3354-Pos Board B401
Development of Cardiac Integrated Database Management System (CIDMS)
Pradeep Kumar Sreenivasaiah, Bharathi Mohanraju, Do Han Kim.
GIST, Gwangju, Republic of Korea.
A major emphasis in systems biology is to unfold the dynamic network properties of complex biological processes through modeling and simulation utilizing various experimental data. CIDMS was developed to provide an integrated bioinformatics platform for systems biology approach studying mammalian cardiac systems. CIDMS has following information which was manually curated from published literatures and public databases: 1) list of genes, proteins, pathways critical for development and functioning of heart; 2) list of human cardiac diseases; and 3) mathematical / computational models employed to study cardiac biological processes and diseases. These lists are supplemented with key annotations like function, regulatory mechanisms, interactions and references to the source of information in context of cardiology. CIDMS also provides a large amount of additional automated data integrated within the framework of manually curated data. Data such as structure, function, domains, location, phenotype, molecular interactions, and pathways from numerous life-science databases are made available. In addition to these qualitative data, CIDMS integrates quantitative data such as reaction kinetics, stoichiometric and model information. This broad data integration has become possible by the state of the art technologies like data-warehouse and through webservices (SOAP interfaces) of well established public database servers ensuring up to the minute information. The comprehensive information can be easily searched and browsed by categories and ontologies with an advanced web interface hosted at http://cidms.org.

3355-Pos Board B402
Towards Unifying Systems Biology - Using Pathway Data in Biopax Format for SBML Simulators
Oliver Ruebenacker, Ion I. Moraru, **Michael L. Blinov**.
University of Connecticut Health Center, Farmington, CT, USA.
Thousands of biochemical interactions are available from public sources in the Biological Pathways Exchange (BioPAX) format. However, the current standard for exchange of simulation-ready biological models is System Biology Markup Language (SBML). This markup language is structurally and semantically different from BioPAX. Some conversion schemes exist, using annotations and based on simple one-to-one mappings between SBML and BioPAX objects, which ignores semantic differences and therefore often leads to significant loss of information or meaning. A comprehensive modeling framework capable of representing the complex relationships between SBML and BioPAX data is needed to take full advantage of existing pathway data in kinetic modeling, thus integrating these two formats by gluing them together.
Here we describe such a framework that we are developing as a part of the Virtual Cell (http://vcell.org) modeling and simulation environment. Systems Biology Linker (SyBiL, http://vcell.org/biopax) is a tool for analyzing and visualizing BioPAX data, and converting them to SBML. Based on the Jena Semantic Web framework for Java, SyBiL supports handling of generic RDF/OWL data (such as visualization and reasoning) as well as functions specific to handling SBML and BioPAX data. SyBiL uses Systems Biology Pathway eXchange, called SBPAX, as a generic approach to integrate model-centric formats similar to SBML with pathway-centric formats similar to BioPAX. SBPAX is an OWL-based schema that serves as a glue to integrate different data formats, despite semantic differences. Effectively, SBPAX provides a bridge between SBML and the Semantic Web world. SyBiL offers various visualization modes showing reaction networks to varying degrees of details, including displaying nodes for reactions only as well as displaying Petri nets consisting of reaction nodes and reaction participants and catalysts.

3356-Pos Board B403
Prediction of Functional Sites in 50,000 Protein Domains Using Dynamics Perturbation Analysis
Judith D. Cohn, Dengming Ming, **Michael E. Wall**.
Los Alamos National Laboratory, Los Alamos, NM, USA.
Dynamics perturbation analysis (DPA) [1-3] finds regions in a protein structure where proteins are ''ticklish,'' *i.e.*, where interactions cause a large change in protein dynamics. Such regions corresponded to native binding sites in a test set of hundreds of protein-ligand complexes [3]. We have further tested DPA for prediction of protein functional sites by applying an accelerated algorithm, Fast DPA [4], to over 50,000 structures of protein domains from the SCOP database [5]. We compared the predicted sites with known functional sites obtained from two sources: catalytic residues in the Catalytic Site Atlas; and putative binding sites identified by finding protein residues near small molecules in crystal structures. Combining predictions with information from sequence conservation and multiple sequence alignments reduced false positive rates. Overall, prediction of functional sites using DPA recapitulates much of the known information about functional sites in SCOP domains, and validates the use of DPA to predict functional sites in proteins. These results further suggest that functional sites in proteins tend to evolve at control points where interactions cause a large change in protein dynamics [1, 3].
1. Ming D. and Wall M.E. Quantifying allosteric effects in proteins. 2005. Proteins 59:697-707.
2. Ming D. and Wall M.E. Allostery in a coarse-grained model of protein dynamics. 2005. Phys Rev Lett 95:198301.
3. Ming D. and Wall M.E. Interactions in native binding sites cause a large change in protein dynamics. 2006. J Mol Biol 358:213-223.
4. Ming D., Cohn J.D. and Wall M.E. Fast dynamics perturbation analysis for prediction of functional sites. 2008. BMC Struct Biol 8:5.
5. Murzin A.G., Brenner S.E., Hubbard T. and Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. 1995. J Mol Biol 247:536-540.

3357-Pos Board B404
From Abstract Graphs To (biophysical) Reality - Graphs In Mutual Information Space Help Identify Highly Correlated Positions In G Protein-coupled Receptors
Sarosh N. Fatakia, Stefano Costanzi, Carson C. Chow, Vipul Periwal.
NIDDK, NIH, Bethesda,, MD, USA.
G protein-coupled receptors (GPCRs) are a superfamily of seven transmembrane-spanning proteins involved in a wide array of physiological functions and are the most common targets of pharmaceuticals. Are there idiomatic relations between positions on these protein sequences that are evident in their co-evolution? We investigate the semantic graph of the constituent amino acid (AA) positions using an information theoretic approach. Using a multiple sequence alignment of the seven transmembrane (7-TM) domains, we calculated the mutual information (MI) between all pairs of aligned positions. Representing TM positions as vertices and pairing them by their MI we compute the planar acyclic graph that maximizes the total MI. The total MI of this graph is much greater than the total MI of a random planar acyclic graph. From this graph, we identify few positions which have a significantly high degree (edges incident to the vertices) when contrasted with the others. The positions, from class A and class C GPCRs, with the leading degree values are found to be associated with the experimentally determined binding pocket, confirming our previous studies involving MI graphs.

3358-Pos Board B405
Local surface shape-based protein function prediction using Zernike descriptors
Daisuke Kihara[1], Lee Sael[1], Rayan Chikhi[2].
[1]Purdue University, West Lafayette, IN, USA, [2]Ecole Normale Superieure de Cachan, Britanny, France.
Structural genomics projects have been solving an increasing number of protein structures with unknown function. As the number of uncharacterized protein structures continues to grow, there is increasing need for effective computational strategies for structure-based protein function prediction. As of now, there are more than 2400 protein structures in the PDB database, which are categorized as unknown function and thus are awaiting functional characterization. Previously, we have reported a fast global protein surface shape comparison method based on 3D Zernike descriptors. 3D Zernike descriptors are series expansions of a given three-dimensional function that compactly represent a protein surface and its corresponding physicochemical properties. Consequently, it takes approximately a minute to compare a query structure against thousands of protein structures in a database.
Here, we apply the method for local protein surface shape comparison, focusing on surface pockets, which are potential ligand molecules binding sites. Applications of the methods presented in this work include classification of ligand binding sites of TIM barrel proteins according to their surface electrostatic potential and classification of protein local surface regions in terms of the similarity of shape and physicochemical properties. We also show that prediction of binding ligand molecule by pocket similarity search. Taken together, we demonstrate the high throughput applicability of Zernike descriptors in recognizing local physicochemical of protein surfaces. This work was supported in part by NIH (R01 GM075004).

3359-Pos Board B406
Large Scale Prediction of Computational Hot Spots in Protein Interfaces
Nurcan Tuncbag[1], **Ozlem Keskin**[1], Ruth Nussinov[2], Attila Gursoy[1].
[1]Koc University, Istanbul, Turkey, [2]SAIC-NCI, Frederick, MD, USA.
Hot spots are residues comprising only a small fraction of interfaces yet accounting for the majority of the binding energy. These residues are critical in

understanding the principles of protein interactions. Experimental studies like alanine scanning mutagenesis require significant effort; therefore, there is a need for computational methods to predict hot spots in protein interfaces. We present a new efficient method to determine computational hot spots based on sequence conservation and solvent accessibility of the interface residues (Tuncbag et al.; Guney et al.). The predicted hot spots are observed to correlate with the experimental hot spots with an accuracy of 71% and a positive predictive value of 79%. Several machine learning methods (SVM, Decision Trees and Decision Lists) are also applied to predict hot spots and compared to our method. The results reveal that our empirical approach performs better. We observe that both the change in accessible surface area upon complexation and residue accessibility in the complex forms improve detection of hot spots. Predicted computational hot spots for all protein interfaces (49512 interfaces as of 2006) are available at HotSprint database. HotSprint (a database of computational hot spots in protein interfaces) can be accessed at http://prism.ccbb.ku.edu.tr/hotsprint.

Guney E, Tuncbag N, Keskin O, Gursoy A: HotSprint: database of computational hot spots in protein interfaces Nucleic Acids Res 2008, 36(Database issue):D662-666.

Tuncbag N, Gursoy A, Guney E, Nussinov R, Keskin O: Architectures and Functional Coverage of Protein-Protein Interfaces J Mol Biol 2008.

3360-Pos Board B407

Ancestral Sequence Reconstruction and Homology Modeling Link Temperature Adaptation and Conservation of Function With Sequence Evolution in Parvalbumin

A. Carl Whittington[1], Timothy S. Moerland[2].

[1]Florida State University, Tallahassee, FL, USA, [2]Kent State University, Kent, OH, USA.

Temperature is a key factor influencing protein structure and function in poikilotherms. Previous studies of enzymes have shown that orthologs from species acclimated to different thermal niches can maintain a relatively similar level of function at those species' respective physiological temperatures. In some well-characterized enzymes, this conservation of function is correlated with differences in primary structure that lie outside active sites. Information gained from thermal adaptation studies of enzymes can be extended to non-catalytic proteins, which are less thoroughly examined, through the study of parvalbumin structure and function. Parvalbumins are intracellular calcium-binding proteins of the EF-hand type that are thought to function in muscle cells as calcium sinks permitting more rapid unloading of troponin-C, leading to more rapid contraction/relaxation cycles. Parvalbumins contain two functional, highly conserved binding sites and one non-functional site, the AB domain, thought to be an important area of modulation for cation binding. Parvalbumins from teleosts of the sub-order Notothenioid and the unrelated Arctic cod, Boreogadus saida, have converged on a common phenotype. That is, they show similar thermal sensitivity patterns of calcium binding. To explore the underlying structural basis of this similarity in phenotype, we have used ancestral sequence reconstruction combined with homology modeling to identify potential changes in primary structure that have allowed parvalbumins from these disparately related groups of fish to function similarly in their polar habitats. For instance, an Asn to Cys change at position 26 (located in the AB domain) in the evolution of B. saida parvalbumin may lead to the loss of a hydrogen bond in the B-helix. This may provide the change in tertiary structure needed for this parvalbumin to function at polar temperatures.

3361-Pos Board B408

Evolutionary Analyses Of KCNQ1 And HERG Voltage-gated Potassium Channel Sequences Reveal Location-specific Susceptibility And Augmented Chemical Severities Of Arrhythmogenic Mutations

Heather Jackson, **Eric Accili**.

UBC, Vancouver, BC, Canada.

Mutations in HERG and KCNQ1 potassium channels have been associated with Long QT syndrome and atrial fibrillation, and more recently with sudden infant death syndrome and sudden unexplained death. In other proteins, disease-associated amino acid mutations have been analyzed according to the chemical severity of the changes and the locations of the altered amino acids according to their conservation over metazoan evolution. Here, we present the first such analysis of arrhythmia-associated mutations (AAMs) in the HERG and KCNQ1 potassium channels. Using evolutionary analyses, AAMs in HERG and KCNQ1 were preferentially found at evolutionarily conserved sites and unevenly distributed among functionally conserved domains. Non-synonymous single nucleotide polymorphisms (nsSNPs) are under-represented at evolutionarily conserved sites in HERG, but distribute randomly in KCNQ1. AAMs are chemically more severe, according to Grantham's Scale, than changes observed in evolution and their severity correlates with the expected chemical severity of the involved codon. Expected chemical severity of a given amino acid also correlates with its relative contribution to arrhythmias. At evolutionarily variable sites, the chemical severity of the changes is also correlated with the expected chemical severity of the involved codon. Unlike nsSNPs, AAMs preferentially locate to evolutionarily conserved, and functionally important, sites and regions within HERG and KCNQ1, and are chemically more severe than changes which occur in evolution. Expected chemical severity may contribute to the overrepresentation of certain residues in AAMs, as well as to evolutionary change.

3362-Pos Board B409

Evolution of Mammalian HCN Channels - Positive Darwinian Selection Identified at the Molecular Level

Andrew Roth, Heather Jackson, Eric Accili.

UBC, Vancouver, BC, Canada.

HCN channels are important for regulating spontaneous electrical activity and membrane potential in excitable cells. We hypothesize that the four mammalian HCN genes were established by duplication after the divergence of urochordates and before the divergence of fish from the tetrapod lineage. A question in which we are interested is how did the differences in structure and function among the four mammalian channels arise? These differences are due to changes in primary sequence that have occurred since duplication. At the molecular level, changes in DNA sequence over the course of evolution may be identified based upon the Neutral Theory of Molecular Evolution, which states that the majority of changes in DNA are neutral. Neutrality can be estimated directly by comparing the number of non-synonymous changes (dN) and synonymous changes (dS) at the DNA level over a given period of evolutionary time. Neutrality implies that dN and dS are equal (dN/dS=1). A ratio less than one implies purifying selection, whereas a ratio of more than one implies positive selection. Here, we use phylogenetic and statistical analyses of mammalian HCN sequences to identify positive selection among the four mammalian HCN isoforms. We find that the HCN2 isoform yields a very high value for dN/dS ($>>1$), with strong statistical support. Further analysis of HCN2 uncovered a number of specific sites that have undergone positive selection, including an unusual triplet of residues in the outer pore. Our results suggest that positive selection has contributed to differences in structure and function between HCN2 and the other three isoforms.

3363-Pos Board B410

Genomic Identification of Transmembrane β-Barrels (TMBBs)

Thomas Freeman, William C. Wimley.

Tulane University, New Orleans, LA, USA.

Transmembrane beta-barrels (TMBBs) are a special structural class of proteins predominately found in the outer membranes of Gram-negative bacteria, mitochondria, and chloroplasts. It is estimated that 2-3% of a bacterial genome encodes TMBBs, yet less than 40 non-redundant structures have been solved. It would be highly advantageous to have methods to rapidly identify TMBBs from increasingly available genomic databases. A prediction algorithm proposed by Wimley in 2002 was based on the physicochemical properties of TMBBs of known structure. This method used relative amino acid abundances to predict the position of beta-strands and beta-hairpins, which are the major structural subunits of TMBBs, and a mathematical simplification of the topology prediction data called a beta-barrel score. To test the accuracy of this algorithm we scored proteins from a non-redundant database of protein sequences from the Protein Data Bank (NRPDB). The results revealed that the algorithm's ability to discriminate true TMBBs from other proteins, while strong, could be significantly improved. First, we updated the relative amino acid abundances to include the latest structural information. Second, we altered the beta-strand prediction method to account for the fact that certain amino acids have a higher propensity to situate near the lipid/water interface than in the hydrophobic core of the bilayer. Third, we adjusted the calculation of the beta-barrel score to address the lowered beta-hairpin density of larger TMBBs such as BtuB. We reanalyzed the NRPDB and the modifications resulted in a 5-fold decrease in the number of false positives, many of which are either non-bacterial proteins or from Gram-positive organisms. We will use this method to analyze the available genomes of Gram-negative bacteria and the results, along with the signal peptide predictions of SignalP (Bendtsen, et al. 2004) will be deposited into a publicly available database.

3364-Pos Board B411

Predicting Binding Sites of EH1-like Motifs from Their Amino Acid Sequences

D.S. Dalafave.

The College of New Jersey, Ewing, NJ, USA.

Bioinformatics and computer modeling were used to predict binding sites of engrailed homology-1 (eh1) -like motifs from their amino acid sequences. According to previous studies, an eh1 motif provides its transcriptional function