the multichromophoric nanoprobe system result in extended photobleaching lifetimes, enhanced brightness and improved localization precision compared to single organic dyes. In this way, we engineered PDNs as enhanced fluorescent probes for super-resolution imaging. PDNs are bright, photostable and nanometer-sized macromolecular probes that can be synthesized to contain a variety of surface chemical functionalities for biological labeling and imaging. Based on these advantages, PDN probes hold strong promise to find pervasive applications as a new class of probes for fluorescence microscopy.

913-Pos Board B699

Analysis of Protein Complex Formation on Membrane Surfaces by Single Molecule Diffusion

Brian P. Ziemba, Jefferson D. Knight, Joseph J. Falke.
University of Colorado, Boulder, CO, USA.

Many signaling pathways are controlled by protein complexes assembled on membrane surfaces via collisions between diffusing membrane components. Subsequently, the active, membrane-bound complex often collides with other effector proteins or lipids on the membrane surface to transmit its essential downstream signal. Thus, 2D-diffusion of membrane-bound proteins, complexes and effectors plays a central role in signaling biology. We have proposed that single molecule analysis of membrane protein diffusion on supported lipid bilayers could provide a powerful approach for studies of complex formation between membrane-associated proteins (Knight, Lerner, Marcano-Velazquez, Pastor & Falke (2010) Biophys J 99:2879-87). In principle the approach could yield information about complex stoichiometry, stability, and kinetics, and about specific lipid requirements for complex formation. Here we test these ideas using a simple model system in which membrane-bound pleckstrin homology (PH) domains are forced to dimerize by the calcium-triggered association of calmodulin (CaM) and its target peptide from skeletal muscle light chain kinase (MLCKp). Two fusion protein constructs (CaM-PH) and (MLCKp-PH) have been generated and bind normally to the PH domain target lipid, PIP3. Single molecule TIRF analysis of CaM-PH diffusion reveals that in the presence of calcium and MLCKp-PH a stable CaM-PH/MLCKp-PH heterodimer is formed on the membrane surface that diffuses approximately 2-fold more slowly than the single CaM-PH molecule. Addition of excess EDTA to chelate calcium reverses the dimer back to monomers. Substoichiometric levels of MLCKp-PH yield a mixed population of monomers and dimers, and the two populations can be resolved by single molecule analysis. Overall, the findings indicate that single molecule diffusion provides a useful new window into many critical, molecular features of protein complex formation on membrane surfaces. [Supported by NIH grant GM063235.]

914-Pos Board B700

On the Hydration of the Phosphocholine Headgroup in Solution

Sylvia McLain[1], Chris Lorenz[2], Aleksandra Dabrowska[2], Jayne Lawrence[2].
[1]University of Oxford, Oxford, United Kingdom, [2]King's College London, London, United Kingdom.

The hydration of the phosphocholine headgroup in 1,2-dipropionyl-sn-glycero-3-phosphocholine in solution has been determined by using neutron diffraction enhanced with isotopic substitution in combination with computer simulation techniques. The atomic scale hydration structure around this head group shows that both the -N(CH(3))(3) and -CH(2) portions of the choline headgroup are strongly associated with water, through a unique hydrogen bonding regime, where specifically a hydrogen bond from the C-H group to water and a strong association between the water oxygen and N(+) atom in solution have both been observed. In addition, both PO(4) oxygens (P=O) and C=O oxygens are oversaturated when compared to bulk water in that the average number of hydrogen bonds from water to both X=O oxygens is about 2.5 for each group. That water binds strongly to the glycerol groups and is suggestive that water may bind to these groups when phosophotidylcholine is embedded in a membrane bilayer.

Computational Biology

915-Pos Board B701

IPAVS: Integrated Pathway Resources, Analysis, and Visualization System

Pradeep Kumar Sreenivasaiah, Shilpa Rani, Joseph Cayetano, Novino Arul, Do Han Kim.
Gwangju Institute of Science and Technology, Gwangju, Korea, Republic of.

IPAVS is an integrated knowledgebase that offers biologists a single point of access to several manually-curated pathway resources along with its own expert-curated pathways. IPAVS currently integrates over 500 human pathways (consisting of 25,000 interactions) and mainly includes metabolic pathways, signaling pathways, disease pathways and several large processes maps curated using CellDesinger software. Systems Biology Graphical Notations (SBGN) and KEGG pathway notations are used for the visual display of pathway information. IPAVS also provides a web-based interface that: 1) allows biologists to browse and search pathway resources; 2) provides tools for data import, management (filter, group), and comparisons; 3) offers intuitive visualization and variety of enrichment analysis methods to support the interpretation of biological data in light of cellular processes; and 4) offers basic features for community participation and collaboration. IPAVS can export pathways into variety of standard data formats such as SBML, BioPax and XGMML that can be readily imported and worked with modeling and simulation, or network analysis and visualization software tools. IPAVS enables biologists to do systems biology and employ systems methodologies in their day to day research and produce testable hypothesis. IPAVS is freely accessible at http://ipavs.cidms.org.(Supported by Korea MEST NRF grant (2011-0002144), the 2011 GIST Systems Biology Infrastructure Establishment Grant and KISTI-KREONET).

916-Pos Board B702

Pathway Commons at Virtual Cell: Use of Pathway Data For Mathematical Modeling

Michael L. Blinov, Oliver Ruebenacker, James C. Schaff, Leslie M. Loew, **Ion I. Moraru**.
University of Connecticut Health Center, Farmington, CT, USA.

Mathematical models of biological systems are rapidly proliferating and many are now publicly available. However, their distribution and sharing is limited because in the absence of extensive accompanying documentation, there is a significant barrier for understanding and reuse. The Virtual Cell (VCell) is a modeling and simulation framework designed for building of a wide range of mathematical models; more than 3,000 users created more than 29,000 models, of which more than 500 were made public in the VCell database. There is a strong need for methods and tools that enable users to easily create well-annotated models. On the other hand, a large number of public databases provide huge collections of well-annotated physical entities and molecular interactions. Most provide data in RDF-based ontology formats (such as BioPAX) which typically cannot be converted to XML-based modeling formats without loss of information. To resolve this issue, we introduced two new concepts (encoded in VCML), a Pathway Model and a Relationship Model, which allow unambiguous storing of data extracted from pathway databases and the mapping of pathway elements to model entities. We used this technology to implement a new interface that makes available the Pathway Commons collection of databases as a resource that every VCell modeler will have at his/her disposal. Using a customized query interface, one can search pathway databases from inside VCell, import and store the entries of interest alongside a mathematical model, and create a new model based on the pathway data or link existing elements of the model to pathway data entities. This capability can help researchers to make annotated models that are easier to share with the community, as well as quickly create first iterations of a quantitative model from a wealth of well-annotated biological data.

917-Pos Board B703

Partial Reconstruction of Biological Networks from Perturbation Experiments

Maxim Artyomov.
Broad Institute, Cambridge, MA, USA.

Screens monitoring the effect of deletion, knock-down or over expression of regulatory genes on their target gene expression are critical for deciphering the organization of complex regulatory networks. However, since perturbation assays cannot distinguish direct from indirect effects, the derived networks are significantly more complex than the true underlying one. Previous approaches to identify a minimal network topology consistent with the results of a perturbation screen only presented approximate methods with major limitations and are often applicable only to simple network topologies. We present an approach to systematically find a family of core networks for an input network of any topology with an arbitrary number of activating and inhibiting interactions. Based on this approach we introduce distance measure between two different network topologies and propose method to address evaluate underlying network structure based on the perturbation experiments.

918-Pos Board B704

A Parallel Algorithm for Reverse Engineering Gene Regulatory Networks

Jason Bazil, Feng Qi, Daniel A. Beard.
Medical College of Wisconsin, Milwaukee, WI, USA.

Dynamic biological systems, such as gene regulatory networks (GRNs) and protein signaling networks, are often represented as systems of ordinary differential equations. Such equations can be utilized in reverse engineering these biological networks, specifically since identifying these networks is challenging due to the cost of the necessary experiments growing with at least the square of the size of the system. Moreover, the number of possible models, proportional to the number of directed graphs connecting nodes representing the variables in the system, suffers from combinatorial explosion as the size of the system grows. Therefore, exhaustive searches for systems of nontrivial complexity are not feasible. Here

we describe a practical and scalable algorithm for determining candidate network interactions based on decomposing an *N*-dimensional system into *N* one-dimensional problems. The algorithm was tested on *in silico* networks based on known biological GRNs. The computational complexity of the network identification is shown to increase as N^2 while a parallel implementation achieves essentially linear speedup with the increasing number of processing cores. For each *in silico* network tested, the algorithm successfully predicts a candidate network that reproduces the network dynamics. This approach dramatically reduces the computational demand required for reverse engineering GRNs and produces a wealth of exploitable information in the process. Moreover, the candidate network topologies returned by the algorithm can be used to design future experiments aimed at gathering informative data capable of further resolving the true network topology.

919-Pos Board B705

Multiple Co-Evolutionary Networks have Evolved on the Common Tertiary Scaffold of the LacI/GalR Proteins

Liskin Swint-Kruse, Daniel J. Parente.
KU Medical Center, Kansas City, KS, USA.

The number of known protein sequences is growing at a phenomenal rate. A first step in understanding the functional implications of novel sequences is to extract information from characterized, homologous proteins via sequence/function analyses. To that end, many new algorithms have been recently developed. Of these, co-evolutionary analysis has emerged as a popular class of methods. These methods identify pairs of positions in a multiple sequence alignment that vary in a coordinated manner. From these calculations, a network of co-evolving positions can be assembled. Strongly coevolving positions are thought to participate in the same functional and/or structural role.

Generally, co-evolution analyses are carried out on protein families with >40% sequence identity. One unresolved question is the extent to which co-evolutionary networks are conserved on a common tertiary architecture. To investigate this question, we are comparing the co-evolutionary networks of various subfamilies in the LacI/GalR transcription repressor family. Each subfamily has >40% sequence identity. In contrast, identity between subfamilies is usually 15-25%. Nevertheless, available structures show that subfamilies have very similar tertiary structures. Co-evolution networks have been generated using five different algorithms. As previously reported, the algorithm outputs differ from each other (e.g., McBASC and OMES identify different co-evolving positions); however, all algorithms find different patterns for each LacI/GalR subfamily. Additional analyses show that (i) each subfamily has a different set of conserved positions; and (ii) positions that are conserved in one subfamily are highly varied in another. Thus, the common tertiary architecture of the LacI/GalR homologs accommodates a variety of functional and/or structural networks that is distinct for divergent subfamilies.

920-Pos Board B706

An Improved Meta-Analysis Procedure Reveals Key Transcriptome Signatures Underlying a Renal Damage Phenotype

Yu Bai, Jennifer Schmahl.
Regeneron, Tarrytown, NY, USA.

Integration of transcriptome profiling data via meta-analysis can significantly leverage statistic power while reducing false positives and thus is widely applied in genomic research. Various meta-analysis procedures, ranging from Venn diagram to score-based gene selections, have been employed to uncover key genes underlying a phenotype or disease. Scores are often adopted to characterize genes beyond the simple knowledge of being perturbed or not and permit statistical control of false positives. Nevertheless, many scoring methods use significance (e.g. p-value, q-value) to characterize the gene perturbation instead of more biology relevant measures such as fold change. Recent work showed the advantage of fold change-ranked scores. However, the rank of genes can vary significantly for small alterations in fold change and therefore noise-sensitive. Herein we developed a new score combining fold change and the occurrence counts to select genes that are most commonly and strongly perturbed, with false discovery rate (FDR) estimated by permutation. The method was evaluated in comparison to previous rank or occurrence scores using synthetic data, where multiple datasets were simulated with predefined meta-genes (noise introduced so not all meta-genes are perturbed in each dataset) in the background of a varying number of other randomly perturbed genes. Our method outcompeted previous occurrence score with higher sensitivity, and the rank score with both higher sensitivity and specificity. We applied the method to dissect a renal damage phenotype observed in a set of KO mice with unknown mechanism, yielding 247 meta-genes (FDR<=0.1) enriched in renal and urological function (pvalue<1e-4, Ingenuity) and significantly correlated with a Cisplatin-induced acute renal failure model (pvalue<1e-35, NextBio). These insightful meta-signatures were not achieved by other methods, highlighting the power of our new method in facilitating deeper understandings of the molecular mechanism underlying disease phenotypes.

921-Pos Board B707

In Silico Reconstruction of HIV Viral Fitness Landscapes

Andrew L. Ferguson, Arup K. Chakraborty.
MIT, Cambridge, MA, USA.

Since its discovery in the early 1980's, HIV/AIDS has claimed over 30 million lives worldwide, with another 33.3 million infected today and no effective vaccine yet available. Contemporary vaccine design efforts adopt a new paradigm to counter the extreme mutability of HIV by targeting vulnerable regions of the proteome in which mutations to evade the immune response fatally compromise viral fitness. In this way HIV is trapped between a rock - vaccine induced immune pressure - and a hard place - crippled mutant strains. We have developed maximum entropy models, which map to an Ising spin glass Hamiltonian, fitted to multiple sequence alignment data, to permit the *in silico* construction of sequence space fitness landscapes of HIV. These quantitative models of viral fitness allow identification of T-cell epitopes that maximally abolish fitness and block viral escape mutations, providing a systematic means to identify promising candidate regions for next generation vaccine trials. Clinical validation of our approach is provided by its identification of vulnerable regions in the *gag* polyprotein that are naturally targeted by rare patients who control HIV infection without therapy, and reproduction within our fitness landscapes of well-documented viral escape pathways.

922-Pos Board B708

How to Find Non Hub Important Nodes in Protein Networks?

Vince Grolmusz, Gabor Ivan, Daniel Banky, Balazs Szerencsi.
Eötvös University, Budapest, Hungary.

Biological network data, such as metabolic-, signaling- or physical interaction graphs of proteins are increasingly available in public repositories for important species. Tools for the quantitative analysis of these networks are being developed today. Protein network-based drug target identification methods usually returns protein hubs with large degrees (also called ''connectivity'') in the networks as potentially important targets. Some known, important protein targets, however, are not hubs at all, and on the other hand, perturbing protein hubs in these networks may have several unwanted physiological effects, due to their interaction with numerous partners.

Here we show a novel method applicable in networks with directed edges (such as metabolic networks) that compensates the low degree (non-hub) vertices in the network, and identifies intrinsically important nodes, totally independently from their hub properties. We demonstrate that the method correctly finds numerous already validated drug targets in distinct organisms (Mycobacterium tuberculosis, Plasmodium falciparum and MRSA Staphylococcus aureus), and consequently, it may suggest new possible protein targets as well.

Our method computes the PageRank for the nodes of the network, and divides the PageRank by the in-degree (i.e., the number of incoming edges) of the node. This quotient is the same in all nodes in an undirected graph (even for large- and low degree nodes, that is, for hubs and non-hubs as well), but may differ significantly from node to node in directed graphs. We suggest to assign importance to non-hub nodes with large PageRank/in-degree quotient.

923-Pos Board B709

A Multiscale View of Protein-Protein Interactions

Franca Fraternali.
King's College, London, United Kingdom.

High-throughput proteomics has allowed for the drafting of large Protein-Interaction Networks and systems biology devotes a constant effort in the statistical analysis of these. PPIN derived from large-scale experiments portrait a global picture of proteins connectivity and offer an intricate and complex picture of the extracted interactions. The detailed analysis of a particular cellular function may benefit from the study of smaller and accurately selected sub-networks, containing high confidence interactions resulting from the cross-mapping of multiple sources of information for the nodes such as tissue-specific expression data, gene expression data, domain profiles and structural information. We call these extracted networks Sub-Networks Knowledge. Analysis of these SNKs provides the basis for the identification of fundamental components of the network and of specific pathways associated with the studied phenomenon that can be targeted to speed-up experimental screening and design new experiments.

By reducing further the scale of details analysed in a given PPIN, one could focus only on special proteins in the network, like the multi-partner ones, called 'hubs'. In particular, we concentrated on protein hubs that have a 3D-structure and on the characterization of their conformational dynamics and flexibility of the multiple interfaces. A large-scale survey of the intrinsic dynamics of protein interfaces has been done using a dataset of 340 proteins classified according to the different number of partners. Conformational ensembles were generated