinformatics, Hong Kong Aug 2007, p. 71]. In this communication, we report a more robust 3D structure of NorA using more advanced program PHYRE-Phyre-Protein Homology/analogY Recognition Engine (that involves the use of PSIPRED secondary structure prediction method) [http://www.sbg.bio.ic.ac.uk/phyre]. The modified 3D structure of NorA confirmed the presence of 12 transmembrane helices interspersed by loops of variable lengths at regular intervals. Further the potential binding sites of modified NorA structure have been deciphered using Cerius2 software. Docking of known inhibitors with predicted binding sites revealed the presence of four important potential binding pockets in this new proposed structure of NorA. Further, docking studies have been carried out using potent NorA inhibitors from IIIM repository and from the results obtained we could decipher that in this new predicted structure, the corroboration with the practical results have improved significantly (up to 70%), which earlier was only 56%.

240: Recompilation of amino acid substitution matrices for sequence searches of a biased genome

Uma Devi Paila, R. Kondam, A. Ranjan

Computational and Functional Genomics Group, Centre for DNA Fingerprinting and Diagnostics, ECIL Road, Nacharam, Hyderabad-76, India

The genomic data uncovered in the recent years encodes many biological insights whose deciphering can be the basis for remarkable scientific success. For biased genomes, annotation is a serious challenge as is the case with the extremely AT-rich *Plasmodium falciparum* genome. This parasite is relatively distant to other eukaryotes with most of its encoded proteins lacking notable sequence similarity to other organisms. Since a nucleotide bias may lead to an overall bias in the amino acid composition of a protein, we were interested in examining the amino acid substitution in *P. falciparum*, which we later observe to be significantly different. To improve sequence searches of *P. falciparum*, we further aimed computing a new amino acid substitution matrix in the context of this genome. The idea was that the present approach would help resolve the enigma of inconsistent target and background frequencies, that is a matter of concern with the standard matrices used to compare non-standard genomes. Here we show that PfSSM (*Plasmodium falciparum* Specific Substitution Matrices) derived from a unique dataset of protein orthologs is more specific for *P. falciparum* related sequence searches. A performance evaluation of our new matrix with sequence alignments showed an improvement in the alignment score and length of ortholog protein pairs. The alignments extended across important motifs of *P. falciparum* proteins while some hypothetical proteins gave potential ortholog hits with database searches. This study has important implications on the annotation of proteins that are of experimental interest but give poor sequence alignments with standard matrices.

241: Structure determination of cysteine conjugate-beta lyase 1 (CCBL1) a key enzyme for nephro/neurotoxicity through computational approaches

Madhumita Pattanayak, Shailendra K. Gupta, Farhat N. Jaffery

Indian Institute of Toxicology Research, M.G. Marg, P.O. Box-80, Lucknow-226001, India

In agriculture, pesticides are a contributing factor towards the success of the green revolution. These pesticides are useful but in most cases they are having undesirable non-target effects and known to produce

adverse effect on the environment. It is indeed important to gather complete knowledge about them irrespective of their widespread usage. An attempt was taken for checking the adverse effects caused by the interaction of some pesticides in animals by using in silico methods. CCBL1 (cysteine conjugate-beta lyase 1) in Mus musculus is a cytosolic enzyme that is responsible for the metabolism of cysteine conjugates of certain halogenated alkenes and alkanes. This metabolism can form reactive metabolites leading to nephrotoxicity and neurotoxicity. CCBL1 catalyzes the irreversible transamination of the L-tryptophan metabolite L-kinurenine to form kynurenic acid (KA). It catalyzes the beta-elimination of S-conjugates and Se-conjugates of L- (seleno) cysteine, resulting in the cleavage of the C–S or C–Se bond. This enzyme has a central role in the bioactivation of nephrotoxic halogenated hydrocarbons. Increased levels of this enzyme have been linked to many nephro/neurotoxicity disorders. Fenvalerate, an organochlorine pesticide increases the activity of CCBL1 thus giving rise to nephrotoxicity and neurotoxicity whereas carbaryl and dimethoate are organophosphorus pesticides that decrease the enzyme's activities. Study was carried out for the in silico prediction of the 3D structure of the CCBL1 enzyme to identify its active sites. The verification of the structure was done by Ramachandran plot and dope score. Best structure will be taken for further study to identify the interacting points with the above pesticides to give a better knowledge of their mode of action.

242: Binding of gastrin-releasing peptide hormone to the membrane bilayers: investigation of preferred orientation using molecular dynamics studies

Priyanka Prakash, R. Sankararamakrishnan

Indian Institute of Technology, Kanpur, Department of Biological Sciences and Bioengineering, India

Gastrin-releasing peptide (GRP) belongs to the family of Bombesin-like peptides (BLPs) which are peptide hormones that interact with G-protein coupled bombesin receptors in order to carry out their physiological actions. BLPs are widely expressed in the central nervous system and gastrointestinal tract. GRP and its receptor, GRPR (Gastrin-releasing peptide receptor belonging to the family of bombesin receptors) have been reported to be involved in various types of cancers and obesity related problems. Therefore, understanding the mechanism of interaction of this peptide hormone with its receptor at molecular level is of great importance. According to 'Membrane compartments theory' [1], the flexible peptide hormones in random conformations first bind to the membrane before interacting with their receptors. The stable conformation induced by the membrane medium then binds to the receptor after two-dimensional diffusion. Several recent experimental studies have rendered support to this hypothesis. Thus the membrane lipid bilayer seems to play a crucial role in the ligand-receptor interactions. However, studying such membrane systems experimentally is a great challenge. Computational tools such as molecular dynamics (MD) simulations are of immense significance in the characterization of such systems where atomistic details are not directly achievable from the experiments.

The structure of one of the BLPs, neuromedin B (NMB), has been determined in a membrane-mimetic medium [2]. We have used the NMB structure to model the C-terminal decapeptide of GRP hormone. The peptide hormone can bind parallel or perpendicular to the membrane surface or in any orientation in between these two. In the current study we have investigated the preferred orientation of the GRP within the bilayer using MD simulations with multiple starting points. Five initial structures differed in the orientation of GRP ligand with respect to the membrane plane were considered. At the end of 10–20 ns simulations, all except one converged to the orientation that is parallel to the membrane surface. We have also found that this observation is independent of the force-field. Analysis of density profiles and solvation [3] of peptide side-chains are being analyzed to find out the factors responsible for the preferred orientation.

[1] Schwyzer, R., Biopolymers, 31, 785 (1991)
[2] Lee and Kim, FEBS Lett., 460, 263 (1999)
[3] Prakash and Sankararamakrishnan, Prot. Pept. Lett., 14, 590 (2007)

243: Functional correlation of cyclooxygenases-1, -2 and -3 from amino acid sequences

[1]**G. V. Reddy**, [1]M. Nagini, [1]P. Aparoy, [2]G. R. Hemalatha, [2]Lalitha Guruprasad, [1]P. Reddanna

[1]School of Life Sciences, University of Hyderabad, Gachi Bowli, Hyderabad, India, [2]School of Chemistry, University of Hyderabad, Gachi Bowli, Hyderabad, India

COX-1, COX-2, COX-3, three isoforms of cyclooxygenase, encoded by the PTGS1 (COX-1 and COX-3) and PTGS2 (COX-2) genes, are differentially expressed. We have analyzed the sequences of these cyclooxygenases and built the three-dimensional model structures for human COX-1, COX-2 and canine COX-3 to characterize the function of cyclooxygenase isozymes based on the sequence and structure information. Sequence analysis reveals that COX-3 shares 90% homology with COX-1 and 60% with COX-2. The COX-1 model has been compared with those of COX-2 and COX-3 and the active site regions have been analyzed. The major differences in the active sites of COX-1 and COX-2 are: Ile 523 in COX-1 is replaced by Val in COX-2, apart from a few mutations at the mouth of the active site. No such differences in the active sites are seen between COX-1 and COX-3 structures and the amino acid residues that differ between COX-1 and COX-3 lie only on the surface of the protein. Therefore, it is hard to explain the specificity of acetaminophen towards COX-3. Further, the intron 1 of canine PTGS1 gene that is retained in COX-3 mRNA transcript codes for a polyproline region that could be responsible for intermolecular interactions.

244: Structural analysis of factor IX variants for genotype phenotype correlation in hemophilia B Patients

[1]**Atreyee Saha**, [1]Saibal Mukherjee, [2]Parbati Biswas, [3]Chhabinath Mandal, [1]Kunal Ray

[1]Molecular and Human Genetics Division, Indian Institute of Chemical Biology (CSIR), Kolkata, India, [2]University of Delhi, New Delhi, India, [3]Structural Biology and Bioinformatics Division, Indian Institute of Chemical Biology (CSIR), Kolkata, India

Factor IX (FIX) is a component protein of blood coagulation pathway. Defective FIX protein resulting from mutation in the corresponding F9 gene causes Hemophilia B, an X-linked bleeding disorder. The disease manifests as mild, moderate or severe forms. The causal mutations for Hemophilia B are heterogeneous, spread over the entire gene. The level of biological activity and the amount of mutant protein are measured by a coagulation assay and by estimation of antigen cross-reacting material, respectively. One of the challenges of human genetics is predicting the functional implication of mutations that might lead to pathogenesis. Nonsynonymous variations are most common among all nucleotide changes in the coding sequence, which could be highly pathogenic or completely innocuous—presented as a

polymorphism or a rare variant. In case of Hemophilia B, one could create a recombinant clone containing the suspect nucleotide variant in the FIX gene and examine the biological activity of the mutant FIX protein by its expression in suitable cells. In this context, we attempted to reach similar conclusion by mapping the mutations on the crystal structure of FIX, and comparing the alteration of the structural parameters relative to those for the normal protein. Therefore, we examined the effect of point mutations on FIX by measuring hydrogen bonding pattern, solvent accessibility and electrostatic potential. Out of a total of 16 severe mutations 14 (88%) showed changes of hydrogen bonding pattern to variable extent. Among the 9 mild Hemophilia B mutations, 6 (i.e. 66.66%) showed no change in hydrogen bonding pattern. Our data suggest that there is a statistically significant correlation between the two groups of mutations as measured by change in the H-bonding pattern. Our study truly represents an initiation of an effort that would provide a framework for first evaluation of suspected mutations by in silico approaches, which could be further validated by other experimental techniques. This study represents the beginning of investigation on structural perturbation as a measure to assess genotype–phenotype correlation in case of Hemophilia B. This study is supported by funds from CSIR, Govt. of India.

245: Mobility in *Mycobacterium tuberculosis* ribosome recycling factor

[1]**M. Selvaraj**, [1]Alok Sharma, [2]Anuradha Seshadri, [2]Umesh Varshney, [1]M. Vijayan

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560 012, India, [2]Microbiology and Cell Biology Laboratory, Indian Institute of Science, Bangalore-560 012, India

We have been pursuing a long range programme involving structural genomics of TB proteins, one element of which is concerned with protein synthesis. As part of this programme, we have elucidated the molecular plasticity of the ribosome recycling factor (RRF) primarily through the X-ray analysis of the protein from *M. tuberculosis* (*Mt*RRF). The plasticity of the molecule involves the bending and twisting motion of the two arms (domains) of the L-shaped molecule and the rotatory motion of the smaller domain. Two potential salt bridges are involved in interactions between the two domains in *Mt*RRF. The mobility of the molecule, which is believed to be important for its function, is likely to be influenced by these salt bridges. In order to explore this influence, structural studies of a mutant which abolishes one of the salt bridges and molecular dynamics simulations on the wild type molecule and the mutant, are being carried out. The mutant structure determined now and the wild type structures studied earlier, appear to indicate that the abolition of one particular salt bridge leads to the strengthening of the other in a partially compensatory manner. However, the mutation results in a slight opening of the L-shaped molecule. MD simulation enables the elaboration of this observation. Further analysis of the structures and the trajectories of movement indicated by simulations are in progress. Additional results resulting from this analysis will be presented.

246: Development of a new SVM-based method for protein fold recognition

Mohammad Tabrez Anwar Shamim, Hampapathalu Adimurthy Nagarajaram

Laboratory of Computational Biology, Centre for DNA Fingerprinting and Diagnostics (CDFD), ECIL X Road, Nacharam, Hyderabad 500 076, India

Fold recognition is an important problem in molecular biology. Although many methods have been developed for protein fold recognition, their accuracies remain low, and can be attributed to the less exploitation of discriminatory features. In this study, we have investigated the discriminatory potential of the secondary structural and solvent accessibility state information of amino acid residues and residue pairs for protein fold recognition. Our studies have revealed that the secondary structural and solvent accessibility state frequencies of amino acids and amino acid pairs collectively give rise to the best fold-discrimination. Evaluation of our method using benchmark dataset yielded an accuracy of more than 70% (Shamim et. al. 2007), which is $\sim$8% higher than the best available method. Our studies have also revealed that the three multi-class classification methods used, namely one versus all, one versus one, and Crammer and Singer method, yield similar predictions. Furthermore, we have increased the coverage of our method by including more SCOP folds to the list of folds used in our earlier publication (Shamim et. al. 2007). Since our SVM-based method outperforms other available methods, we are using it for fold-wise classification of unknown proteins discovered in mycobacterial genomes.

* Shamim et. al., 2007, Bioinformatics, 23(24), 3320–27.

247: Computational methods of protein network identification in *Acinetobacter baylyi*

Shailza Singh, B. A. Chopade

Institute of Bioinformatics and Biotechnology, University of Pune, Pune, India

Genome-wide functional linkages among proteins in cellular complexes and metabolic pathways can be inferred from high throughput experimentation, such as DNA microarrays, or from bioinformatic analyzes. The genome of *Acinetobacter baylyi* was analyzed using recently developed computational approaches to infer protein function and protein linkages. As judged by the nucleotide distance between genes in the same genomic orientation and combined this method with those of the Rosetta Stone, Phylogenetic Profile and Conserved Gene Neighbor computational methods for the inference of protein function a method was evaluated to infer genes likely to belong to the same operon. This method involves the construction of a genome-wide functional linkage map where each significant functional linkage between a pair of proteins is displayed on a two-dimensional scatterplot and organized according to the order of genes along the chromosome. Subsequent hierarchical clustering of the map reveals clusters of genes with similar functional linkage profiles. This facilitates the inference of protein function and the discovery of functionally linked gene clusters throughout the genome. Cellular functions were assigned to previously uncharacterized proteins involved in cell wall biosynthesis, signal transduction, chaperone activity, energy metabolism and polysaccharide biosynthesis.

248: Substrate, positional and stereospecificity of 12R-LOX : positional and functional correlation studies

[1]**Agarwal Smita**, [1]Meruvu Sunitha, [2]Kuhn Hartmun, [1]Reddanna Pallu

[1]School of Life Sciences, University of Hyderabad, Gachibowli, Hyderabad, India, [2]University of Berlin Charite, Monbijoustrasse 2, 10117 Berlin, Germany

Lipoxygenases (LOX) represent a widespread family of non heme, non sulphur, iron containing dioxygenases that catalyze the

regioselective and stereo selective dioxygenation of fatty acid substrates containing one or more (Z, Z)-1,4-pentadiene moieties. Mammalian lipoxygenases (LOXs) are categorized with respect to their positional specificity of arachidonic acid oxygenation, 12R-LOX representing the first mammalian LOX isoenzymes with pure R-chirality. The substrate specificity of murine 12R-LOX was determined by involving site directed mutagenetic studies. To search for sequence determinants of murine (12R)-LOX and human 12(R)-LOX, we carried out multiple amino acid sequence alignments and found that Phe390, Gly441, Ala455, and Val631 align with previously identified positional determinants of S-LOX isoforms. Multiple site-directed mutagenesis studies on Phe390 and Ala455 did not induce specific alterations in the reaction specificity, but yielded enzyme species with reduced specific activities and stereo random product patterns. Mutation of Gly441 to Ala, which caused drastic alterations in the reaction specificity of other LOX isoforms, failed to induce major alterations in the positional specificity of mouse (12R)-LOX, but markedly modified the enantioselectivity of the enzyme. This result obtained in murine 12(R)-LOX is expected to be consistent with human 12(R)-LOX due to the high sequence identity at that specific domain. When Val631, which aligns with the positional determinant Ile593 of rabbit 15-LOX, was mutated to a less space-filling residue (Ala or Gly), we obtained an enzyme species with augmented catalytic activity and specifically altered reaction characteristics (major formation of chiral (11R)-hydroxyeicosatetraenoic acid methyl ester). The importance of Val631 for the stereo control of murine (12R)-LOX was confirmed with other substrates such as methyl linoleate and 20-hydroxyeicosatetraenoic acid methyl ester. These data identify Val631 as the major sequence determinant for the specificity of murine (12R)-LOX. Furthermore, we conclude that substrate fatty acids may adopt different catalytically productive arrangements at the active site of murine (12R)-LOX and that each of these arrangements may lead to the formation of chiral oxygenation products.