

Catalyzing Inquiry
at the Interface of
**Computing
and
Biology**

John C. Wooley and Herbert S. Lin, editors

Committee on Frontiers at the Interface of Computing and Biology

Computer Science and Telecommunications Board

Division on Engineering and Physical Sciences

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

Support for this project was provided by the Defense Advanced Research Projects Agency under Contract No. MDA972-00-1-0005, the National Science Foundation under Contract No. DBI-0094528, the Department of Health and Human Services/National Institutes of Health (including the National Institute of General Medical Sciences and the National Center for Research Resources) under Contract No. N01-OD-4-2139, the Department of Energy under Contract No. DE-FG02-02ER63336, the Department of Energy's Office of Science (BER) under Interagency Agreement No. DE-FG02-04ER63934, and National Research Council funds. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number 0-309-09612-X
Library of Congress Control Number: 2005936580

Cover designed by Jennifer M. Bishop.

This report is available from
Computer Science and Telecommunications Board
National Research Council
500 Fifth Street, N.W.
Washington, DC 20001

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, <http://www.nap.edu>.

Copyright 2005 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

**COMMITTEE ON FRONTIERS AT THE INTERFACE OF
COMPUTING AND BIOLOGY**

JOHN C. WOOLEY, University of California at San Diego, *Chair*
ADAM P. ARKIN, University of California at Berkeley and Lawrence Berkeley
National Laboratory
ERIC BRILL, Microsoft Research Labs
ROBERT M. CORN, University of California at Irvine
CHRIS DIORIO, University of Washington
LEAH EDELSTEIN-KESHET, University of British Columbia
MARK H. ELLISMAN, University of California at San Diego
MARCUS W. FELDMAN, Stanford University
DAVID K. GIFFORD, Massachusetts Institute of Technology
TAKEO KANADE, Carnegie Mellon University
STEPHEN S. LADERMAN, Agilent Laboratories
JAMES S. SCHWABER, Thomas Jefferson Medical College

Staff

Herbert Lin, Senior Scientist and Study Director
Geoff Cohen, Consultant to CSTB
Mitchell Waldrop, Consultant to CSTB
Daehee Hwang, Consultant to Board on Biology
Robin Schoen, Senior Staff Officer
Elizabeth Grossman, Senior Staff Officer (through March 2001)
Jennifer Bishop, Program Associate
D.C. Drake, Senior Program Assistant (through March 2003)

COMPUTER SCIENCE AND TELECOMMUNICATIONS BOARD

JOSEPH TRAUB, Columbia University, *Chair*
ERIC BENHAMOU, Benhamou Global Ventures, LLC
DAVID D. CLARK, Massachusetts Institute of Technology, *CSTB Chair Emeritus*
WILLIAM DALLY, Stanford University
MARK E. DEAN, IBM Almaden Research Center
DEBORAH ESTRIN, University of California, Los Angeles
JOAN FEIGENBAUM, Yale University
HECTOR GARCIA-MOLINA, Stanford University
KEVIN KAHN, Intel Corporation
JAMES KAJIYA, Microsoft Corporation
MICHAEL KATZ, University of California, Berkeley
RANDY H. KATZ, University of California, Berkeley
WENDY A. KELLOGG, IBM T.J. Watson Research Center
SARA KIESLER, Carnegie Mellon University
BUTLER W. LAMPSON, Microsoft Corporation, *CSTB Member Emeritus*
TERESA H. MENG, Stanford University
TOM M. MITCHELL, Carnegie Mellon University
DANIEL PIKE, GCI Cable and Entertainment
ERIC SCHMIDT, Google Inc.
FRED B. SCHNEIDER, Cornell University
WILLIAM STEAD, Vanderbilt University
ANDREW J. VITERBI, Viterbi Group, LLC
JEANNETTE M. WING, Carnegie Mellon University

RICHARD ROWBERG, Acting Director
KRISTEN BATCH, Research Associate
JENNIFER M. BISHOP, Program Associate
JANET BRISCOE, Manager, Program Operations
JON EISENBERG, Senior Program Officer and Associate Director
RENEE HAWKINS, Financial Associate
MARGARET MARSH HUYNH, Senior Program Assistant
HERBERT S. LIN, Senior Scientist
LYNETTE I. MILLETT, Senior Program Officer
JANICE SABUDA, Senior Program Assistant
GLORIA WESTBROOK, Senior Program Assistant
BRANDYE WILLIAMS, Staff Assistant

For more information on CSTB, see its Web site at <http://www.cstb.org>, write to CSTB, National Research Council, 500 Fifth Street, N.W., Washington, DC 20001, or call (202) 334-2605, or e-mail the CSTB at cstb@nas.edu.

Preface

In the last decade of the 20th century, computer science and biology both emerged as fields capable of remarkable and rapid change. Moreover, they evolved as fields of inquiry in ways that draw attention to their areas of intersection. The continuing advancements in technology and the pace of scientific research present the means for computing to help answer fundamental questions in the biological sciences and for biology to demonstrate that new approaches to computing are possible.

Advances in the power and ease of use of computing and communications systems have fueled computational biology (e.g., genomics) and bioinformatics (e.g., database development and analysis). Modeling and simulation of biological entities such as cells have joined biologists and computer scientists (and mathematicians, physicists, and statisticians too) to work together on activities from pharmaceutical design to environmental analysis.

On the other side, computer scientists have pondered the significance of biology for their field. For example, computer scientists have explored the use of DNA as a substrate for new computing hardware and the use of biological approaches in solving hard computing problems. Exploration of biological computation suggests a potential for insight into the nature of and alternative processes for computation, and it also gives rise to questions about hybrid systems that achieve some kind of synergy of biological and computational systems. And there is also the fact that biological systems exhibit characteristics such as adaptability, self-healing, evolution, and learning that would be desirable in the information technologies that humans use.

Making the most of the research opportunities at the interface of computing and biology—what we are calling the BioComp interface—requires illuminating what they are and effectively engaging people from both computing and biology. As in other contexts, the challenges of interdisciplinary education and of collaboration are significant, and each will require attention, together with substantive work from both policy makers and researchers. At the start of the 1990s, attempts were made to stimulate mutual interest and collaboration among young researchers in computing and biology. Those early efforts yielded nontrivial successes, but in retrospect represented a Version 1.0 prototype for the potential in bringing the two fields together. Circumstances today seem much more favorable for progress. New research teams and training programs have been formed as individual investigators from the respective communities, government agencies, and private foundations have become increasingly engaged. Similarly, some larger groups of investigators from different backgrounds have been able to

obtain funding to work together to address cross-disciplinary research problems. It is against this background that the committee sees a Version 2.0 of the BioComp interface emerging that will yield unprecedented progress and advance.

The range of possible activities at the BioComp interface is broad, and accordingly so is the range of interested agencies, which include the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation (NSF), the Department of Energy (DOE), and the National Institutes of Health (NIH). These agencies have, to varying degrees, recognized that truly cross-disciplinary work would build on both computing and biology, and they have sought to advance activities at the interface.

This report by the Committee on Frontiers at the Interface of Computing and Biology seeks to establish the intellectual legitimacy of a fundamentally cross-disciplinary collaboration between biologists and computer scientists. That is, while some universities are increasingly favorable to research at the intersection, life science researchers at other universities are strongly impeded in their efforts to collaborate. This report addresses these impediments and describes some strategies for overcoming them.

In addition, this report provides a wealth of well-documented examples. As a rule, these examples have generally been selected to illustrate the breadth of the topic in question, rather than to identify the most important areas of activity. That is, the appropriate spirit in which to view these examples is “let a thousand flowers bloom,” rather than one of “finding the prettiest flowers.” It is hoped that these examples will encourage students in the life sciences to start or to continue study in computer science that will enable them to be more effective users of computing in their future biological studies. In the opposite direction, the report seeks to describe a rich and diverse domain—biology—within which computer scientists can find worthy problems that challenge current knowledge in computing. It is hoped that this awareness will motivate interested computer scientists to learn about biological phenomena, data, experimentation, and the like—so that they can engage biologists more effectively.

To gather information on such a broad area, the committee took input from a wide variety of sources. The committee convened two workshops in March 2001 and May 2001, and committee members or staff attended relevant workshops sponsored by other groups. The committee mined the published literature extensively. It solicited input from other scientists known to be active in BioComp research. An early draft of the report was examined by a number of reviewers far larger than usual for National Research Council (NRC) reports, and the draft was modified in accordance with their extensive input, which helped the committee to sharpen its message and strengthen its presentation.

The result of these efforts is the first comprehensive NRC study that suggests a high-level intellectual structure for federal agencies for supporting work at the BioComp interface. Although workshop reports have been supported by individual agencies on the subject of computing applied to various aspects of biological inquiry, the NRC has not until now undertaken a study whose intent was to be inclusive.

Within the NRC, the lead unit on this project was the Computer Science and Telecommunications Board (CSTB), and Marjory Blumenthal and Elizabeth Grossman launched the project. The committee also acknowledges with gratitude the contribution of the Board on Biology—Robin Schoen continued work on the project after Elizabeth Grossman’s departure. Geoff Cohen and Mitch Waldrop, consultants to CSTB, made major substantive contributions to this report. A variety of project assistants, including D.C. Drake, Jennifer Bishop, Gloria Westbrook, and Margaret Huynh, provided research and administrative support. Finally, grateful thanks are offered to DARPA, NIH, NSF, and DOE for their financial support for this project as well as their patience in awaiting the final report. No single agency can respond to the challenges and opportunities at the interface, and the committee hopes that its analysis will facilitate agency efforts to define their own priorities, set their own path, and participate in what will be a continuing adventure along the frontier at this exciting and promising interface, which will continue to develop throughout the 21st century.

A Personal Note from the Chair

The committee found the scope of the study and the need to achieve an adequate level of balance in both directions around the BioComp interface to be a challenge. This challenge, I hope, has been met, but this was only possible due to the recruitment of an outstanding physicist turned computer science policy expert from the NRC. Specifically, after the original series of meetings, Herb Lin from the CSTB side of the NRC joined the effort, and most notably, followed up on the committee's earlier analyses by interviewing numerous individuals engaged in both biocomputing (applications of biology to computing) and computational biology (applications of computing to biology). This was invaluable, as was Herb's never ending enthusiasm, insight into the nature of the interdisciplinary discussions that are growing, and his willingness to engage in learning a lot about biology. The report could never have been completed without his persistence. His expertise in editing and analytical treatment of policy and technical material allowed us to sustain a broad vision. (Even with the length and breadth of this study, we were able to cover only selected areas at the interface.) The committee's efforts were sustained and accelerated by Herb's determination that we stay the course despite the size of the task, and by his insightful comments, criticisms, and suggestions on every aspect of the study and the report.

John Wooley, *Chair*
Committee on Frontiers at the Interface
of Computing and Biology

Acknowledgment of Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

Harold Abelson, Massachusetts Institute of Technology,
Eric Benhamou, Benhamou Global Ventures, LLC,
Mina Bissell, Lawrence Berkeley National Laboratory,
Gaetano Borriello, University of Washington,
Dennis Bray, University of Cambridge,
Steve Burbeck, IBM,
Andrea Califano, Columbia University,
Charles Cantor, Boston University,
David D. Clark, Massachusetts Institute of Technology,
G. Bard Ermentrout, University of Pittsburgh,
Lisa Fauci, Tulane University,
David Galas, Keck Graduate Institute,
Leon Glass, McGill University,
Mark D. Hill, University of Wisconsin-Madison,
Tony Hunter, The Salk Institute for Biological Studies,
Sara Kiesler, Carnegie Mellon University,
Isaac Kohane, Children's Hospital,
Nancy Kopell, Boston University,
Bud Mishra, New York University,
William Noble, University of Washington,

Alan S. Perelson, Los Alamos National Laboratory,
Robert J. Robbins, Fred Hutchinson Cancer Research Center,
Lee Segel, The Weizmann Institute of Science,
Larry L. Smarr, University of California, San Diego,
Sylvia Spengler, National Science Foundation,
William Stead, Vanderbilt University,
Suresh Subramani, University of California, San Diego,
Charles Taylor, University of California, Los Angeles, and
Andrew J. Viterbi, Viterbi Group, LLC.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by Russ Altman, Stanford University. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Contents

EXECUTIVE SUMMARY	1
1 INTRODUCTION	9
1.1 Excitement at the Interface of Computing and Biology, 9	
1.2 Perspectives on the BioComp Interface, 10	
1.2.1 From the Biology Side, 11	
1.2.2 From the Computing Side, 12	
1.2.3 The Role of Organization and Culture, 13	
1.3 Imagine What's Next, 14	
1.4 Some Relevant History in Building the Interface, 16	
1.4.1 The Human Genome Project, 16	
1.4.2 The Computing-to-Biology Interface, 16	
1.4.3 The Biology-to-Computing Interface, 17	
1.5 Background, Organization, and Approach of This Report, 19	
2 21st CENTURY BIOLOGY	23
2.1 What Kind of Science?, 23	
2.1.1 The Roots of Biological Culture, 23	
2.1.2 Molecular Biology and the Biochemical Basis of Life, 24	
2.1.3 Biological Components and Processes in Context, and Biological Complexity, 25	
2.2 Toward a Biology of the 21st Century, 27	
2.3 Roles for Computing and Information Technology in Biology, 31	
2.3.1 Biology as an Information Science, 31	
2.3.2 Computational Tools, 33	
2.3.3 Computational Models, 33	
2.3.4 A Computational Perspective on Biology, 33	
2.3.5 Cyberinfrastructure and Data Acquisition, 34	
2.4 Challenges to Biological Epistemology, 34	

3	ON THE NATURE OF BIOLOGICAL DATA	35
3.1	Data Heterogeneity, 35	
3.2	Data in High Volume, 37	
3.3	Data Accuracy and Consistency, 38	
3.4	Data Organization, 40	
3.5	Data Sharing, 44	
3.6	Data Integration, 47	
3.7	Data Curation and Provenance, 49	
4	COMPUTATIONAL TOOLS	57
4.1	The Role of Computational Tools, 57	
4.2	Tools for Data Integration, 58	
4.2.1	Desiderata, 59	
4.2.2	Data Standards, 60	
4.2.3	Data Normalization, 60	
4.2.4	Data Warehousing, 62	
4.2.5	Data Federation, 62	
4.2.6	Data Mediators/Middleware, 65	
4.2.7	Databases as Models, 65	
4.2.8	Ontologies, 67	
4.2.8.1	Ontologies for Common Terminology and Descriptions, 67	
4.2.8.2	Ontologies for Automated Reasoning, 69	
4.2.9	Annotations and Metadata, 73	
4.2.10	A Case Study: The Cell Centered Database, 75	
4.2.11	A Case Study: Ecological and Evolutionary Databases, 79	
4.3	Data Presentation, 81	
4.3.1	Graphical Interfaces, 81	
4.3.2	Tangible Physical Interfaces, 83	
4.3.3	Automated Literature Searching, 84	
4.4	Algorithms for Operating on Biological Data, 87	
4.4.1	Preliminaries: DNA Sequence as a Digital String, 87	
4.4.2	Proteins as Labeled Graphs, 88	
4.4.3	Algorithms and Voluminous Datasets, 89	
4.4.4	Gene Recognition, 89	
4.4.5	Sequence Alignment and Evolutionary Relationships, 92	
4.4.6	Mapping Genetic Variation Within a Species, 94	
4.4.7	Analysis of Gene Expression Data, 97	
4.4.8	Data Mining and Discovery, 100	
4.4.8.1	The First Known Biological Discovery from Mining Databases, 100	
4.4.8.2	A Contemporary Example: Protein Family Classification and Data Integration for Functional Analysis of Proteins, 101	
4.4.9	Determination of Three-dimensional Protein Structure, 103	
4.4.10	Protein Identification and Quantification from Mass Spectrometry, 106	
4.4.11	Pharmacological Screening of Potential Drug Compounds, 107	
4.4.12	Algorithms Related to Imaging, 107	
4.4.12.1	Image Rendering, 110	
4.4.12.2	Image Segmentation, 110	
4.4.12.3	Image Registration, 113	
4.4.12.4	Image Classification, 114	
4.5	Developing Computational Tools, 114	

5	COMPUTATIONAL MODELING AND SIMULATION AS ENABLERS FOR BIOLOGICAL DISCOVERY	117
5.1	On Models in Biology, 117	
5.2	Why Biological Models Can Be Useful, 119	
5.2.1	Models Provide a Coherent Framework for Interpreting Data, 120	
5.2.2	Models Highlight Basic Concepts of Wide Applicability, 120	
5.2.3	Models Uncover New Phenomena or Concepts to Explore, 121	
5.2.4	Models Identify Key Factors or Components of a System, 121	
5.2.5	Models Can Link Levels of Detail (Individual to Population), 122	
5.2.6	Models Enable the Formalization of Intuitive Understandings, 122	
5.2.7	Models Can Be Used as a Tool for Helping to Screen Unpromising Hypotheses, 122	
5.2.8	Models Inform Experimental Design, 122	
5.2.9	Models Can Predict Variables Inaccessible to Measurement, 123	
5.2.10	Models Can Link What Is Known to What Is Yet Unknown, 124	
5.2.11	Models Can Be Used to Generate Accurate Quantitative Predictions, 124	
5.2.12	Models Expand the Range of Questions That Can Meaningfully Be Asked, 124	
5.3	Types of Models, 125	
5.3.1	From Qualitative Model to Computational Simulation, 125	
5.3.2	Hybrid Models, 129	
5.3.3	Multiscale Models, 130	
5.3.4	Model Comparison and Evaluation, 131	
5.4	Modeling and Simulation in Action, 134	
5.4.1	Molecular and Structural Biology, 134	
5.4.1.1	Predicting Complex Protein Structures, 134	
5.4.1.2	A Method to Discern a Functional Class of Proteins, 134	
5.4.1.3	Molecular Docking, 136	
5.4.1.4	Computational Analysis and Recognition of Functional and Structural Sites in Protein Structures, 136	
5.4.2	Cell Biology and Physiology, 139	
5.4.2.1	Cellular Modeling and Simulation Efforts, 139	
5.4.2.2	Cell Cycle Regulation, 146	
5.4.2.3	A Computational Model to Determine the Effects of SNPs in Human Pathophysiology of Red Blood Cells, 148	
5.4.2.4	Spatial Inhomogeneities in Cellular Development, 149	
5.4.2.4.1	Unraveling the Physical Basis of Microtubule Structure and Stability, 149	
5.4.2.4.2	The Movement of <i>Listeria</i> Bacteria, 150	
5.4.2.4.3	Morphological Control of Spatiotemporal Patterns of Intracellular Signaling, 151	
5.4.3	Genetic Regulation, 152	
5.4.3.1	<i>Cis</i> -regulation of Transcription Activity as Process Control Computing, 152	
5.4.3.2	Genetic Regulatory Networks as Finite-state Automata, 153	
5.4.3.3	Genetic Regulation as Circuits, 157	
5.4.3.4	Combinatorial Synthesis of Genetic Networks, 158	
5.4.3.5	Identifying Systems Responses by Combining Experimental Data with Biological Network Information, 159	
5.4.4	Organ Physiology, 161	
5.4.4.1	Multiscale Physiological Modeling, 161	
5.4.4.2	Hematology (Leukemia), 162	

5.4.4.3	Immunology, 163	
5.4.4.4	The Heart, 166	
5.4.5	Neuroscience, 172	
5.4.5.1	The Broad Landscape of Computational Neuroscience, 172	
5.4.5.2	Large-scale Neural Modeling, 173	
5.4.5.3	Muscular Control, 175	
5.4.5.4	Synaptic Transmission, 181	
5.4.5.5	Neuropsychiatry, 187	
5.4.6	Virology, 189	
5.4.7	Epidemiology, 191	
5.4.8	Evolution and Ecology, 193	
5.4.8.1	Commonalities Between Evolution and Ecology, 193	
5.4.8.2	Examples from Evolution, 194	
5.4.8.2.1	Reconstruction of the <i>Saccharomyces</i> Phylogenetic Tree, 195	
5.4.8.2.2	Modeling of Myxomatosis Evolution in Australia, 197	
5.4.8.2.3	The Evolution of Proteins, 198	
5.4.8.2.4	The Emergence of Complex Genomes, 199	
5.4.8.3	Examples from Ecology, 200	
5.4.8.3.1	Impact of Spatial Distribution in Ecosystems, 200	
5.4.8.3.2	Forest Dynamics, 201	
5.5	Technical Challenges Related to Modeling, 202	
6	A COMPUTATIONAL AND ENGINEERING VIEW OF BIOLOGY	205
6.1	Biological Information Processing, 205	
6.2	An Engineering Perspective on Biological Organisms, 210	
6.2.1	Biological Organisms as Engineered Entities, 210	
6.2.2	Biology as Reverse Engineering, 211	
6.2.3	Modularity in Biological Entities, 213	
6.2.4	Robustness in Biological Entities, 217	
6.2.5	Noise in Biological Phenomena, 220	
6.3	A Computational Metaphor for Biology, 223	
7	CYBERINFRASTRUCTURE AND DATA ACQUISITION	227
7.1	Cyberinfrastructure for 21st Century Biology, 227	
7.1.1	What Is Cyberinfrastructure? 227	
7.1.2	Why Is Cyberinfrastructure Relevant? 228	
7.1.3	The Role of High-performance computing, 231	
7.1.4	The Role of Networking, 235	
7.1.5	An Example of Using Cyberinfrastructure for Neuroscience Research, 235	
7.2	Data Acquisition and Laboratory Automation, 237	
7.2.1	Today's Technologies for Data Acquisition, 237	
7.2.2	Examples of Future Technologies, 241	
7.2.3	Future Challenges, 245	
8	BIOLOGICAL INSPIRATION FOR COMPUTING	247
8.1	The Impact of Biology on Computing, 247	
8.1.1	Biology and Computing: Promise and Skepticism, 247	
8.1.2	The Meaning of Biological Inspiration, 249	
8.1.3	Multiple Roles: Biology for Computing Insight, 250	

- 8.2 Examples of Biology as a Source of Principles for Computing, 253
 - 8.2.1 Swarm Intelligence and Particle Swarm Optimization, 253
 - 8.2.2 Robotics 1: The Subsumption Architecture, 255
 - 8.2.3 Robotics 2: Bacterium-inspired Chemotaxis in Robots, 256
 - 8.2.4 Self-Healing Systems, 257
 - 8.2.5 Immunology and Computer Security, 259
 - 8.2.5.1 Why Immunology Might Be Relevant, 259
 - 8.2.5.2 Some Possible Applications of Immunology-based Computer Security, 259
 - 8.2.5.3 Immunological Design Principles for Computer Security, 260
 - 8.2.5.4 An Example: Immunology and Intruder Detection, 262
 - 8.2.5.5 Interesting Questions and Challenges, 263
 - 8.2.5.5.1 Definition of Self, 263
 - 8.2.5.5.2 More Immunological Mechanisms, 263
 - 8.2.5.6 Some Possible Difficulties with an Immunological Approach, 264
 - 8.2.6 Amorphous Computing, 264
- 8.3 Biology as Implementer of Mechanisms for Computing, 265
 - 8.3.1 Evolutionary Computation, 265
 - 8.3.1.1 What Is Evolutionary Computation? 265
 - 8.3.1.2 Suitability of Problems for Evolutionary Computation, 267
 - 8.3.1.3 Correctness of a Solution, 268
 - 8.3.1.4 Solution Representation, 269
 - 8.3.1.5 Selection of Primitives, 269
 - 8.3.1.6 More Evolutionary Mechanisms, 270
 - 8.3.1.6.1 Coevolution, 270
 - 8.3.1.6.2 Development, 270
 - 8.3.1.7 Behavior of Evolutionary Processes, 271
 - 8.3.2 Robotics 3: Energy and Compliance Management, 272
 - 8.3.3 Neuroscience and Computing, 273
 - 8.3.3.1 Neuroscience and Architecture in Broad Strokes, 274
 - 8.3.3.2 Neural Networks, 274
 - 8.3.3.3 Neurally Inspired Sensors, 277
 - 8.3.4 Ant Algorithms, 277
 - 8.3.4.1 Ant Colony Optimization, 278
 - 8.3.4.2 Other Ant Algorithms, 279
- 8.4 Biology as Physical Substrate for Computing, 280
 - 8.4.1 Biomolecular Computing, 280
 - 8.4.1.1 Description, 281
 - 8.4.1.2 Potential Application Domains, 284
 - 8.4.1.3 Challenges, 285
 - 8.4.1.4 Future Directions, 286
 - 8.4.2 Synthetic Biology, 287
 - 8.4.2.1 An Engineering Approach to Building Living Systems, 288
 - 8.4.2.2 Cellular Logic Gates, 288
 - 8.4.2.3 Broader Views of Synthetic Biology, 290
 - 8.4.2.4 Applications, 291
 - 8.4.2.5 Challenges, 291
 - 8.4.3 Nanofabrication and DNA Self-Assembly, 292
 - 8.4.3.1 Rationale, 292
 - 8.4.3.2 Applications, 296

8.4.3.3	Prospects, 297	
8.4.3.4	Hybrid Systems, 298	
9	ILLUSTRATIVE PROBLEM DOMAINS AT THE INTERFACE OF COMPUTING AND BIOLOGY	299
9.1	Why Problem-focused Research? 299	
9.2	Cellular and Organismal Modeling, 300	
9.3	A Synthetic Cell with Physical Form, 303	
9.4	Neural Information Processing and Neural Prosthetics, 306	
9.5	Evolutionary Biology, 311	
9.6	Computational Ecology, 313	
9.7	Genome-enabled Individualized Medicine, 317	
9.7.1	Disease Susceptibility, 318	
9.7.2	Drug Response and Pharmacogenomics, 320	
9.7.3	Nutritional Genomics, 322	
9.8	A Digital Human on Which a Surgeon Can Operate Virtually, 323	
9.9	Computational Theories of Self-assembly and Self-modification, 325	
9.10	A Theory of Biological Information and Complexity, 327	
10	CULTURE AND RESEARCH INFRASTRUCTURE	331
10.1	Setting the Context, 331	
10.2	Organizations and Institutions, 332	
10.2.1	The Nature of the Community, 332	
10.2.2	Education and Training, 333	
10.2.2.1	General Considerations, 333	
10.2.2.2	Undergraduate Programs, 334	
10.2.2.3	The BIO2010 Report, 335	
10.2.2.3.1	Engineering, 336	
10.2.2.3.2	Quantitative Training, 336	
10.2.2.3.3	Computer Science, 337	
10.2.2.4	Graduate Programs, 341	
10.2.2.5	Postdoctoral Programs, 343	
10.2.2.5.1	The Sloan/DOE Postdoctoral Awards for Computational Molecular Biology, 343	
10.2.2.5.2	The Burroughs-Wellcome Career Awards at the Scientific Interface, 344	
10.2.2.5.3	Keck Center for Computational and Structural Biology: The Research Training Program, 344	
10.2.2.6	Faculty Retraining in Midcareer, 345	
10.2.3	Academic Organizations, 346	
10.2.4	Industry, 349	
10.2.4.1	Major IT Corporations, 350	
10.2.4.2	Major Life Science Corporations, 350	
10.2.4.3	Start-up and Smaller Companies, 351	
10.2.5	Funding and Support, 352	
10.2.5.1	General Considerations, 352	
10.2.5.1.1	The Role of Funding Institutions, 352	
10.2.5.1.2	The Review Process, 352	

10.2.5.2	Federal Support, 353	
10.2.5.2.1	National Institutes of Health, 353	
10.2.5.2.2	National Science Foundation, 356	
10.2.5.2.3	Department of Energy, 357	
10.2.5.2.4	Defense Advanced Research Projects Agency, 359	
10.3	Barriers, 361	
10.3.1	Differences in Intellectual Style, 361	
10.3.1.1	Historical Origins and Intellectual Traditions, 361	
10.3.1.2	Different Approaches to Education and Training, 362	
10.3.1.3	The Role of Theory, 363	
10.3.1.4	Data and Experimentation, 365	
10.3.1.5	A Caricature of Intellectual Differences, 367	
10.3.2	Differences in Culture, 367	
10.3.2.1	The Nature of the Research Enterprise, 367	
10.3.2.2	Publication Venue, 369	
10.3.2.3	Organization of Human Resources, 369	
10.3.2.4	Devaluing the Contributions of the Other, 369	
10.3.2.5	Attitudinal Issues, 370	
10.3.3	Barriers in Academia, 371	
10.3.3.1	Academic Disciplines and Departmental Structure, 371	
10.3.3.2	Structure of Educational Programs, 372	
10.3.3.3	Coordination Costs, 373	
10.3.3.4	Risks of Retraining and Conversion, 374	
10.3.3.5	Rapid But Uneven Changes in Biology, 374	
10.3.3.6	Funding Risk, 375	
10.3.3.7	Local Cyberinfrastructure, 375	
10.3.4	Barriers in Commerce and Business, 375	
10.3.4.1	Importance Assigned to Short-term Payoffs, 375	
10.3.4.2	Reduced Workforces, 376	
10.3.4.3	Proprietary Systems, 376	
10.3.4.4	Cultural Differences Between Industry and Academia, 376	
10.3.5	Issues Related to Funding Policies and Review Mechanisms, 377	
10.3.5.1	Scope of Supported Work, 377	
10.3.5.2	Scale of Supported Work, 379	
10.3.5.3	The Review Process, 380	
10.3.6	Issues Related to Intellectual Property and Publication Credit, 381	
11	CONCLUSIONS AND RECOMMENDATIONS	383
11.1	Disciplinary Perspectives, 383	
11.1.1	The Biology-Computing Interface, 383	
11.1.2	Other Emerging Fields at the BioComp Interface, 384	
11.2	Moving Forward, 385	
11.2.1	Building a New Community, 386	
11.2.2	Core Principles for Practitioners, 387	
11.2.3	Core Principles for Research Institutions, 388	
11.3	The Special Significance of Educational Innovation at the BioComp Interface, 389	
11.3.1	Content, 389	
11.3.2	Mechanisms, 390	

- 11.4 Recommendations for Research Funding Agencies, 392
 - 11.4.1 Core Principles for Funding Agencies, 392
 - 11.4.2 National Institutes of Health, 395
 - 11.4.3 National Science Foundation, 397
 - 11.4.4 Department of Energy, 397
 - 11.4.5 Defense Advanced Research Projects Agency, 398
- 11.5 Conclusions Regarding Industry, 398
- 11.6 Closing Thoughts, 399

APPENDIXES

A	The Secrets of Life: A Mathematician's Introduction to Molecular Biology	403
B	Challenge Problems in Bioinformatics and Computational Biology from Other Reports	429
C	Biographies of Committee Members and Staff	437
D	Workshop Participants	443
	What Is CSTB?	445

Catalyzing Inquiry
at the Interface of

Computing and Biology

