
MEETING REPORT

Modeling of Protein Interactions in Genomes

Sandor Vajda,^{*,1} Ilya A. Vakser,² Michael J.E. Sternberg,³ and Joël Janin⁴

¹*Biomedical Engineering, Boston University, Boston, Massachusetts*

²*Department of Cell and Molecular Pharmacology, Medical University of South Carolina, Charleston, South Carolina*

³*Structural Bioinformatics, Department of Biological Sciences, Imperial College, London, United Kingdom*

⁴*Laboratoire d'Enzymologie et Biochimie Structurales, CNRS, Gif-sur-Yvette, France*

INTRODUCTION

Protein interactions play pivotal roles in various aspects of the structural and functional organization of the cell, and their elucidation sheds light on the molecular mechanisms of biologic processes. Genome-wide interaction studies fostering our understanding of the cell as molecular machinery will play an important role in functional genomics. An important challenge is the development of suitable 3-D modeling tools to elucidate the details of specific interactions at the atomic level and the ability to perform such modeling on a genomic scale. In connection with these challenges, the first Conference on Modeling of Protein Interactions in Genomes was held in Charleston, South Carolina, on June 16–19, 2001 (<http://reco3.ams.sunysb.edu/conference/>) to discuss computational procedures that can be used for reconstruction and analysis of the network of connections between proteins in genomes, including structural, genomic, and knowledge-based approaches.

The Conference was organized by Ilya Vakser (Medical University of South Carolina) and Sandor Vajda (Boston University). It included sessions on protein–protein docking, energetics and protein structure–function relationships, protein–small molecule interactions, and identification of interactions and pathways, as well as a poster session. An introductory lecture was presented by Jeffrey Skolnick (Donald Danforth Plant Science Center), who gave a broad overview of the field and described his approach to function determination in a database of proteins with predicted folds. Joel Janin (CNRS, Gif-sur-Yvette) described some of the common structural principles governing protein–protein and protein–DNA recognition.

PROTEIN–PROTEIN DOCKING

Several groups presented new results on rigid-body docking based on the correlation by the fast Fourier transform (FFT) method, introduced by Katchalski-Katzir and co-workers in 1992. The original shape complementarity target function has now been extended to include electrostatic interactions (Michael Sternberg, Imperial Cancer Research Fund; Lynn Ten Eyck, San Diego

Supercomputer Center) or both electrostatic and solvation terms (Zhiping Weng, Boston University). The FFT-based procedure was further extended to include the low-resolution docking of protein models (Ilya Vakser, Medical University of South Carolina). Other approaches to protein–protein docking involve machine vision tools (Ruth Nussinov, National Institutes of Health), Monte-Carlo and genetic algorithms implemented in the SurfDock and AutoDock programs (Arthur Olson, Scripps Research Institute), and a modification of the DOCK program that performs side-chain prediction in parallel with the rigid-body search (Brian Shoichet, Northwestern University).

Current docking methods need improved scoring procedures to discriminate against false-positive predictions. Scoring by empirical free energy functions, while not eliminating all false positives, improves the ranking (Michael Sternberg, Imperial Cancer Research Fund; David Gatchell, Boston University). Another promising direction is the refinement by flexible docking (Carlos Camacho, Boston University). Docking methods were also shown to be successful in building supramolecular structures (Andrew McCammon, University of California San Diego). The kinetics of protein–protein association, from basic understanding to rational design, were discussed by Gideon Schreiber (Weizmann Institute).

ENERGETICS OF PROTEIN INTERACTIONS AND PROTEIN STRUCTURE–FUNCTION RELATIONSHIPS

The theoretical prediction of the structure of a molecule or an assembly of molecules frequently involves the minimization of a function representing the free energy of the system. For single proteins, the free energy surface has

*Correspondence to: S. Vajda, Biomedical Engineering, Boston University, 44 Cummings Street, Boston, MA 02215. E-mail: SV vajda@bu.edu

I.A. Vakser is currently at Bioinformatics Laboratory, Department of Applied Mathematics and Statistics, SUNY at Stony Brook, Stony Brook, NY 11794-3600.

been well studied to predict structures and rationalize folding mechanisms. This approach has now been extended to the analysis of protein–protein association, both in terms of the thermodynamics and kinetics of complex formation. Harold Scheraga (Cornell University) reviewed the relationship between the energetics of protein folding and protein–protein interactions and described applications that involve both folding and docking. Barry Honig (Columbia University) demonstrated how various interactions, primarily electrostatics, effect protein–protein recognition and described rules that help select the correct solution in docking. Glen Kellogg (Virginia Commonwealth University) presented new results on the energetics of biomolecular interactions using an empirical hydrophathy model. The possibility of predicting functional residues without using sequence data was addressed by Adrian Elcock (University of Iowa).

INTERACTIONS OF SMALL MOLECULES WITH PROTEINS

Challenges in small molecule–protein interactions go beyond the well-established docking and scoring methods generally used in drug design-related research and relate to protein–protein docking. Gennady Verkhivker (Agouron Pharmaceuticals) spoke about the universality and diversity of protein folding and molecular recognition mechanisms focusing on peptide–protein interactions. Christophe Verlinde (University of Washington) described a stochastic approximation method that has been used for docking of small molecules to proteins. A central topic in this session was the modulation of protein–protein interactions by small molecules. Small ligands can be important in drug design and studies of cell-signaling processes, reviewed by Andrea Cochran (Genentech). While many drug design methods map the protein surface for the binding sites of molecular probes (small ligands and functional groups), standard docking techniques used in pharmaceutical applications are generally unable to reproduce the experimental data on the binding of organic solvents. Sandor Vajda (Boston University) described his algorithm for site mapping.

IDENTIFICATION OF INTERACTIONS AND PATHWAYS IN GENOMES

The concluding session dealt with the prediction of interacting proteins in genomes and computational problems related to large-scale screening of protein–protein interactions. Genomic data and large-scale screening, primarily by yeast two-hybrid analysis, provide unprecedented new information on proteins and their interactions. They create new challenges for prediction, retrieval, and analysis of the resulting interaction networks and understanding their biologic significance. Structure-based methods will need substantial improvement to enhance genome-wide interaction data, and knowledge-based approaches have major limitations. Thus, experiments must be integrated with a battery of different computational approaches.

Michael Laskowski (Purdue University) showed how the reactivity of a large family of protease inhibitors could be predicted from the sequence of their binding loop. Shoshana Wodak (University Libre de Bruxelles) described a database to retrieve molecular activities and cellular processes. Edward Marcotte (University of Texas) presented the approaches to the construction of genome-wide protein networks implemented in the DIP database. Benno Schwikowski (Institute for Systems Biology) discussed the challenge in visualizing such interaction networks. Christopher Hogue (Mount Sinai Hospital, Toronto) reviewed the BIND protein interaction database. Andrei Tovchigrechko (Medical University of South Carolina) described a database of protein–protein interactions that focuses on structural aspects.

ASSESSING AND BENCHMARKING PROTEIN DOCKING METHODS

John Moult (Center for Advanced Research in Biotechnology) chaired the discussion of a protein docking challenge similar to CASP (Critical Assessment of Structure Prediction), which he organized. CASP includes blind predictions by homology modeling, threading, or ab initio methods of protein structures being determined by X-ray crystallography or nuclear magnetic resonance (NMR). In 1996, it also included docking, but with only one protein–protein target and a small number of participating groups. Noting that the number of potential participants has expanded significantly, and interest in docking techniques has grown together with the need for a structural interpretation of genomic information, participants of the Conference decided to launch an experiment similar to CASP and devoted to protein docking. It was given the name CAPRI: Critical Assessment of Predicted Interactions.

CAPRI had a successful start soon after the Conference with three target complexes and 19 participating groups. Its continuation now depends on experimentalists making targets available, and a call for targets has been issued in *Proteins* (vol. 47, pp. 257, 2002). Additional information on CAPRI is available at <http://capri.ebi.ac.uk>.

In parallel to the blind test performed in CAPRI, docking algorithms will be tested on a benchmark set of some 50 published protein–protein complexes, representing different levels of difficulty for prediction by docking. Zhiping Weng (Boston University) will coordinate the evaluation. Developers should submit to her the docking procedures to be evaluated, which she will run on the benchmark set. Further details of the benchmarking experiment will be available at <http://capri.ebi.ac.uk>.

RECOMMENDATIONS

The most important current development related to protein–protein interaction has been the large-scale screening by yeast two-hybrid essays, mass spectrometry, and other experimental approaches. Computational approaches, including those based on 3-D structure that

provide interaction information at the atomic level, will increasingly contribute as better docking tools are being developed. While substantial progress has been made during the last few years, docking procedures are still too slow for application to large sets of proteins and their capacity to identify near-native structures among a large number of docked complexes is still low. The Conference provided a forum for researchers to discuss the current state of the computational techniques for modeling protein interactions, identify unresolved prob-

lems, and formulate the current priorities in this rapidly developing field.

ACKNOWLEDGMENTS

The Conference was supported by grants from the National Science Foundation, Department of Energy, National Institutes of Health, and the Information Technology Laboratory at the Medical University of South Carolina.