Pacific Symposium on Biocomputing 12:470-471(2007)

DNA-PROTEIN INTERACTIONS: INTEGRATING STRUCTURE, SEQUENCE, AND FUNCTION

MARTHA L. BULYK
Brigham & Women's Hospital and Harvard Medical School
Boston, MA 02115

ALEXANDER J. HARTEMINK
Duke University
Durham, NC 27708

ERNEST FRAENKEL
Massachusetts Institute of Technology
Cambridge, MA 0213

GARY STORMO
Washington University
St. Louis, MO 63108

Recent technological advances have enabled many different types of data to be collected at a genome-wide and proteome-wide scale, including: DNA sequence from various genomes, gene expression data, protein-protein and protein-ligand interactions, and protein-DNA binding data. In addition, efforts in structural biology are yielding structural data on proteins, protein complexes, and protein-ligand interactions. These data provide us for the first time with the opportunity to integrate data from functional studies with structural data, in order to understand how the biophysical aspects of protein-DNA interactions affect their functions. Indeed, much recent work has been devoted to analyzing these data for various focused aspects of this purpose, such as either the regulatory aspects of protein-DNA interactions or the structural aspects of protein-DNA recognition. However, very few studies have integrated these various types of data in order to bridge this divide.

This nascent research area builds on recent developments in diverse areas including DNA motif discovery, modeling of transcriptional regulatory networks, multiple sequence alignments, structural genomics, and structural and evolutionary studies of proteins and DNA. While each of these specific aspects of protein-DNA interactions have been studied previously, these different aspects have just recently begun to be considered together. This session focuses on methods that bridge structure, sequence, and function to infer previously undiscovered associations between these different aspects of protein-DNA interactions.

Methods that employ structure, sequence, and function have several key advantages. First, structural data alone often do not permit the inference of biological function. Second, experimental genomic datasets often contain errors arising from imperfections in the applied technology. Third, functional studies typically do not connect function to structure. Indeed, there has been only a small amount of work that addresses how to take advantage of these currently separate areas of research on protein-DNA interactions. We anticipate that combining these different types of data will allow us to identify essential biological associations, and ultimately to model and predict these interactions.

We accepted three papers for this new session. In the first paper, Liu and Bader predict transcription factor binding sites by calculating binding free energies, starting with the 3D structure of another protein-DNA complex from the same structural class of protein. They apply their approach to homeodomain and bZIP proteins, and it could be applied to proteins of other structural classes as well. In the second paper, Leung and Chin present an algorithm for improved motif discovery that takes advantages of pattern characteristics of different transcription factor binding site motif classes. In the third paper, Zhao and colleagues develop an approach to predict the biological pathway in a target genome that is orthologous to that in a query genome by considering the protein-DNA interactions and operon structures of the pathway genes.

Further progress in these areas may further improve the ability to predict interactions between transcription factors and their DNA binding sites. In addition, there are numerous other challenges in this nascent research area aside from those addressed in the accepted papers for this session. Future work may address questions such as: Do certain types of domains of DNA binding proteins confer particular biophysical properties, either in terms of kinetics or ligand specificity? Has there been an evolutionary selection for the usage of certain structural classes of DNA binding proteins in particular types of biological pathways? How are affinities of protein-DNA interactions tied to function? We believe that as more types of data become widely available, integrative approaches will likely produce great insights into the biophysical, evolutionary, and functional aspects of these important biomolecular interactions.

Acknowledgments

We are grateful to those who submitted manuscripts for this session, and we thank the numerous reviewers for their valuable expertise and time in the peer review process.