PROTEIN INTERACTIONS AND DISEASE

MARICEL KANN
National Center for Biotechnology Information, NIH
Rockville, MD 20814, U.S.A.

YANAY OFRAN
Department of Biochemistry & Molecular Biophysics, Columbia University
New York, NY 10032, U.S.A.

MARCO PUNTA
Department of Biochemistry & Molecular Biophysics, Columbia University
New York, NY 10032, U.S.A.

PREDRAG RADIVOJAC
School of Informatics, Indiana University
Bloomington, IN 47408, U.S.A.

One of the ultimate goals of biological sciences, and certainly one with a high impact on society, is to improve our understanding of the processes and events related to diseases. Molecular biologists, who traditionally study the structure and function of individual proteins and genes, have gained insight and introduced several discoveries that have ultimately reached the bedside. However, biological processes are not realized by a single molecule, but rather by the complex interaction of proteins with their environment, including nucleic acids, ions, lipids, membranes and, of course, other proteins. Thus, while the analysis of the structure and function of individual proteins is crucial for the understanding of their role in biological processes, it has a limited capability to explain the processes themselves.

Some diseases are caused by simple genomic events that mutate or eliminate specific genes (e.g. frame shift mutations or insertion of viral genes). On the other end of the spectrum are "complex diseases", such as heart or psychiatric diseases that are multifactorial and arguably could not be fully accounted for by simple molecular processes. Between these two ends lie many, if not most, of the pathologies and illnesses, including AIDS, cancer, and Alzheimer's disease.

The goal of this session is to discuss the latest progress in the study of protein interactions and disease. The eight papers accepted cover a wide range of approaches and demonstrate readiness of the community to address this problem.

Troncale et al. used hybrid Petri nets to model the regulation of human hematopoieses. Their work demonstrates a noteworthy role of *in silico* approaches for developing an understanding of the subtle events that determine the fate of stem cells (differentiation vs. self-renewal) and that underlie many hematological diseases and their treatment. Bandyopadhyay et al. used a network analysis of gene expression and protein interaction data to identify active pathways related to HIV pathogenesis. A functional analysis of the detected subnetworks provides useful insights into various stages of the HIV replication cycle.

The papers by Ye et al. and Singh et al. explore the relationship between a disease and protein interactions using structural analysis of the interacting proteins. The work by Ye et al. analyzed distributions of non-synonymous SNPs and found that disease-related mutations tend to cluster at the protein surface. They hypothesized that these sites may be involved in protein binding and suggested that one could use comparative modeling to elucidate the mechanism of protein malfunctions. Singh et al. combined structural data with other sources of information to improve the quality of protein-protein interaction predictions. Their results suggest that the predictive power of structural information is high and that structural information alone can be used for successful inference of protein interactions.

Chen et al. developed a framework to mine disease-related proteins from OMIM and protein interaction data. They demonstrate the power of their method by applying it to Alzheimer's disease. The key to their method is a scoring function that ranks proteins according to their relevance to a particular disease pathway. Terribilini et al. used a machine learning approach to classify amino acids on binding sites. In particular, they applied their methodology to Rev proteins of HIV and found good agreement with experimental data.

The remaining two papers focused on computational aspects of predicting protein-protein and protein-nucleic acid interactions. Gunewardena et al. addressed the problem of predicting binding affinity of protein-DNA interactions, while Robinson et al. explored computational aspects related to the correct prediction of protein interactions.

In recent years, the emergence of new experimental protocols and techniques such as RNA, DNA and protein microarrays, two-hybrid systems, and mass spectrometry, as well as the explosion on the number and size of sequence and structure databases, have changed biomedical science. By taking advantage of the enormous amount of data generated by all these techniques computational biology can now attempt to capture more of the complexity of a biological process. The increasing number of computational studies of protein networks, pathways, protein-protein, protein-metabolite and protein-DNA/RNA interactions

indicates that it is now possible to address the connections between protein interactions and diseases. We are confident that the papers presented in this session will contribute to further advance this field, which may soon become a major area of biomedical research.

Acknowledgements
The session co-chairs would like to thank the numerous reviewers for their help in selecting the best papers among many excellent submissions.