BIOLOGICAL AND MEDICAL PHYSICS, BIOMEDICAL ENGINEERING

BIOLOGICAL AND MEDICAL PHYSICS BIOMEDICAL ENGINEERING

Volumes Published in This Series:

The Physics of Cerebrovascular Diseases, Hademenos, G.J., and Massoud, T.F., 1997

Lipid Bilayers: Structure and Interactions, Katsaras, J., 1999

Physics with Illustrative Examples from Medicine and Biology: Mechanics, Second Edition, Benedek, G.B., and Villars, F.M.H., 2000

Physics with Illustrative Examples from Medicine and Biology: Statistical Physics, Second Edition, Benedek, G.B., and Villars, F.M.H., 2000

Physics with Illustrative Examples from Medicine and Biology: Electricity and Magnetism, Second Edition, Benedek, G.B., and Villars, F.M.H., 2000

Physics of Pulsatile Flow, Zamir, M., 2000

Molecular Engineering of Nanosystems, Rietman, E.A., 2001

Biological Systems Under Extreme Conditions: Structure and Function, Taniguchi, Y. et al., 2001

Intermediate Physics for Medicine and Biology, Third Edition, Hobbie, R.K., 2001

Epilepsy as a Dynamic Disease, Milton, J., and Jung, P. (Eds), 2002

Photonics of Biopolymers, Vekshin, N.L., 2002

Photocatalysis: Science and Technology, Kaneko, M., and Okura, I., 2002

E. coli in Motion, Berg, H.C., 2004

Biochips: Technology and Applications, Xing, W.-L., and Cheng, J. (Eds.), 2003

Laser-Tissue Interactions: Fundamentals and Applications, Niemz, M., 2003

Medical Applications of Nuclear Physics, Bethge, K., 2004

Biological Imaging and Sensing, Furukawa, T. (Ed.), 2004

Biomaterials and Tissue Engineering, Shi, D., 2004

Biomedical Devices and Their Applications, Shi, D., 2004

Microarray Technology and Its Applications, Müller, U.R., and Nicolau, D.V. (Eds), 2004

Emergent Computation: Emphasizing Bioinformatics, Simon, M., 2005

Molecular and Cellular Signaling, Beckerman, M., March 22, 2005

The Physics of Coronary Blood Flow, Zamir, M., May, 2005

The Physics of Birdsong Mindlin, G.B., Laje, R., August, 2005

Radiation Physics for Medical Physicists Podgorsak, E.B., September 2005

Neutron Scattering in Biology—Techniques and Applications Fitter, J., Gutberlet, T., Katsaras, J. (Eds.), January 2006

Forthcoming Titles

Topology in Molecular Biology: DNA and Proteins Monastyrsky, M.I. (Ed.), 2006

Optical Polarization in Biomedical Applications Tuchin, V.V., Wang, L. (et al.), 2006

Continued After Index

Ying Xu, Dong Xu, and
Jie Liang (Eds.)

Computational Methods for Protein Structure Prediction and Modeling

Volume 2: Structure Prediction

Springer

Ying Xu
Department of Biochemistry
and Molecular Biology
University of Georgia
120 Green Street
Athens, GA 30602
USA
email: xyn@bmb.uga.edu

Dong Xu
Department of Computer Science
Digital Biology Laboratory
University of Missouri–Columbia
201 Engineering Building West
Columbia, MO 65211
USA
email: xudong@missouri.edu

Jie Liang
Department of Bioengineering
Center for Bioinformatics
University of Illinois at Chicago
851 S. Morgan Street
Chicago, IL 60607-063
USA
email: jliang@uic.edu

Library of Congress Control Number: 2006929615

ISBN 10: 0-387-33321-5
ISBN 13: 978-0387-33321-2

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

Preface

An ultimate goal of modern biology is to understand how the genetic blueprint of cells (genotype) determines the structure, function, and behavior of a living organism (phenotype). At the center of this scientific endeavor is characterizing the biochemical and cellular roles of proteins, the working molecules of the machinery of life. A key to understanding of functional proteins is the knowledge of their folded structures in a cell, as the structures provide the basis for studying proteins' functions and functional mechanisms at the molecular level.

Researchers working on structure determination have traditionally selected individual proteins due to their functional importance in a biological process or pathway of particular interest. Major research organizations often have their own protein X-ray crystallographic or/and nuclear magnetic resonance facilities for structure determination, which have been conducted at a rate of a few to dozens of structures a year. Realizing the widening gap between the rates of protein identification (through DNA sequencing and identification of potential genes through bioinformatics analysis) and the determination of protein structures, a number of large scientific initiatives have been launched in the past few years by government funding agencies in the United States, Europe, and Japan, with the intention to solve protein structures *en masse,* an effort called *structural genomics*. A number of structural genomics centers (factory-like facilities) have been established that promise to produce solved protein structures in a similar fashion to DNA sequencing. These efforts as well as the growth in the size of the community and the substantive increases in the ease of structure determination, powered with a new generation of technologies such as synchrotron radiation sources and high-resolution NMR, have accelerated the rate of protein structure determination over the past decade. As of January 2006, the protein structure database PDB contained ~34,500 protein structures.

The role of structure for biological sciences and research has grown considerably since the advent of systems biology and the increased emphasis on understanding molecular mechanisms from basic biology to clinical medicine. Just as every geneticist or cell biologist needed in the 1990s to obtain the sequence of the gene whose product or function they were studying, increasingly, those biologists will need to know the structure of the gene product for their research programs in this century. One can anticipate that the rate of structure determination will continue to grow. However, the large expenses and technical details of structure determination mean that it will remain difficult to obtain experimental structures for more than a small fraction of the proteins of interest to biologists. In contrast, DNA sequence determination has doubled routinely in output for a couple of decades. The genome projects have led to the production of 100 gigabytes of DNA data in Genbank, and

as the cost of sequencing continues to drop and the rate continues to accelerate, the scientific community anticipates a day when every individual has the genes of their interest and the genomes of all related major organisms sequenced.

Structure determination of proteins began before nucleic acids could be sequenced, which now appears almost ironic. As microchemistry technologies continue to mature, ever more powerful DNA sequencing instruments and new methods for preparation of suitable quantities of DNA and cheaper, higher sequencing throughput, while enabling a revolution in the biological and biomedical sciences, also left structure determination way behind. As sequencing capacity matured in the last few decades of the twentieth century, DNA sequences exceeded protein structures by 10-fold, then 100-fold, and now there is a 1000-fold difference between the number of genes in Genbank and the number of structures in the PDB. The order of magnitude difference is about to jump again, in the era of metagenomics, as the analyses of communities of largely unculturable organisms in their natural states come to dominate sequence production. The J. Craig Venter Institute's Sargasso Sea experiment and other early metagenomics experiments at least doubled the number of known open reading frames (ORFs) and potential genes, but the more recent ocean voyage data (or GOS) multipled the number on the order of another 10-fold, probably more. The rate of discovery of novel genes and correspondingly novel proteins has not leveled off, since nearly half of new microbial genomes turn out to be novel. Furthermore, in the metagenomics data, new families of proteins are discovered directly proportional to the rate of gene (ORF) discovery.

The bottom line is quite simple. Despite the several fold reduction in cost in structure determination due to the structural genomics projects—the NIH Protein Structure Initiative and comparable initiatives around the world—and the steady increase in the rate of protein structure determination, the number of proteins with unknown structures will continue to grow vastly faster. At an early structural genomics meeting in Avalon, New Jersey, the experimental community voted in favor of experimentally solving 100,000 structures of proteins with less than 30% sequence identity to proteins with known structures. This seemed to some theoreticians at the time as solving "the protein structure problem" and removing the need for theory, simulation, and prediction. Now, while it appears that this goal is aiming too high for just the initiative alone, certainly, the structural community will have 100,000 structures in the PDB not long after the end of this decade—and probably sooner than expected as costs continue to go down and technologies continue to advance. Yet, those 100,000 structures will be significantly less than 1% of the known ORFs genes! The problem, therefore, is *not* about having structures to predict, but having robust enough methods to make predictions that are useful at deep levels in biology, from helping us infer function and directing experimental efforts to providing insight into ligand binding, molecular recognition, drug discovery, and so on. The kind of success in terms of "reasonable" accuracy for "most" targets has been the grand success of the CASP competition (see Chapter 1) but is completely inadequate for the biology of the twenty-first century and the expectations of both basic and applied life sciences. Prediction is not at the requisite level of comprehensive robustness yet, and therein is one of the features of critical importance of the discussions in this book.

Computational methods for predicting protein structure have been actively pursued for some time. Their acceptance and importance grew rapidly after the establishment of a blind competition for predicting protein structure, namely, CASP. CASP involves theoreticians predicting then-unknown protein structures and their verification and analysis following subsequent experimental determination. The validation of the general approach both enhanced funding and brought participants to the field and pointed to the limitations of current methods and the value of extensive research into advanced computational tools. Overall, the rapidly growing importance of structural data for biology fueled the emergence of a new branch of computational biology and of structural biology, an interface between the methods of bioinformatics and molecular biophysics, namely, *structural bioinformatics*. Similar to genomic sequence analysis, bioinformatic studies of protein structures could lead to both deep and general or broad insights about aspects such as the folding, evolution, and function of proteins, the nature of protein–ligand and protein–protein interactions, and the mechanisms by which proteins act. The success of such studies could have immense impacts not just on science but on the whole society through providing insight into the molecular etiology of diseases, developing novel, effective therapeutic agents and treatment regimens, and engineering biological molecules for novel or enhanced biochemical functions.

As one of the most active research fields in bioinformatics, structural bioinformatics addresses a wide spectrum of scientific issues, including the computational prediction of protein secondary and tertiary structures, protein docking with small molecules and with macromolecules (i.e., DNA, RNA, and proteins), simulation of dynamic behaviors of proteins, protein structure characterization and classification, and study of structure–function relationships. While proteins were viewed as essentially static three-dimensional structures up until the 1980s, the establishment of computational methods, and subsequent advances in experimental probes that could provide data at suitable time scales, led to a revolution in how biologists think about proteins. Indeed, over the past few decades, computational studies using molecular dynamics simulations of protein structure have played essential roles in understanding the detailed functional mechanisms of proteins important in a wide variety of biological processes. Within the applied life sciences, protein docking has been extensively applied in the drug discovery pipeline in the pharmaceutical and biotech industry.

Protein structure prediction and modeling tools are becoming an integral part of the standard toolkit in biological and biomedical research. Similar to sequence analysis tools, such as BLAST for sequence comparison, the new methods for structure prediction are now among the first approaches used when starting a biological investigation, conducted prior to actual experimental design. That computational analysis would become the first step for experimentalists represents a major paradigm shift that is still occurring but is clearly essential to deal with the maturation of the field, the large quantities of data, and the complexity of biology itself as reflected in the requirement for today's powerful experimental probes used to address sophisticated questions in biology. This paradigm shift was noted first by Wally Gilbert, in a prescient article fifteen years ago ("Toward a new paradigm for molecular biology,"

Nature 1991, 349:99), who asserted that biologists would have to change their mode of approach to studying nature and to begin each experimental project with a bioinformatics analysis of extant literature and other computational approaches. This paradigm shift is deeply interconnected with the increased emphasis on computational tools and the expectation for robust methods for structure prediction.

Similar to other fields of bioinformatics, structural bioinformatics is a rapidly growing science. New computational techniques and new research foci emerge every few months, which makes the writing of textbooks a challenging problem. While a number of books have been published covering various aspects of protein structure prediction and modeling, it is widely recognized that the field lacks a comprehensive and coherent overview of the science of "protein structure prediction and modeling," which span a range from very basic problems (around physical and chemical properties and principles), such as the potential function and free energies that determine the folded shape of a protein, to the algorithmic techniques for solving various structure prediction problems, to the engineering aspects of implementation of computer prediction software, and to applications of prediction capabilities for investigations focused on functional properties. As educators at universities, we feel that there is an urgent need for a well-written, comprehensive textbook, one that proverbially goes from soup to nuts, and that this requirement is most critical for beginners entering this field as young students or as experienced researchers coming from other disciplines.

This book is an attempt to fill this gap by providing systematic expositions of the computational methods for all major aspects of protein structure analysis, prediction, and modeling. We have designed the chapters to address comprehensively the main topics of the field. In addition, chapters have been connected seamlessly through a systematic design of the overall structure of the book. We have selected individual topics carefully so that the book would be useful to a broad readership, including students, postdoctoral fellows, research scientists moving into the field, as well as professional practitioners/bioinformatics experts who want to brush up on topics related to their own research areas. We expect that the book can be used as a textbook for upper undergraduate-level or graduate-level bioinformatics courses. Extensive prior knowledge is not required to read and comprehend the information presented. In other words, a dedicated reader with a college degree in computational, biological, or physical science should be able to follow the book without much difficulty. To facilitate learning and to articulate clearly to the reader what background is needed to obtain the maximum benefit from the book, we have included four appendices describing the prerequisites in (1) biology, (2) computer science, (3) physics and chemistry, and (4) mathematics and statistics. If a reader lacks knowledge in a particular area, he or she could benefit by starting from the references provided in the corresponding appendix.

While the chapters are organized in a logical order, each chapter in the book is a self-contained review of a specific subject. Hence, a reader does not need to read through the chapters sequentially. Each chapter is designed to cover the following material: (1) the problem definition and a historical perspective, (2) a mathematical or computational formulation of the problem, (3) the computational methods and

algorithms, (4) the performance results, (5) the existing software packages, (6) the strengths, pitfalls, and challenges in current research, and (7) the most promising future directions. Since this is a rapidly developing field that encompasses an exceptionally wide range of research topics, it is difficult for any individual to write a comprehensive textbook on the entire field. We have been fortunate in assembling a team of experts to write this book. The authors are actively doing research at the forefront of the major areas of the field and bring extensive experience and insight into the central intellectual methods and ideas in the subdomain and its difficulties, accomplishments, and potential for the future.

Chapter 1 (A Historical Perspective and Overview of Protein Structure Prediction) gives a perspective on the methods for the prediction of protein structure and the progress that has been achieved. It also discusses recent advances and the role of protein structure modeling and prediction today, as well as touching briefly on important goals and directions for the future.

Chapter 2 (Empirical Force Fields) addresses the physical force fields used in the atomic modeling of proteins, including bond, bond-angle, dihedral, electrostatic, van der Waals, and solvation energy. Several widely used physical force fields are introduced, including CHARMM, AMBER, and GROMOS.

Chapter 3 (Knowledge-Based Energy Functions for Computational Studies of Proteins) discusses the theoretical framework and methods for developing knowledge-based potential functions essential for protein structure prediction, protein–protein interaction, and protein sequence design. Empirical scoring functions including single-body energy function, statistical method for pairwise interaction between amino acids, and scoring function based on optimization are addressed.

Chapter 4 (Computational Methods for Domain Partitioning of Protein Structures) covers the basic concept of protein structural domains and practical applications. A number of computational techniques for domain partition are described, along with their applications to protein structure prediction. Also described are a few, widely used, protein domain databases and associated analysis tools.

Chapter 5 (Protein Structure Comparison and Classification) discusses the basic problem of protein structure comparison and applications, and computational approaches for aligning two protein structures. Applications of the structure–structure alignment algorithms to protein structure search against the PDB and to protein structural motif search in the PDB are also discussed.

Chapter 6 (Computation of Protein Geometry and Its Applications: Packing and Function Prediction) treats protein structures as 3D geometrical objects, and discusses structural issues from a geometric point of view, such as (1) the union of ball models, molecular surface, and solvent-accessible surface, (2) geometric constructs such as Voronoi diagram, Delaunay triangulation, alpha shape, surface geometry (including cavities and pockets) and their computation, (3) local surface similarity measure in terms of shape and sequence, and (4) function prediction based on protein surface patterns. Also described are the application issues of these computational techniques.

Chapter 7 (Local Structure Prediction of Proteins) covers protein secondary structure prediction, supersecondary structure prediction, prediction of disordered

regions, and applications to tertiary structure prediction. A number of popular prediction software packages are described.

Chapter 8 (Protein Contact Maps Prediction) describes the basic principles for residue contact predictions, and computational approaches for prediction of residue–residue contacts. Also discussed is the relevance to tertiary structure prediction. A number of popular prediction programs are introduced.

Chapter 9 (Modeling Protein Aggregate Assembly and Structure) describes the basic problem of structure misfolding and implications, experimental approach for data collection in support of computational modeling, computational approaches to prediction of misfolded structures, and related applications.

Chapter 10 (Homology-Based Modeling of Protein Structure) presents the foundation for homology modeling, computational methods for sequence–sequence alignment and constructing atomic models, structural model assessment, and manual tuning of homology models. A number of popular modeling packages are introduced.

Chapter 11 (Modeling Protein Structures Based on Density Maps at Intermediate Resolutions) discusses methods for constructing atomic models from density maps of proteins at intermediate resolution, such as those obtained from electron cryomicrosopy. Details of application of computational tools for identifying α-helices, ß-sheets, as well as geometric analysis are described.

Chapter 12 (Protein Structure Prediction by Protein Threading) describes the threading approach for predicting protein structure. It discusses the basic concepts of protein folds, an empirical energy function, and optimal methods for fitting a protein sequence to a structural template, including the divide-and-conquer, the integer programming, and tree-decomposition approaches. This chapter also gives practical guidance, along with a list of resources, on using threading for structure prediction.

Chapter 13 (*De Novo* Protein Structure Prediction) describes protein folding and free energy minimization, lattice model and search algorithms, off-lattice model and search algorithms, and mini-threading. Benchmark performance of various tools in CASP is described.

Chapter 14 (Structure Prediction of Membrane Proteins) covers the methods for prediction of secondary structure and topology of membrane proteins, as well as prediction of their tertiary structure. A list of useful resources for membrane protein structure prediction is also provided.

Chapter 15 (Structure Prediction of Protein Complexes) describes computational issues for docking, including protein–protein docking (both rigid body and flexible docking), protein–DNA docking, and protein–ligand docking. It covers computational representation for biomolecular surface, various docking algorithms, clustering docking results, scoring function for ranking docking results, and start-of-the-art benchmarks.

Chapter 16 (Structure-Based Drug Design) describes computational issues for rational drug design based on protein structures, including protein therapeutics based on cytokines, antibodies, and engineered enzymes, docking in structure-based drug design as a virtual screening tool in lead discovery and optimization, and ligand-based drug design using pharmacophore modeling and quantitative

structure–activity relationship. A number of software packages for structure-based design are compared.

Chapter 17 (Protein Structure Prediction as a Systems Problem) provides a novel systematic view on solving the complex problem of protein structure prediction. It introduces consensus-based approach, pipeline approach, and expert system for predicting protein structure and for inferring protein functions. This chapter also discusses issues such as benchmark data and evaluation metrics. An example of protein structure prediction at genome-wide scale is also given.

Chapter 18 (Resources and Infrastructure for Structural Bioinformatics) describes tools, databases, and other resources of protein structure analysis and prediction available on the Internet. These include the PDB and related databases and servers, structural visualization tools, protein sequence and function databases, as well as resources for RNA structure modeling and prediction. It also gives information on major journals, professional societies, and conferences of the field.

Appendix 1 (Biological and Chemical Basics Related to Protein Structures) introduces central dogma of molecular biology, macromolecules in the cell (DNA, RNA, protein), amino acid residues, peptide chain, primary, secondary, tertiary, and quaternary structure of proteins, and protein evolution.

Appendix 2 (Computer Science for Structural Informatics) discusses computer science concepts that are essential for effective computation for protein structure prediction. These include efficient data structure, computational complexity and NP-hardness, various algorithmic techniques, parallel computing, and programming.

Appendix 3 (Physical and Chemical Basis for Structural Bioinformatics) covers basic concepts of our physical world, including unit system, coordinate systems, and energy surfaces. It also describes biochemical and biophysical concepts such as chemical reaction, peptide bonds, covalent bonds, hydrogen bonds, electrostatic interactions, van der Waals interactions, as well as hydrophobic interactions. In addition, this chapter discusses basic concepts from thermodynamics and statistical mechanics. Computational sampling techniques such as molecular dynamics and Monte Carlo method are also discussed.

Appendix 4 (Mathematics and Statistics for Studying Protein Structures) covers various basic concepts in mathematics and statistics, often used in structural bioinformatics studies such as probability distributions (uniform, Gaussian, binomial and multinomial, Dirichlet and gamma, extreme value distribution), basics of information theory including entropy, relative entropy, and mutual information, Markovian process and hidden Markov model, hypothesis testing, statistical inference (maximum likelihood, expectation maximization, and Bayesian approach), and statistical sampling (rejection sampling, Gibbs sampling, and Metropolis–Hastings algorithm).

Ying Xu
Dong Xu
Jie Liang
John Wooley

April 2006

Acknowledgments

During the editing of this book, we, the editors, have received tremendous help from many friends, colleagues, and families, to whom we would like to take this opportunity to express our deep gratitude and appreciation. First we would like to thank Dr. Eli Greenbaum of Oak Ridge National Laboratory, who encouraged us to start this book project and contacted the publisher at Springer on our behalf. We are very grateful to the following colleagues who have critically reviewed the drafts of the chapters of the book at various stages: Nick Alexandrov, Nir Ben-Tal, Natasja Brooijmans, Chris Bystroff, Pablo Chacon, Luonan Chen, Zhong Chen, Yong Duan, Roland Dunbrack, Daniel Fischer, Juntao Guo, Jaap Heringa, Xiche Hu, Ana Kitazono, Ioan Kosztin, Sandeep Kumar, Xiang Li, Guohui Lin, Zhijie Liu, Hui Lu, Alex Mackerell, Kunbin Qu, Robert C. Rizzo, Ilya Shindyalov, Ambuj Singh, Alex Tropsha, Iosif Vaisman, Ilya Vakser, Stella Veretnik, Björn Wallner, Jin Wang, Zhexin Xiang, Yang Dai, Xin Yuan, and Yaoqi Zhou. Their invaluable input on the scientific content, on the pedagogical style, and on the writing style helped to improve these book chapters significantly. We also want to thank Ms. Joan Yantko of the University of Georgia for her tireless help on numerous fronts in this book project, including taking care of a large number of email communications between the editors and the authors and chasing busy authors to get their revisions and other material. Last but not least, we want to thank our families for their constant support and encouragement during the process of us working on this book project.

Contents

Contributors

Natasja Brooijmans

Chemical and Screening Sciences
Wyeth Research
Pearl River, New York 10965

Christopher Bystroff

Department of Biology
Rensselaer Polytechnic Institute
Troy, New York 12180

Liming Cai

Department of Computer Science
University of Georgia
Athens, Georgia 30602-7404

Orhan Camoglu

Department of Computer Science
University of California Santa Barbara
Santa Barbara, California 93106

Yang Dai

Department of Bioengineering
University of Illinois at Chicago
Chicago, Illinois 60607-7052

Haobo Guo

Department of Biochemistry and
Cellular and Molecular Biology
University of Tennessee
Knoxville, Tennessee 37996

Hong Guo

Department of Biochemistry and
Cellular and Molecular
Biology
University of Tennessee
Knoxville, Tennessee 37996

Jun-tao Guo

Department of Biochemistry and
Molecular Biology
University of Georgia
Athens, Georgia 30602-7229

Carol K. Hall

Department of Chemical and
Biomolecular Engineering
North Carolina State University
Raleigh, North Carolina 27695

Jaap Heringa

Centre for Integrative Bioinformatics
Vrije Universiteit
1081 HV Amsterdam, The
Netherlands

Xiche Hu

Department of Chemistry
University of Toledo
Toledo, Ohio 43606

Ling-Hong Hung

Department of Microbiology
University of Washington
Seattle, Washington 98195-7242

Xiang Li

Department of Bioengineering
University of Illinois at Chicago
Chicago, Illinois 60607-7052

Jie Liang

Department of Bioengineering
University of Illinois at Chicago
Chicago, Illinois 60607-7052

Guohui Lin

Department of Computing Science
University of Alberta
Edmonton, Alberta T6G 2E8, Canada

Zhijie Liu

Department of Biochemistry and
Molecular Biology
University of Georgia
Athens, Georgia 30602-7229

Hui Lu

Department of Bioengineering
University of Illinois at Chicago
Chicago, Illinois 60607-7052

Jianpeng Ma

Department of Biochemistry and
Molecular Biology
Baylor College of Medicine
Houston, Texas 77030
and
Department of Bioengineering
Rice University
Houston, Texas 77005

Alexander D. MacKerell, Jr.

Department of Pharmaceutical
Chemistry
School of Pharmacy
University of Maryland
Baltimore, Maryland 21201

Shing-Chung Ngan

Department of Microbiology
University of Washington
Seattle, Washington 98195-7242

Ognjen Perišić

Department of Bioengineering
University of Illinois at Chicago
Chicago, Illinois 60607-7052

Brian Pierce

Department of Biomedical Engineering
Boston University
Boston, Massachusetts 02215

Kunbin Qu

Department of Chemistry
Rigel Pharmaceuticals, Inc.
San Francisco, California 94080

Ram Samudrala

Department of Microbiology
University of Washington
Seattle, Washington 98195-7242

Ilya Shindyalov

San Diego Supercomputer Center
University of California San Diego
San Diego, California 92093-0505

Victor A. Simossis

Centre for Integrative Bioinformatics
Vrije Universiteit
1081 HV Amsterdam, The Netherlands

Ambuj K. Singh

Department of Computer Science
University of California Santa Barbara
Santa Barbara, California 93106

Stella Veretnik

San Diego Supercomputer Center
University of California San Diego
San Diego, California 92093-0505

Zhiping Weng

Department of Biomedical Engineering
Boston University
Boston, Massachusetts 02215

Ronald B. Wetzel

Department of Structural Biology
Pittsburgh Institute for Neurodegenerative Diseases
University of Pittsburgh School of Medicine
Pittsburgh, Pennsylvania 15260

John C. Wooley

Associate Vice Chancellor for Research
University of California San Diego
San Diego, California 92093-0043

Zhexin Xiang

Center for Molecular Modeling
Center for Information Technology
National Institutes of Health
Bethesda, Maryland 20892-5624

Dong Xu

Computer Science Department
University of Missouri—Columbia
Columbia, Missouri 65211-2060

Ying Xu

Institute of Bioinformatics and
Department of Biochemistry
and Molecular Biology
University of Georgia
Athens, Georgia 30602-7229

Yuzhen Ye

Bioinformatics and Systems Biology
Department
The Burnham Institute for Medical
Research
La Jolla, California 92037

Xin Yuan

Department of Computer Science
Florida State University
Tallahassee, Florida 32306