PM13: ISMB2006 Tutorial

Title: "Protein-protein interactions: structure and systems approaches to analyze diverse genomic data"

Topic Area:

- Systems Biology (including Pathways and Networks) (50%)
- Structural Bioinformatics (50%)

Main Presenter:
Dr. Anna Panchenko, Staff Scientist
National Center for Biotechnology Information, NIH
Bldg. 38A, Rm. 8S814, 8600 Rockville Pike, Bethesda, MD 20894
panch@ncbi.nlm.nih.gov, tel: (301)435-5891, fax: (301)480-4637,
http://www.ncbi.nlm.nih.gov/CBBresearch/Panchenko/
Teaching experience: teaching the course "Introduction to Bioinformatics" (http://www.seas.gwu.edu/~madej/cs177/fall2005/crshome0.htm) at George Washington University; the course "Computational Aspects of Molecular Structure" (http://www.apl.jhu.edu/~przytyck/class605_791_7.html) at Johns Hopkins University.

Second Presenter:
Dr. Benjamin Shoemaker, Staff Scientist
National Center for Biotechnology Information, NIH
Bldg. 38A, Rm. 8N811Q, 8600 Rockville Pike, Bethesda, MD 20894
shoemake@mail.nih.gov, tel: (301)594-8093, fax: (301)480-4637,
http://www.ncbi.nlm.nih.gov/Structure/RESEARCH/res.shtml

Abstract:

This tutorial will survey available databases and computational resources for studying protein interactions and discuss the theory behind various approaches of organizing the interactions. The challenge is to find novel and relevant interactions from large experimental sets. The focus will be on network analysis, verification and prediction of protein interactions.

Tutorial level: Intermediate
Prior knowledge required: We expect that participants have prior knowledge of protein structure and have been exposed or used different tools for protein sequence and structure analysis. Suitable for computational biologists, computer scientists, experimental biologists and others working on or interested in protein-protein interaction networks.

Suitability of this tutorial to ISMB:
This session is timely and important for the researchers for the following reasons:

- It is still poorly understood how different proteins interact with each other to provide specific cellular functions.
- Progress and high interest in the area of proteomics and systems biology, as is apparent in recent ISMB conferences.
- Various experimental methods provide a large amount of data on protein-protein interactions; there is need to develop computational methods to verify these data.
- Accelerating pace of the structural genomics initiative provides numerous structures which can be used for verification and prediction of protein-protein and domain-domain interactions.
- Increasing demand in the computational algorithms and methods for exploring the protein-protein interaction space.

Profile of presenter 1.
Anna R Panchenko, Ph.D., is a staff scientist at the National Center for Biotechnology Information, NIH. Her research interests include protein-protein interactions, protein domain classification, protein structure evolution and prediction of protein structure and functional sites. She is one of the organizers of the DIMACS-2006 workshop on protein function prediction. Dr Panchenko is an adjunct faculty member at George Washington University and Johns Hopkins University; she is currently teaching the courses "Introduction to Bioinformatics" and "Computational aspects of molecular structure".

Profile of presenter 2.
Benjamin A. Shoemaker, Ph.D., is a staff scientist at the National Center for Biotechnology Information, NIH. His research interests include protein-protein interactions, development of databases for functional annotation and protein domain classification, and methods of protein interaction prediction. He has created a database of protein domain interactions which is based on the analysis of structurally conserved patterns. He also helped to develop the NCBI Conserved Domain Database.

Tutorial outline.

In this tutorial we will go over the resources available for studying protein-protein interaction data. The challenge is to find novel and relevant interactions from large sets which include significant numbers of false positives. In particular we will look at using the detail of protein structure data as well as other contextual information to assist in prioritizing interaction networks. We will survey available databases and computational resources and discuss the theory behind various approaches of organizing interactions. We will also work through several examples to better understand how a researcher might glean useful information from the abundance of experimental data, perhaps for the study of a particular pathway or as a means of understanding larger evolutionary trends.

Table of contents:

1. **Experimental studies of protein-protein interactions; protein interaction databases (45 min).**
   Systematic analyses of entire genomes reveal the patterns of interactions between proteins and other components of the cell responsible for the properties of an organism. High-throughput experiments are designed to explore the potential interaction patterns within the genomes, among them yeast two-hybrid, mass spectrometry, phage display, X-ray diffraction and other methods. Publicly available databases of protein-protein interactions such as DIP, BIND, MIPS, 3did, InterDOM and others provide experimental data for detailed analysis and standardization of diverse interaction patterns.

2. **Protein domain interactions revealed from protein structure data (30 min).**
   Complementing the large number of interaction data found by yeast two hybrid and other experiments, the X-ray diffraction data provide the most detailed protein-protein interaction information available. Protein structures allow us to sort through the large number of domain-domain and protein-protein interactions and distinguish biologically relevant from non-biological interactions. Because of the reliability of this data, the details of interaction geometries and interfaces can be explored at the atomic level.

3. **Methods of verification of protein interactions (40 min).**
   Despite a large amount of experimental data, the reproducibility of interactions is an issue and different methods can produce quite different interaction maps for the same organism. While some of these discrepancies may be attributed to novel interactions, the challenges associated with the experimental studies can result in a considerable number of false positives. These discrepancies demand the development of computational and statistical methods for verification of interactions in order to avoid further propagation of errors. The quality of interaction data can be assessed, for example, by the analyses of expression distances between interacting partners, the conservation of interactions among homologs or by studying the properties of protein interaction networks.

4. **Analysis of properties of protein interaction networks (20 min).**
   Graph-theoretical analyses of network topologies reveal interesting features of scale-free behavior, modular composition, asymmetric distribution and conservation of certain protein interaction patterns among different organisms.

5. **Evolutionary conservation of interaction interfaces and network topologies (15 min).**
   Evolutionary studies of protein interactions show that many interacting partners conserve the sequence and structure at the interaction interfaces; and at the same time the same interaction surface can be reused by different interacting domains.

6. **Inferring protein domain interactions from experimental protein-protein interaction data (30 min).**
   Although protein-protein interactions can be determined experimentally, highly specific domain interactions between proteins are difficult to infer from sources other than X-ray diffraction experiments. Statistical analysis of domain co-occurrence in interacting proteins can be applied in these cases, which aims to find pairs of domains co-occurring more often in protein-protein interaction pairs compared to non-interacting pairs.

7. **Prediction of protein interactions, homology modeling (30 min).**
   One of the advantages to having structural interaction data is that domain interactions could be predicted or modeled for protein/domain pairs not observed in the structural database. Based on the previous observation that the interaction interface is conserved among close homologs, one can make an assumption that proteins clustered together on a phylogenetic tree should also exhibit similar interaction patterns.