Transforming Glycoscience

A ROADMAP FOR THE FUTURE

Committee on Assessing the Importance and Impact of
Glycomics and Glycosciences

Board on Chemical Sciences and Technology
Board on Life Sciences

Division on Earth and Life Studies

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

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

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW 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.

This project was supported by the National Institutes of Health under contract N01-OD-4-2139, TO#251, the National Science Foundation under grant CHE-1138764, the U.S. Department of Energy under contract DE-SC0007069, the Food and Drug Administration under contract HHSF223200810020I, TO#HHSF22301023, and the Howard Hughes Medical Institute. The views expressed herein are those of the authors and do not necessarily reflect the views of the organizations or agencies that provided support for the project. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

International Standard Book Number-13: 978-0-309-26083-1
International Standard Book Number-10: 0-309-26083-3

Additional copies of the report are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Cover: Imaging of glycans in Chinese hamster ovary cells. Glycans are labeled in red and cell nuclei and Golgi apparatus are labeled in blue and green, respectively. Image courtesy of Carolyn Bertozzi, Scott Laughlin, and Jeremy Baskin. Source: Baskin J. M., J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli, and C. R. Bertozzi. 2007. Copper-free click chemistry for dynamic in vivo imaging. *Proceedings of the National Academy of Sciences of the United States of America* 104(43):16793-16797.

Copyright 2012 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. Charles M. Vest 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. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON ASSESSING THE IMPORTANCE AND IMPACT OF GLYCOMICS AND GLYCOSCIENCES

DAVID WALT (*Chair*), Department of Chemistry, Tufts University
KIYOKO F. AOKI-KINOSHITA, Department of Bioinformatics, Soka University, Japan
BRAD BENDIAK, University of Colorado, Denver
CAROLYN R. BERTOZZI, University of California, Berkeley
GEERT-JAN BOONS, Complex Carbohydrate Research Center, University of Georgia
ALAN DARVILL, Complex Carbohydrate Research Center, University of Georgia
GERALD HART, Department of Biological Chemistry, Johns Hopkins University School of Medicine
LAURA L. KIESSLING, Department of Chemistry, University of Wisconsin
JOHN LOWE, Genentech, Inc.
ROBERT J. MOON, Forest Products Laboratory, U.S. Forest Service
JAMES C. PAULSON, Departments of Chemical Physiology and Molecular Biology, The Scripps Research Institute
RAM SASISEKHARAN, Massachusetts Institute of Technology
AJIT P. VARKI, Glycobiology Research and Training Center, University of California, San Diego, School of Medicine
CHI-HUEY WONG, Academia Sinica, Taiwan, and The Scripps Research Institute

Staff

KATHERINE BOWMAN, Co-Study Director and Senior Program Officer, Board on Life Sciences
DOUGLAS FRIEDMAN, Co-Study Director and Program Officer, Board on Chemical Sciences and Technology
SHEENA SIDDIQUI, Senior Program Associate, Board on Chemical Sciences and Technology
RACHEL YANCEY, Senior Program Assistant, Board on Chemical Sciences and Technology
JOE ALPER, Consulting Science Writer

BOARD ON CHEMICAL SCIENCES AND TECHNOLOGY

Members

PABLO DEBENEDETTI *(Co-chair)*, Princeton University, New Jersey
C. DALE POULTER *(Co-chair)*, University of Utah, Salt Lake City
ZHENAN BAO, Stanford University, California
ROBERT BERGMAN, University of California, Berkeley
HENRY BRYNDZA, E. I. du Pont de Nemours & Company, Wilmington, Delaware
EMILY CARTER, Princeton University, New Jersey
DAVID CHRISTIANSON, University of Pennsylvania
MARY JANE HAGENSON, Chevron Phillips Chemical Company, LLC, The Woodlands, Texas
CAROL J. HENRY, The George Washington University, Washington, D.C.
JILL HRUBY, Sandia National Laboratories, Albuquerque, New Mexico
MICHAEL KERBY, ExxonMobil Chemical Company, Baytown, Texas
CHARLES E. KOLB, Aerodyne Research, Inc., Billerica, Massachusetts
JOSEF MICHL, University of Colorado, Boulder
SANDER MILLS, Merck, Sharp, & Dohme Corporation, Kenilworth, New Jersey
DAVID MORSE, Corning International, Corning, New York
ROBERT E. ROBERTS, Institute for Defense Analyses, Washington, D.C.
DARLENE J. S. SOLOMON, Aligent Laboratories, Santa Clara, California
JEAN TOM, Bristol-Myers Squibb, West Windsor, New Jersey
DAVID WALT, Tufts University, Medford, Massachusetts

National Research Council Staff

DOROTHY ZOLANDZ, Director
AMANDA CLINE, Administrative Assistant
DOUGLAS FRIEDMAN, Program Officer
KATHRYN HUGHES, Senior Program Officer
TINA M. MASCIANGIOLI, Senior Program Officer
SHEENA SIDDIQUI, Senior Program Associate
RACHEL YANCEY, Senior Program Assistant

BOARD ON LIFE SCIENCES

JO HANDELSMAN (*Chair*), Yale University, New Haven, Connecticut
VICKI L. CHANDLER, Gordon and Betty Moore Foundation, Palo Alto, California
SEAN EDDY, HHMI Janelia Farm Research Campus, Ashburn, Virginia
SARAH C. R. ELGIN, Washington University, St. Louis, Missouri
DAVID R. FRANZ, Former Cdr USAMRIID, Frederick, Maryland
LOUIS J. GROSS, University of Tennessee, Knoxville, Tennessee
RICHARD A. JOHNSON, Arnold & Porter, LLC, Washington, D.C.
JUDITH KIMBLE, University of Wisconsin, Madison, Wisconsin
CATO T. LAURENCIN, University of Connecticut Health Center, Farmington, Connecticut
ALAN I. LESHNER, American Association for the Advancement of Science, Washington, D.C.
BERNARD LO, University of California, San Francisco, California
KAREN E. NELSON, J. Craig Venter Institute, Rockville, Maryland
ROBERT M. NEREM, Georgia Institute of Technology, Atlanta, Georgia
CAMILLE PARMESAN, University of Texas, Austin, Texas
ALISON G. POWER, Cornell University, Ithaca, New York
MARGARET RILEY, University of Massachusetts, Amherst, Massachusetts
BRUCE W. STILLMAN, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
JANIS C. WEEKS, University of Oregon, Eugene, Oregon
CYNTHIA WOLBERGER, Johns Hopkins University School of Medicine, Baltimore, Maryland
MARY WOOLLEY, Research!America, Alexandria, Virginia

Staff

FRANCES E. SHARPLES, Director
JO L. HUSBANDS, Scholar/Senior Project Director
JAY B. LABOV, Senior Scientist/Program Director for Biology Education
KATHERINE W. BOWMAN, Senior Program Officer
INDIA HOOK-BARNARD, Senior Program Officer
MARILEE K. SHELTON-DAVENPORT, Senior Program Officer
KEEGAN SAWYER, Program Officer
BETHELHEM M. BANJAW, Financial Associate
ORIN E. LUKE, Senior Program Assistant
CARL G. ANDERSON, Program Associate
SAYYEDA AYESHA AHMED, Senior Program Assistant

Preface

Although I was trained as a synthetic organic chemist and was involved in carbohydrate research early in my scientific career, my research has primarily been focused on developing new technologies for making analytical measurements. This work has led to the development and commercialization of some of the technologies that are presently used for the revolution in genetics and genomics that has taken place over the past decade. I have seen the transformation in scientific capability enabled by these new genetic tools. Access to both the tools and the public databases by virtually any scientist and engineer has democratized the field and has made genetic information an essential component of many fields of science. Science has benefitted tremendously, and many fields are decades ahead of where they would have been without these capabilities. In addition, genetic technologies are beginning to have a big impact on practical applications—diagnostics, therapeutics, and animal breeding to name a few. The economic benefit is in the billions of dollars per year and growing.

This study can be viewed as an opportunity to elevate the importance and possibilities of glycoscience, which is equally pervasive and certainly more directly linked to biological activity than genetics. For example, glycans are responsible for virtually all cell-cell recognition. Moreover, they play a central role in recent burgeoning biofuels efforts. But glycoscience has much more to offer, as described in this report. It was identifying these opportunities and providing a roadmap that was the challenge to

the Committee on Assessing the Importance and Impact of Glycomics and Glycosciences.

The National Academies assembled a stellar group of glycoscientists for this committee. They came from disparate fields—biology, chemistry, and computer science—and work on equally diverse problems in fundamental biology, synthetic chemistry, health, energy, and materials science. I have been so impressed with the passion of these glycoscience committee members for their field. They have worked for many years to advance this important yet underappreciated area—and, despite my limited knowledge of the field, they welcomed me both as a colleague and a friend. It has been a genuine pleasure to work with this dedicated and passionate group of scientists. They have worked tirelessly to help advance the field and, more importantly, science in general with their contributions to this study and to this report. The community is indebted to their service.

The National Academies staff are the real heroes. In particular, Dr. Katherine Bowman and Dr. Douglas Friedman were essential to the success of this study. Katie and Doug pushed the committee to meet deadlines, dealt with the challenging logistics of committee members spanning 12 time zones, helped pull the report together, and worked tirelessly. Even with difficult deadlines, I never heard them complain. They brought ideas and creativity to the discussions. Their selfless dedication to science is admirable and should be a model for us all. In addition to Katie and Doug, Sheena Siddiqui and Rachel Yancey provided superb administrative support. I also want to thank Dr. Fran Sharples, director of the Board on Life Sciences, and Dr. Dorothy Zolandz, director of the Board on Chemical Sciences and Technology, for their support and vision.

This report has the potential to transform the field of glycoscience, but—more significantly—it should transform science in dramatic ways. Sugars are ubiquitous, and scientists in all fields will realize the full potential of their research only by embracing and incorporating glycoscience. The tools for realizing this potential are not available yet. It is the hope of the committee that this report will bring glycoscience into the scientific mainstream.

David Walt, *Chair*
Committee on Assessing the Importance and
Impact of Glycomics and Glycosciences

Acknowledgments

This report has been reviewed in draft form by persons 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 the published report as sound as possible and to ensure that it meets institutional standards of 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 for their review of this report:

Richard D. Cummings, *Emory University*
Samuel Danishefsky, *Memorial Sloan-Kettering Cancer Center*
Anne Dell, *Imperial College London*
Stephen S. Kelley, *North Carolina State University*
Nicolle Packer, *Macquarie University (Sydney, Australia)*
Robert Sackstein, *Harvard Medical School*
Chris Somerville, *University of California, Berkeley*
George M. Whitesides, *Harvard University*

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 **Kenneth**

Moloy, *Dupont Central Research and Development* and **Johanna Dwyer**, *Tufts Medical Center*. Appointed by the National Research Council, they 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 authors and the institution.

Contents

SUMMARY 1

1 Introduction **13**

1.1 Understanding the Language of Life: The Centrality of Sugars, 13
1.2 Genes and Proteins Are Not Enough: The Rich Information Content of Glycans, 21
1.3 How Glycoscience Builds on Genomics and Proteomics, 22
1.4 Why Now? The Case for Change, 23
1.5 Charge to the Committee, 25
1.6 Organization of the Report, 26

2 The Landscape of Current Research in Glycoscience **29**

2.1 An Overview of Glycoscience Worldwide, 30
2.2 An "Omics" Field—Glycoscience in Its Infancy, 33
2.3 Common Concerns Among U.S. and International Glycoscientists, 34
2.4 Conclusion, 36

3 Glycoscience in Health, Energy, and Materials **37**

3.1 Glycoscience and Health, 38
3.1.1 Glycans' Regulation of Inflammation, 38
3.1.2 Glycans' Essential Role in Regulation of the Immune System, 40

3.1.3 Glycans' Key Role in Infectious Diseases and Vaccine Development, 43
3.1.4 Glycans' Multifaceted Role in Cardiovascular Disease, 46
3.1.5 Glycans and the Molecular Mechanisms of Chronic Diseases, 47
3.1.6 Glycans' Roles in Cancer Progression and Early Detection, 49
3.1.7 Critical Roles of Glycans in Human Development, 51
3.1.8 Bioactivity and Pharmacokinetics of Drugs, 52
3.1.9 Key Messages on Glycoscience and Health, 54
3.2 Glycoscience and Energy, 55
3.2.1 Biomass—Plant Cell Walls, 56
3.2.2 Recalcitrance to Degradation of Biomass Feedstock, 58
3.2.3 Key Messages on Glycoscience and Energy, 60
3.3 Glycoscience and Materials, 61
3.3.1 Fine Chemicals and Feedstocks, 62
3.3.2 Polymeric Materials, 64
3.3.3 Nanomaterials, 65
3.3.4 Key Messages on Glycoscience and Materials, 69
3.4 Summary, 70

4 Examples of Outstanding Questions in Glycoscience 71
4.1 What Are the Mechanisms and Roles of Glycan Diversification in Evolution?, 72
4.2 How Can Single Glycoforms and Polysaccharides Be Synthesized and How Can Specific Glycans at Specific Sites on Glycoproteins Be Modified?, 73
4.3 How Does Glycan Microheterogeneity Occur, What Does It Do, and What Is Its Impact?, 74
4.4 What Are the Three-Dimensional Structures of Intact Glycoproteins?, 75
4.5 How Does Nuclear and Cytoplasmic Protein Glycosylation Regulate Cellular Physiology?, 75
4.6 How Does the Glycocalyx Affect the Organization of Molecules on the Cell Surface?, 76
4.7 How Can the Glycans and Glycoproteins on a Single Cell Be Determined?, 77
4.8 What Are the Functions of Microbial and Host Interactions Involving Glycans?, 78
4.9 How Do Glycan Binding Proteins Decode the Glycome?, 79
4.10 How Can Plant Recalcitrance to Degradation Be Understood and Overcome?, 80

4.11 How Can Sugars Be Reassembled to Develop Materials with Tailored Properties and Functionality?, 81
4.12 Summary, 82

5 The Toolkit of Glycoscience 85
5.1 Synthesis, 86
5.1.1 General Aspects, 86
5.1.2 Synthetic Tools, 90
5.1.3 Manipulating Glycans by Pathway Engineering, 95
5.1.4 Synthesis of Standards for Mass Spectrometry, 97
5.1.5 Key Messages on Glycan Synthesis, 98
5.2 Analysis, 98
5.2.1 Analysis of Primary Glycan Structures, 100
5.2.2 Analysis of Glycoconjugates, 108
5.2.3 Analysis of Glycan-Protein Interactions, 109
5.2.4 Analysis of the Roles of Some Glycans in Metabolic Pathways Related to Energy Metabolism, 111
5.2.5 Analysis Techniques That Relate Glycan Structures and Their Synthetic Enzymes, 111
5.2.6 Analysis of Locations of Specific Glycan Structures in Organisms Through Various Imaging Techniques, 112
5.2.7 Key Messages on Glycan Analysis, 114
5.3 Computational Modeling, 114
5.3.1 Computational Modeling of Oligo- and Polysaccharides, 114
5.3.2 Protein-Glycan Interactions, 115
5.3.3 Atomistic Modeling of Crystalline Cellulose, 116
5.3.4 Key Messages on Computational Analysis of Glycans, 118
5.4 Glycoenzymes, 118
5.4.1 Classes of Glycoenzymes, 118
5.4.2 Applications of Glycosyltransferases and Other Glycoenzymes, 120
5.4.3 Key Messages on Glycoenzymes, 125
5.5 Systems Glycobiology, 125
5.6 Informatics and Databases, 126
5.6.1 Limited Successes in Developing Broadly Available Informatics Tools, 127
5.6.2 Critical Need for Development of a Single Integrated Database, 129
5.6.3 Key Messages on Glycan Bioinformatics and Databases, 132
5.7 Summary and Findings, 133

6 Deciphering the Glycome for Human Health and Sustainability: Findings, Recommendations, and Roadmap 135

REFERENCES 143

APPENDIXES

A Committee Member Biographies 159
B The Landscape of Current Research in Glycoscience: Additional Information 167
C Workshop on the Future of Glycoscience: Agenda and Participants 177
D Input Received Online and Through Other Data Gathering 185
E Glossary 187

Summary

In response to a request from the National Institutes of Health (NIH), the Food and Drug Administration (FDA), the U.S. Department of Energy (DOE), and the National Science Foundation (NSF), the National Research Council convened a committee to assess the importance and impact of glycoscience, explore the landscape of current research, and identify the challenges that will need to be addressed to enable the field to move forward. The committee was charged to "articulate a unified vision for the field on glycoscience and glycomics" and to "develop a roadmap with concrete research goals to significantly advance [the field]" (see Statement of Task, Box 1-5). The committee's consensus findings, conclusions, and recommendations in addressing this charge are summarized below.

WHY GLYCOSCIENCE?

Glycans are one of the four fundamental classes of macromolecules that comprise living systems, along with nucleic acids, proteins, and lipids, and are made up of individual sugar units linked to one another in a multitude of ways. Understanding the structures and functions of glycans is central to understanding biology. One of the most common reactions on the planet—photosynthesis—uses energy from sunlight to ultimately combine carbon dioxide and water into polymers of sugars such as starch, glycogen, or cellulose—glycans used in our metabolic pathways to provide us with energy, that provide structural support in such materials as wood, and that other animals are able to use as energy sources.

BOX S-1
Carbohydrate, Glycan, Saccharide, or Sugar?

Carbohydrate: A generic term used interchangeably in this report with sugar, saccharide, or glycan. This term includes monosaccharides, oligosaccharides, and polysaccharides as well as derivatives of these compounds.

Glycan: A generic term for any sugar or assembly of sugars, in free form or attached to another molecule.

Saccharide: A generic term for any carbohydrate or assembly of carbohydrates, in free form or attached to another molecule.

Sugar: A generic term often used to refer to any carbohydrate, but most frequently to low molecular weight carbohydrates that are sweet in taste.

Glycans (see Box S-1) are ubiquitous. All living cells are coated on their cell membranes with glycans or include glycan polymers as integral components of their cell walls. They play diverse roles, including critical functions in the areas of cell signaling, molecular recognition, immunity, and inflammation. They are the cell surface molecules that define the ABO blood groups, influencing an individual's ability to receive another's blood. Glycans are attached to specific locations on many proteins, modulating aspects of their biological activity through molecular recognition or affecting their circulation time in blood. The difference between glycan molecules added by humans when they naturally produce the protein erythropoietin, which affects red blood cell production, and glycan molecules present when this protein drug is produced commercially in cell culture, serves as the basis for antidoping tests in athletes. They are also central components of plant cell walls, which enable plants to grow upright and to resist degradation from the environment and from microbes.

Advances in the life sciences over the past several decades have led to a greater understanding of many of the basic mechanisms present in biological systems. Stimulated by the Human Genome Project, there have been improvements in understanding the central dogma of molecular biology. Sequences of DNA—genes—are transcribed into RNA, which in turn are translated to form proteins. This basic understanding, along with advances in the tools used to study biology, underpins the expansion of both genomics and proteomics. The wide array of posttranslational modifications that occur on proteins are also part of this increasingly clear picture. Protein glycosylation, one of the most common forms of post-

translational modification, is important for many biological processes and often serves as an analog switch that is capable of carefully modulating protein activity.

Relatively little attention has been paid to this class of molecules, and glycoscience remains a relatively understudied field. It is hard to predict what advances in glycoscience will bring as the contributions from the life sciences and chemical sciences to numerous areas of applied science continue to expand. This report provides an overview of the current knowledge and state of glycoscience and illustrates why glycoscience is central to multiple avenues of research. An expanding understanding of glycan functions and structures will complement and strengthen other areas of research, building on advances made in such fields as genomics, proteomics, chemical synthesis, materials science, and engineering. Understanding glycans and applying this knowledge can help find problem-driven solutions to a diverse set of challenges. Examples include the early detection of cancer and other diseases through identification of disease biomarkers, protection against infectious diseases such as influenza through increased understanding of the role of glycans in host-pathogen interactions and the immune response, and creation of products and fuels derived from carbohydrate raw materials.

Much of the fundamental biology and chemistry being explored in glycoscience has the ability to influence what are often viewed as disparate fields. Researchers in health, energy, and materials science can leverage discoveries in each other's disciplines to help strengthen the field as a whole. For example, efforts to understand the biochemical pathways of glycans and the roles of carbohydrate polymers inside cells are of use to scientists working to better understand cancer biology and plant biology alike. The conversion of biomass into novel starting materials can have implications for both materials scientists working to develop new plastics based on renewable resources or synthetic chemists working to synthesize novel drug targets. This report provides a holistic vision for glycoscience by suggesting a research roadmap for the scientific community that, while undoubtedly challenging, may ultimately help democratize the field and help realize the broad benefits from this important area. This roadmap will enable the tools to address glycoscience questions to be available to scientists and engineers who wish to incorporate them into their research. To address the roadmap goals, glycoscience will require input from researchers not currently working in this field and glycoscientists will need to reach out to bring these researchers into their community.

WHY NOW?

While genomics and proteomics have advanced rapidly, glycoscience and glycomics have made strides that are enabling scientists to better understand the role that glycans play in biological systems. Glycoscience researchers have already developed a fundamental knowledge base that can be utilized to help address many of today's major research problems. This knowledge base, when combined with the current set of tools available to probe glycan structure and function, is a powerful resource to better understand human, plant, and microbial biology.

Glycoscience has, until recently, been explored by a small group of experts, working with a more limited set of information and resources than are available in fields such as genomics and proteomics. What is known about glycoscience and glycomics, the study of the complete set of glycans in an organism, is still incomplete. But current knowledge now makes it possible to integrate glycoscience broadly into the fields of health, energy, and materials science, and the set of available tools, while not perfect, provides a base to enable further development and discovery.

A CENTRAL FIELD WITH LINKS TO MANY DISCIPLINES

Glycoscience is a highly interdisciplinary field that aims to better understand the structures and functions of glycans and how they can be used. It is a global field with a dedicated community of researchers in the United States and abroad. Glycoscientists do not have a single training/education background. They come from various fields, including **physiology** and **developmental biology**, where glycans are involved in processes such as cell movement and tissue development. They are in **medicine**, where glycans are involved in the development and progression of chronic and infectious diseases. In **microbiology,** glycans are key players in interactions among and between microbes and host cells. Glycoscientists are **chemists** developing new synthetic and analytical methods for glycans, and **biochemists** working to understand glycan synthesis and metabolism. In **materials science**, glycans can be used as polymeric materials having a wide range of properties. In **computational science** and informatics, modeling studies and the effective analysis of large amounts of experimental data are also necessary to better understanding glycans.

CONTRIBUTIONS TO IMPROVING HEALTH, DEVELOPING ALTERNATIVE FORMS OF ENERGY, AND CREATING NEW MATERIALS

This report focuses on three areas in which glycoscience can make significant contributions: health, energy, and materials science. The com-

mittee identified these three areas because they illustrate the diverse roles played by glycans and because glycoscience is relevant to researchers from a range of backgrounds. These focus areas demonstrate how improved understanding of glycans can make concrete impacts in society, particularly as part of the development of a bio-enabled innovation economy, as recently articulated by both the Organisation for Economic Co-operation and Development and the White House. This report does not address the roles of carbohydrates as food sources and nutritional supplements. Although these are also important areas to be explored, they are outside the scope of this study and outside the expertise of the study committee.

In human health, glycans are involved in myriad processes that are part of normal physiology, development, and cell signaling, along with the development of both chronic and infectious diseases. For example, glycans on cell surfaces are important in molecular recognition. One example of this function is their role in the movement of white blood cells through the body to a site of infection, enabling the immune system to respond where needed. Much of the information content in cells is encompassed in the glycome. Glycans contain key biological information that complements the information stored in DNA to help complete the link between genotype and phenotype or between the genome and expressed traits. Many advances in understanding human health and diseases are the result of current knowledge about nucleic acids, proteins, and glycans and how these vary in different circumstances and in different people. However, much is still unknown. Continued advances in understanding the biological roles played by glycans, along with the factors that influence or alter their functions, will have consequences for the fundamental understanding of biology and will contribute to the development of new therapeutic medicines.

Carbohydrates are fundamental to plant biology. Constituents of plant cell walls include glycans such as cellulose and hemi-cellulose combined in a matrix of other biopolymers. As society explores sources of energy that can provide alternatives to fossil fuels, harnessing the energy stored in these plant carbohydrates is one attractive option. Effectively converting plant glycans into liquid biofuels requires breaking down the structures of plant cell walls in order to release the constituent carbohydrate molecules for subsequent processing. Advances in understanding the glycans that comprise the cell wall, the enzymes that help assemble and degrade it, and how it can be altered to improve the degradation process can all make significant contributions to improving the feasibility of this energy source.

Glycans such as cellulose, starch, chitin, and others also provide the basis for creating new materials with useful physical and chemical prop-

erties. Such materials can take the form of bulk polymers or be processed into forms such as nanoparticles. In addition, other molecules can be attached to the glycans to alter the functional properties of the material or to affect how the polymer interacts in biological systems. These glycan-based materials provide potential substitutes for many petroleum-based plastics and have wide-ranging uses in medicine and industrial applications. For example, they can serve as carriers to encapsulate and deliver drugs and as scaffolds for tissue engineering, and they can be used in flexible coatings and films.

A TOOLKIT THAT INCLUDES MANY COMPONENTS BUT THAT ALSO HAS KEY GAPS

Because glycans are made of different types of individual sugar units linked in multiple ways, large numbers of different glycan structures can be created from the same constituent carbohydrate molecules. Unlike DNA and proteins, glycans are not created by following a sequence template but rather through enzymatic reactions that depend on several factors, including the concentrations of many different enzymes and many different substrates. The diversity of possible glycan structures makes them scientifically interesting. The large number of structures and the various ways in which glycans interact with other biological molecules create diversity beyond what can be encoded in an organism's genome alone. However, these characteristics also pose challenges to probing glycan structure and function and to being able to control and manipulate them in research.

The explosion in genetic research and understanding of gene functions that has occurred over the past 25 years was enabled by the development of new tools, such as high-throughput DNA sequencers and synthesizers. Tools to study DNA are now part of the repertoire of many biologists and chemists. Glycoscience, too, relies on a toolkit of techniques that enable key questions to be explored and answered. Although much can be accomplished by using existing tools, large gaps remain in such areas as the chemical and enzymatic synthesis of glycans and analytical techniques to determine glycan structures and functions. Glycoscience also lacks accessible, integrated, and well-annotated databases similar to those that exist for proteins and nucleic acids. New tools and techniques will be needed to enable glycoscience to live up to its potential to contribute to areas in health, energy, and materials. Creating these new tools and techniques will require engaging scientists and engineers from multiple disciplines who can bring new ideas and solutions to the field to help fill these identified gaps.

OVERARCHING FINDINGS

Glycoscience is a broad field, and the committee's findings capture only an overview of the information the committee considered in making its recommendations and developing a roadmap for the field. The findings are organized under the topics of human health, energy, materials science, and the toolkit needed to advance the field.

Health

- Glycans are directly involved in the pathophysiology of every major disease.
- Additional knowledge from glycoscience will be needed to realize the goals of personalized medicine and to take advantage of the substantial investments in human genome and proteome research and its impact on human health.
- Glycans are increasingly important in pharmaceutical development.

Energy

- Plant cell walls, made mostly of glycans, represent the planet's dominant source of biological carbon sequestration, or biomass, and are a potentially sustainable and economical source of non-petroleum-based energy.
- Understanding cell wall structure and biosynthesis and overcoming the recalcitrance of plant cell walls to conversion into feedstocks that can be transformed into liquid fuels and other energy sources will be important to achieving a sustainable energy revolution. Glycoscience research will be necessary to advance this area.
- Glycoscience can contribute significantly to bioenergy development by advancing the understanding of how to increase biomass production per hectare and how to increase the yield of fermentable sugar per ton of biomass.

Materials Science

- By fostering a greater understanding of the properties of glycans and of plant cell wall construction and deconstruction, glycoscience can play an important role in the development of non-petroleum-based sustainable new materials.

- Glycan-based materials have wide-ranging uses in such areas as fine chemicals and feedstocks, polymeric materials, and nanomaterials.
- There are many pathways to create a variety of functionalities on a glycan, creating a wide range of options for tailoring material properties.

Based on the above, the committee makes the following findings regarding the toolkit needed to advance glycoscience:

- Scientists and engineers need access to a broad array of chemically well-defined glycans.
- Over the past 30 years, tremendous advances have been made in chemical and enzymatic synthesis of glycans, but these methods remain relegated to specialized laboratories capable of producing only small quantities of a given glycan. For glycoscience to advance, significant further progress in glycan synthesis is needed to create widely applicable methodologies that generate both large and small quantities of any glycan on demand.
- A suite of widely applicable tools, analogous to those available for studying nucleic acids and proteins, is needed to detect, describe, and fully purify glycans from natural sources and then to characterize their chemical composition and structure.
- Continued advances in molecular modeling, verified by advanced chemical analysis and solution characterization tools, can generate insights for understanding glycan structures and properties.
- An expanded toolbox of enzymes and enzyme inhibitors for manipulating glycans would drive progress in many areas of glycoscience.
- A centralized accessible database linked to other molecular databases is needed to fully realize advancements in knowledge generated by an expanded effort in glycoscience. Glycan information is not currently accessible to the research community in an integrated and centralized manner similar to other biological information.

A ROADMAP TO ADVANCE GLYCOSCIENCE

Based on these findings, the committee makes the following recommendations in order to achieve a more complete understanding of glycoscience and to realize its impacts on health, energy, and materials science. Each recommendation is followed by a series of roadmap goals.

The capabilities created by the achievement of these recommen-

dations will ensure that all interested researchers can efficiently and effectively incorporate glycoscience into their work.

> **1. The committee recommends that the development of transformative methods for the facile synthesis of carbohydrates and glycoconjugates be a high priority for NIH, NSF, DOE, and other relevant stakeholders.**

Roadmap Goals

Within 7 years, have synthetic tools to be able to synthesize all known carbohydrates of up to octasaccharides, including substituents (e.g., acetyl, sulfate groups). This goal encompasses human glycoprotein and glycolipid glycans and proteoglycans, which are currently estimated to be 10,000 to 20,000 structures, along with plant and microbial glycans and polymers.

Within 10 years, have synthetic tools to be able to synthesize uniform batches, in milligram quantities, of all linear and branched glycans that will enable glycan arrays for identifying protein binding epitopes, provide standards for analytical methods development, and enable improved polysaccharide materials engineering and systematic studies for all fields to be conducted. This includes methods for synthesis of structures with isotopic enrichment of specific desired atoms that may be needed for a wide variety of studies.

Within 15 years, be able to synthesize any glycoconjugate or carbohydrate in milligram to gram quantities using routine procedures. Community access should be available through a web ordering system with rapid delivery.

> **2. The committee recommends that the development of transformative tools for detection, imaging, separation, and high-resolution structure determination of carbohydrate structures and complex mixtures be a high priority for NIH, NSF, DOE, FDA, and other relevant stakeholders.**

Roadmap Goals

Over the next 5-10 years, develop the technology to purify, identify, and determine the structures of all the important glycoproteins, glycolipids, and polysaccharides in any biological sample. For glycoproteins, determine the significant glycans present at each glycosylation site. Develop agreed upon criteria for what constitutes the acceptable level of structural detail and purity.

Within 10 years, have the ability to undertake high-throughput sequencing of all *N*- and *O*-linked glycans from a single type of cell in a single week.

Within 10 years, have the ability to routinely determine the complete carbohydrate structure of any glycan or polymer repeat sequence including branching, anomeric linkages between glycans, and substituents.

Within 15 years, have the ability to determine glycoforms (a complete description of molecular species within a population that have the same polypeptide sequence) of any glycoprotein in a biological sample.

For example, one specific achievable step could be to apply the tools developed in the roadmap to characterize the set of glycomes in blood, including those of blood cells and plasma.

3. The committee recommends that the development of transformative capabilities for perturbing carbohydrate and glycoconjugate structure, recognition, metabolism, and biosynthesis be a high priority for NIH, NSF, DOE, and other relevant stakeholders.

Roadmap Goals

Within 5 years, identify the genes involved in glycan and glycoconjugate metabolism in any organism whose genome has been sequenced, and identify the activities of at least 1,000 enzymes that may have utility as synthetic and research tools.

Within 10 years, be able to use all glyco-metabolic enzymes (e.g., glycosyltransferases, glycosidases) as well as other state of the art tools for perturbing and modifying glyco-metabolic pathways (knockouts, siRNAs, etc.) of utility to the biomedical and plant research communities.

Within 10 years, develop methods for creating specific inhibitors to any human, plant, or microbial glycosyltransferase suitable for in vitro and in vivo studies in order to perturb the biology mediated by these enzymes.

Within 15 years, develop imaging methods for studying glycan structure, localization, and metabolism in both living and non-living systems.

4. The committee recommends that robust, validated informatics tools be developed in order to enable accurate carbohydrate and glycoconjugate structural prediction, computational modeling, and data mining. This capability will broaden access of glycoscience data to the entire scientific community.

Roadmap Goals

Within 5 years, develop an open-source software package that can automatically annotate an entire glycan profile (such as from a mass spectrometry experiment) with minimal user interaction.

Within 5 years, develop the technology to perform computer simulations of carbohydrate interactions with other entities such as proteins and nucleic acids.

Within 10 years, develop the software to simulate a cellular system to predict the effects of perturbations in glycosylation of particular glycoconjugates and polysaccharides.

5. The committee recommends that a long-term-funded, stable, integrated, centralized database, including mammalian, plant, and microbial carbohydrates and glycoconjugates, be established as a collaborative effort by all stakeholders. The carbohydrate structural database needs to be fully cross-referenced with databases that provide complementary biological information (e.g., PDB and GenBank). Furthermore, there should be a requirement for deposition of new structures into the database using a reporting standard for minimal information.

Roadmap Goals

Within 5 years, develop a long-term-funded, centralized glycan structure database with each entry highly annotated using standards adopted by the community and all the world's repositories of glycan structures. The database should be cross-referenced and open source to allow the community to develop database resources that draw on this resource and improve its utility to investigators that wish to incorporate glycoscience in their work.

Within 5 years, employ an active curation system to automatically validate glycan structures deposited into a database so that journals can provide authors with an easily accessible interface for submitting new glycan structures to the database.

To achieve the roadmap goals articulated in its recommendations, the committee notes that it will be of critical importance for the field to reach agreement on the standards of evidence and the nature of the assumptions that will be used to annotate and validate glycan structures within the next 2 to 3 years.

Finally, the committee notes that there is widespread lack of understanding and appreciation of glycoscience in the scientific and medical communities and among the general public. Glycans are integral components of living organisms, whether human, animal, plant, or microbe,

and glycan products have applications in health, energy, and materials science.

> **The committee concludes that integrating glycoscience into relevant disciplines in high school, undergraduate, and graduate education, and developing curricula and standardized testing for science competency would increase public as well as professional awareness.**

Roadmap Goals

Within 5 years, integration of glycoscience as a significant part of the science curriculum would include glycoscience as both lecture materials and hands-on experiments or activities.

Within 10 years, glycoscience will be integrated and taught at every level wherever it is relevant to understand the scientific content. Competency in glycoscience could also be included in all standardized testing wherever relevant (for example, as part of the SAT and GRE Subject Tests, the MCAT, and Medical Board Exams).

CONCLUSION

Glycoscience is a vibrant field filled with challenging problems. It can make contributions toward understanding and improving human health, creating next-generation fuels and materials, and contributing to economic innovation and development. Now is the time for glycoscience to be embraced broadly by the research community. Drawing in members from the full spectrum of chemistry, biology, materials science, engineering, medicine, and other disciplines will be needed to address the technical challenges described here. Although these challenges are substantial and complex, the results of achieving these goals have the potential to impact science in exciting ways.

1

Introduction

1.1 UNDERSTANDING THE LANGUAGE OF LIFE: THE CENTRALITY OF SUGARS

Sugars (see Box 1-1) are everywhere. They are the foundation of all life on Earth. The most important biochemical process on Earth is photosynthesis—plants, algae, and other similar organisms using the energy in sunlight to combine carbon dioxide and water to make sugars. Many of the resulting sugars in plants end up as either starch or cellulose, both polymers of the sugar glucose. Such polymerized sugars—called oligosaccharides, polysaccharides, carbohydrates, or, generically, glycans—are the most abundant molecules on the planet. Cellulose is a polymer of glucose that provides the structural support for all plants and trees, as well as the raw material for clothing, paper products, and wood products. While humans cannot digest cellulose—it is an important part of the indigestible "fiber" in our diets—grazing animals can, and it serves as their major source of energy. Starch is another glucose polymer. It differs only subtly from cellulose, yet humans can digest it into its component glucose molecules, the central feedstock for our metabolic pathways. Human metabolism, and the metabolism of virtually all living things, harvests energy by breaking down glucose into water and carbon dioxide, which is then ready to undergo another round of fixation by photosynthesis.

Glucose is key to life, but it is also central to disease. Diabetes, for example, results when glucose is not properly controlled by normal metabolic mechanisms. High concentrations of glucose can result in organ damage, while low concentrations can lead to loss of consciousness and

BOX 1-1
Carbohydrate, Glycan, Saccharide, or Sugar?

Carbohydrate: A generic term used interchangeably in this report with sugar, saccharide, or glycan. This term includes monosaccharides, oligosaccharides, and polysaccharides as well as derivatives of these compounds.

Glycan: A generic term for any sugar or assembly of sugars, in free form or attached to another molecule.

Saccharide: A generic term for any carbohydrate or assembly of carbohydrates, in free form or attached to another molecule.

Sugar: A generic term often used to refer to any carbohydrate, but most frequently to low molecular weight carbohydrates that are sweet in taste.

sudden death due to inadequate energy. Diabetics must measure their blood sugar frequently to ensure proper glucose levels. Such measurements account for a significant number of the total number of diagnostic tests conducted each year in developed countries.

But glucose is not the only sugar molecule of importance to human health. Our cells carry complex sugars that comprise individual sugar molecules linked to one another in a multitude of ways. These complex sugars are usually referred to as glycans. Glycans are one of the four major classes of macromolecules—nucleic acids, proteins, and lipids being the other three—that are essential for life and are involved in every aspect of biology, medicine, and a number of practical applications. These other three classes often incorporate or rely on glycans for their activity—nucleic acids contain the carbohydrates ribose or deoxyribose, whereas proteins and lipids often require appended glycans for activity (glycoproteins and glycolipids, respectively). These structures, and combinations of these structures, contain information that is used for a wide variety of biological processes. Key facts about glycans and glycoscience are given in Box 1-2.

For example, one result of 3 billion years of evolution is that every cell of every organism is coated with a layer of glycans—the glycocalyx in animals or the cell wall in prokaryotes, plants, and fungi (see examples in Figure 1-1). The glycocalyx/cell wall contains high information content. On red blood cells the different sugars of the glycocalyx are responsible for the different blood groups—A, B, AB, and O (see Box 1-3). On cells of organs, these and other aspects of the glycocalyx can determine whether a

BOX 1-2
Important Facts About Glycans

General

1. Glycans are the most abundant family of organic molecules on the planet.
2. The potential information content of glycans vastly exceeds that of any other class of macromolecules.
3. Every living cell on the planet is covered with a dense and complex array of glycans. These glycans form the glycocalyx in many types of cells (such as in humans) and comprise the cell wall in others (such as plants). Some cells do not have a nucleus, but all have a glycocalyx or cell wall.
4. Every molecule, cell, or organism that interacts with a cell must do so in the context of the glycocalyx or cell wall.
5. The vast majority of cellular and secreted proteins are modified with glycans, which modify, alter, and/or control their functions.

Health

1. Elimination of any single major class of glycans from an organism results in death.
2. Every disease that affects humans significantly involves glycans.
3. A great majority of host-pathogen interactions involve glycans, via recognition, degradation, or molecular mimicry.
4. Most protein therapeutics must be glycosylated properly to be functionally effective.
5. Altered glycosylation is a universal feature of cancer and contributes to pathogenesis and progression.
6. Many vaccines are glycan based.

Sustainability

1. Glycoscience is one of the only fields that directly impacts both the pharmaceutical and energy industries.
2. The majority of solar energy trapped as cellular energy is converted to carbohydrates.
3. There are no other candidate classes of molecules that can solve our energy and materials needs.
4. Petroleum resources have finite lifetimes, but polysaccharide resources are continually being created with the sun's energy.
5. Nitrogen fixation in plants depends on carbohydrate signaling between bacteria and plant roots.

particular person in need of a heart, liver, or kidney transplant can receive an organ from a particular donor.

Indeed, cell surface glycosylation (i.e., the process by which cells create and display their glycocalyx) is as important to understanding life as is the genetic code, yet our understanding of the information contained in glycosylation is rudimentary at best. In large part this lack of knowl-

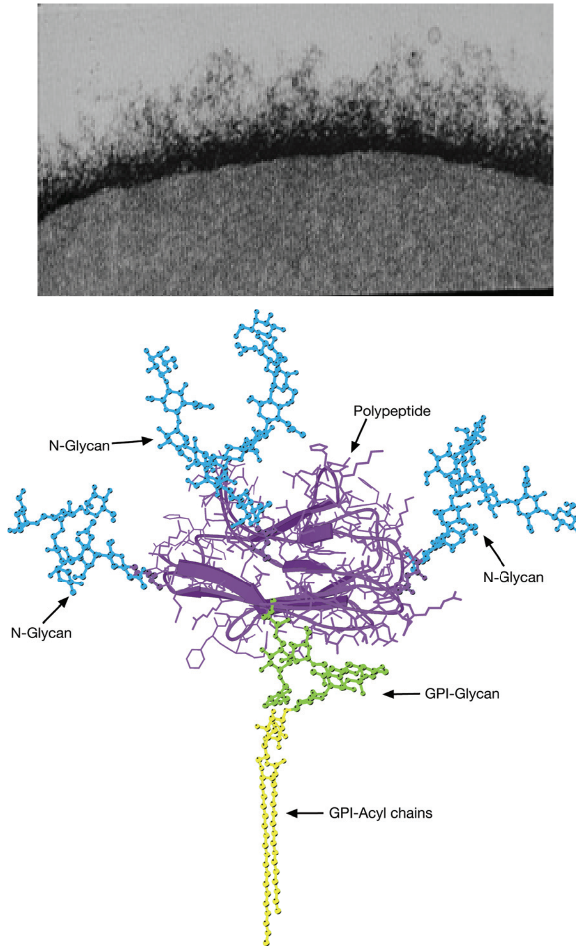
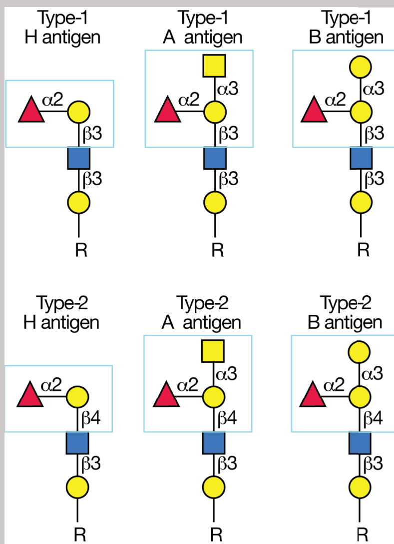


FIGURE 1-1 Glycans are significant components on biological surfaces and as parts of biological molecules. *Top*, Image of a red blood cell showing the glycocalyx extending from the membrane surface. SOURCE: Voet and Voet 2010, used with permission. *Bottom*, Scale model of a protein showing the relative sizes of the *N*-linked glycans and GPI-anchors that are attached to it. SOURCE: Varki et al. 2009, used with permission.

edge results from two factors: (1) the remarkable structural complexity of glycans found on cell surfaces and (2) a lack of tools for deciphering glycosylation patterns. Glycans thus got "left behind" in the initial phase of the modern revolution in molecular and cellular biology, resulting in a generation of scientists who may be largely unfamiliar with and untrained in the study of these key molecules of life.

BOX 1-3
ABO Blood Groups

One of the most familiar ways in which the glycan information of a cell influences phenotype is the ABO blood grouping, which is a significant factor in determining which blood transfusions can be carried out. With rare exceptions, human red blood cells contain on their surfaces a core carbohydrate sequence (called the "H antigen"). The familiar ABO blood types derive from further modifications to this H carbohydrate chain. In the genome, the locus that determines ABO type encodes for a glycosyltransferase. Different variants of this enzyme either are non-functional and therefore don't alter the H carbohydrate (type O) or add slightly different sugars to it (type A and type B; see image). Because a person receives DNA from both parents, the four possible blood types are O, A, B, and AB. Immune antibodies can form against the types of sugar chains that an individual does not have on his or her red blood cells. Thus, a person with type O blood may form anti-A and anti-B antibodies that prevent him or her from successfully receiving blood from anyone other than a similar type O donor. On the other hand, a person with both type A and type B carbohydrate chains will not form antibodies against either and can receive blood from any ABO source. As a caveat, it is important to recognize that the ABO system is not the only factor that determines transfusion acceptance and thus the above description is not absolute. For example, humans also have red blood cell proteins that influence transfusion acceptance (for example, Rh factor). However, the ABO system helps illustrate how small differences in glycans translate to practical, physiological differences. The possibility of modifying the surface glycans on red blood cells to avoid ABO incompatibilities is also being explored (Olsson and Clausen 2008; Liu et al. 2007).



Representation of ABO sugars on red blood cells. SOURCE: Varki et al. 2009, used with permission.

The complexity and high information content of glycans result from the many ways in which they can be assembled from simple sugar building blocks. This is in contrast to the simple ways that building blocks of proteins and nucleic acids—the amino acids and nucleotides, respectively—are linked together. Protein and nucleic acid biopolymers are linear, and every building block is linked to the next through the same kind of connection. By contrast, sugar building blocks can be linked together at many different sites and in different spatial orientations (i.e., stereochemistries), creating both linear and branched polymers with a wide variety of shapes (see Figure 1-2). Between the combination of structural diversity and different possible connection sites, the complexity of glycans increases rapidly. This diversity not only gives rise to many important and interesting biological functions and chemical properties but also creates challenges for synthesis, purification, and characterization—structure elucidation challenges discussed in detail later in this report.

The tools available today for fully characterizing the complex structures of glycans at low levels are mostly destructive, making it largely impossible to follow the changes in glycosylation that occur on a cell's surface over time. In addition, the diversity of glycan structures makes full characterization of the cell surface glycome (i.e., the totality of glycans with which a cell is coated) an incredible challenge, one beyond

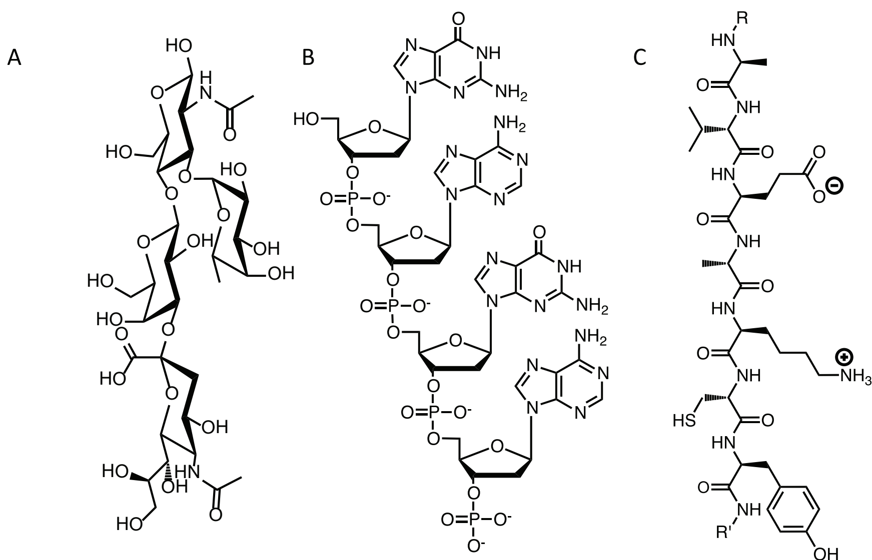


FIGURE 1-2 Comparison of nucleic acids, proteins, and glycans. A, glycan; B, nucleic acid; C, protein.

the capabilities of current technology. Today, it is possible to obtain only a general idea of the composition of the glycocalyx or cell wall, rather than a detailed molecular-level description. Yet these surface glycans are essential to both understanding and treating many diseases. The pattern of sugars on a cell causes pathogens—viruses and bacteria—to attack certain cell types. Many bacteria and viruses recognize specific sugars on particular cell types. In turn, a person's immune system generates antibodies to these invaders based largely on the glycans on these pathogens. Adding complexity, many pathogens carry out molecular mimicry of host glycans in order to evade immune responses. In addition, there is growing evidence that the glycans on cancer cells differ from those on normal cells, presenting a promising opportunity for diagnosis, imaging, and therapy. In addition to their roles on cell surfaces, glycans play important roles in biological communication and signaling (see Box 1-4).

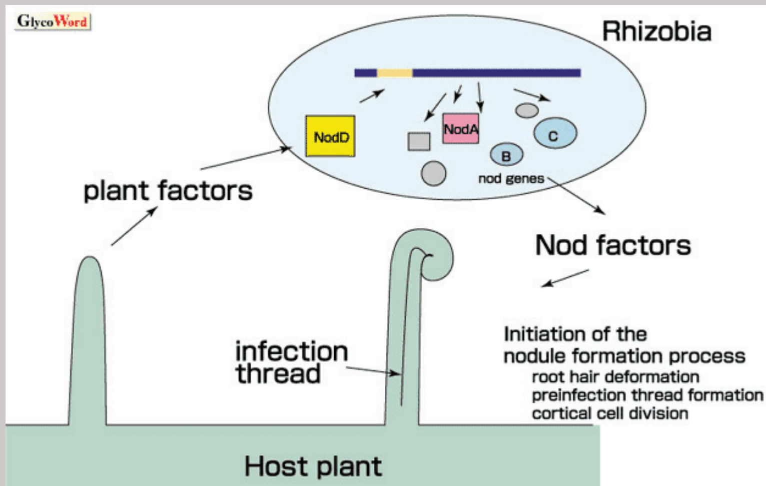
In the area of energy, sugars play an increasingly important role as scientific innovations drive advances in developing energy sources that will be renewable and contribute less to global climate change. Complex glycans, such as the starches and cellulose in plant cell walls (referred to as biomass), are Earth's primary storage location for the products of fixation of carbon into molecules via photosynthesis. These glycans are being exploited as renewable sources of liquid biofuels, such as ethanol. As described above, these materials ultimately can trace their energy content to the sun, so they can be thought of as a form of solar energy—and just as renewable. The challenge is to efficiently harvest the energy contained in the large amount of glycans produced by plants.

Glycoscience is uniquely poised to make significant contributions to this need. The polysaccharide components of the insoluble cell walls include cellulose, hemicelluloses, and pectins—polymers of sugars that are sometimes linear (cellulose) and sometimes branched (hemicelluloses and pectins). These walls have a generalized global structure, with cellulose embedded in a matrix of other molecules, although the fine details of wall structure differ across plant species, across different plant tissues and organs, and indeed across walls in single cells. A major challenge to plant glycoscientists is to understand how these cell wall components are biosynthesized and how they are put together with lignin to form insoluble plant biomass, as well as how to manipulate and break down biomass more effectively in order to release the sugars for development into fuels.

Glycans can also be used as important materials—for example, as gelling agents in foods—and as a renewable resource for high-value chemicals, plastics, and pharmaceuticals. Wood, comprised of lignocelluloses, is a major building material and is used in myriad applications. Other materials, such as most plastics, are derived primarily from petroleum. Glycans can play an important role either as a starting material to

BOX 1-4
Glycan Signaling in Nitrogen Fixation

Nitrogen is an essential element in biological systems and is a key component of proteins and other molecules. To be usable by most organisms, however, the nitrogen available in the atmosphere must first be fixed or converted into ammonium. Before the development of chemical fertilizers, all nitrogen fixation occurred biologically through the action of bacteria capable of undertaking these reactions. Biological nitrogen fixation remains a significant source of bioavailable nitrogen. Although several types of bacteria can fix nitrogen, one important example is the symbiotic relationship that exists between species of Rhizobia bacteria and the roots of legumes. Chemical signals (flavanoids) released by plant roots activate Nod genes in the bacteria. Turning on these genes leads to the production and release of a glycoconjugate called Nod factor that binds to receptors on plant root cells, leading to changes such as nodule formation and the ability of the bacteria to enter the root. Inside the root nodule the bacteria carry out the nitrogen fixing reaction. The symbiotic process depends on communication between bacteria and plant root through the Nod factor, which is an acylated chitin oligosaccharide molecule that includes lipid and carbohydrate components. This familiar example highlights one of the many ways in which glycans play key roles in biological signaling.



Communication between plant and bacteria during the process of nitrogen fixation. SOURCE: http://www.glycoforum.gr.jp/science/word/saccharide/SA-A02E.html; accessed June 12, 2012.

the same types of feedstocks that are presently obtained from petroleum or as alternative materials that can be converted directly into plastics with similar or even superior properties to those of today's synthetic materials. As the ability to engineer polysaccharides and tailor their chemical structures and properties advances, the capacity to design new biochemicals and materials with properties that are unachievable today also will greatly expand.

1.2 GENES AND PROTEINS ARE NOT ENOUGH: THE RICH INFORMATION CONTENT OF GLYCANS

The current view of information flow in biological systems starts with the nucleic acid genome, which codes for proteins that function as parts of networks and whose own roles are still being actively studied. After proteins have been assembled, they are nearly always modified—a process generically called posttranslational modification. The terminal stage in this information flow is often the addition of glycans to proteins (glycosylation), which modulates the proteins' activity. One way of looking at this process is that the instructions in the genome encodes the properties that will ultimately be observable in an organism (phenotype), whereas the proteome predicts the phenotype. The glycome, however, is the phenotype. The system can also be compared to a switchboard, with the sugars being the "on" and "off" switches or turn pots that modulate the functions of glycoproteins and other molecules and help control the activity of the network. Beyond this digital view of biology, glycans also serve major analog functions, allowing modulating ranges of functions of glycoproteins and other molecules as well as metabolic circuits and networks. Working backward to understand biological systems will require starting with glycobiology, just as working forward requires starting with genomics.

Unlike nucleic acids and proteins, the structures of glycans are not "hard-wired" in the genome. Because of the multiple linkages that sugars can engage in that produce isomers and branching patterns, glycan structures cannot accurately be described as simple linear sequences of building blocks. Rather, a glycan's most basic structure must be described in three dimensions. Because glycan structures are not template encoded, they are plastic, reflecting myriad factors determined by cellular metabolism, cell type, developmental stage, nutrient availability, other cues from the cell's environment (Rudd and Dwek 1997; Varki et al. 2009), and stochastic events. As a result, the potential information content of glycosylation is far greater than for all the other types of posttranslational protein modifications combined. It is precisely this enormous diversity and plasticity that are critical to the many biological functions of glycans,

particularly their modulation of glycoprotein activity or localization and their roles in mediating cell-cell or cell-matrix interactions that are key to both normal physiological development and diseases such as cancer.

1.3 HOW GLYCOSCIENCE BUILDS ON GENOMICS AND PROTEOMICS

Today, the glycoscience field is at a place similar to where genetics was at the conception of the Human Genome Project. At that time there was enough of an understanding of genetics to know that a concerted effort to sequence the human genome would lead to both fundamental advances in our understanding of genetics and practical applications that would benefit all fields of science. When this enormous effort began in the 1990s, many scientists questioned if it was even feasible to sequence the 3 billion bases in a human genome. Ten years and $2 billion later, the Human Genome Project not only had sequenced a single human genome but had also spawned a technological revolution that today makes it possible to sequence a human genome in only a week at a cost of $1,000. Similarly, the cost of identifying a single nucleotide polymorphism (SNP), a commonly used marker for genetic traits such as disease, fell from $1 per SNP to $0.004 per SNP, opening the door to a wide range of biological questions inconceivable even 10 years ago.

Another impact of the Human Genome Project has been the democratization of genomics. The result is a revolution in our understanding of genetics that spans the simplest single-celled organisms to the characterization of human variation and disease. Sequencing instruments used to be huge and expensive, and, as a result, sequencing was done only at regional centers. Today, sequencing instruments can sit on a benchtop in any laboratory. Now, any laboratory can get DNA sequenced; computer programs can predict structures from sequences for DNA, RNA, and proteins; and DNA or RNA can be ordered online and delivered the next day.

How did all of this happen in such a short period of time? The transformation of genomics, and the generation of an entire new industry, started with the research community issuing a grand challenge that was a huge leap, something beyond any technical capability available at the time. In the end, *the tools that were developed to meet this grand challenge now enable and drive the science.* The tools of genomics have democratized the field in such a way that thousands of laboratories are now able to ask and address questions that were previously the realm of only a few specialized facilities. Any scientist interested in getting sequence information can do so. Today, because of incredible success at developing sequencing tools, the real cost of sequencing a genome is dominated by informatics, not by

the physical process of sequencing. Making sense of genomic data costs far more than acquiring the data.

Glycoscience needs to similarly catalyze its transformation from the realm of a few specialists to a core science practiced by many. To accomplish this transformation, new technologies are needed to thoroughly characterize glycomolecules and synthesize them. Both genomics and proteomics have methods for automated synthesis, sequencing, and amplification. The emerging field of glycomics does not. There are large libraries of genes and proteins available for study but only small libraries of glycans and glycoconjugates. Genetic manipulation of genes and proteins is easy but is hard for glycans and glycoconjugates. Finally, the number of enzymes available for manipulating genes and proteins is far larger than the number of glycosidases and glycosyltransferases available. Learning from the experience of genomics, glycomics will need many new and sophisticated informatics solutions to stay abreast of technological developments and avoid the bottlenecks that now limit the advances that come from modern genomics and proteomics.

1.4 WHY NOW? THE CASE FOR CHANGE

To fully understand the workings of living organisms and to fully realize the promise of genomics and proteomics, it will be imperative that science now turn its efforts to deciphering the complexity of glycomics. Unless attention is paid to glycans, a major component of biology will be missed. Glycoscience cannot be overlooked. Without a better understanding of the glycome, a clear understanding of cancer, infectious diseases, and the immune response will not be possible. Glycoscience knowledge will be similarly needed in the exploration of improved biofuels and alternative sources of carbohydrate-based energy and in the development of carbohydrate-based materials with functional new properties. It will not be possible to take full advantage of the revolution in genomics and realize the full potential of the Human Genome Project unless close attention is given to glycomics and how cells make and use the myriad complex glycans that decorate their surfaces. At the same time, advances in genomics resulting from the Human Genome Project provide a major opportunity to understand how mutations alter glycan pathways with functional consequences. Indeed, the time is right for the glycoscience community to initiate an undertaking that leads those conducting biological studies to seriously consider incorporating glycoscience into their work.

Several recent advances make now the time to examine challenges and opportunities in glycoscience and outline a possible roadmap forward. In health, for example, changes in glycosylation are common in tumor cells and specific glycans have been identified as biomarkers for

a variety of cancers (Adamczyk et al. 2012). In some cases, this information is being combined with array technologies to provide a base from which to explore key questions in cancer biology. Do particular glycosylation changes play a role in cancer outcome? Which glycans can serve as the most effective biomarkers for different stages and different types of cancer?

In 2011, the U.S. Department of Energy released an update to the Billion-Ton Study, which re-emphasized the significance of biomass feedstocks from non-food crops for energy and materials (DOE 2011). Many of the energy-rich, non-food crops require the conversion of recalcitrant cellulose into useful chemical precursors. Discoveries in the biological pathways by which plant cell walls are synthesized and deconstructed are similarly providing a compelling base from which to further advance the applications of glycoscience to these fields.

Just as studies of nucleic acids and proteins rely on a suite of tools that allow a broad range of researchers to effectively investigate these molecules, so too does glycoscience rely on its own toolkit. Over the past decade, developments in synthetic and analytical methods such as glycan microarrays are enabling high-throughput analysis of the interactions of glycans with proteins, lipids, and other glycan molecules (Rillahan and Paulson 2011). These data are increasingly being combined into glycan databases, to share and aggregate research results within the glycoscience community (Frank and Schloissnig 2010).

Genomics and proteomics have advanced rapidly. Glycoscience and glycomics also have made strides in enabling scientists to understand the role that glycans play in biological systems. Glycoscience researchers have been developing a fundamental knowledge base that can be utilized to help address many of today's major research problems. This knowledge base, when combined with the current set of available tools to probe glycan structure and function, is a powerful resource to better understand human, plant, and microbial biology.

Glycoscience has, until recently, been explored by only a small group of experts, working with more limited information and resources than are available in fields such as genomics and proteomics. What is known about glycoscience and glycomics, the study of the complete set of glycans in an organism, is still incomplete. But the knowledge currently available now makes it possible to integrate glycoscience broadly into the fields of human health, energy, and materials science, and the set of tools, while not perfect, provides a base to enable further development and discovery.

1.5 CHARGE TO THE COMMITTEE

Recognizing that glycoscience presents a frontier for discoveries across many fields, the National Institutes of Health, Food and Drug Administration, U.S. Department of Energy, and National Science Foundation asked the National Research Council to convene a committee to explore advances in glycoscience and challenges that must be overcome to move the field forward. The committee was also tasked with articulating a roadmap and a vision for future development of the field (see Box 1-5).

The committee deliberated at three in-person meetings and held numerous teleconferences to address its charge and produce the present

BOX 1-5
Statement of Task

The National Research Council of the National Academy of Sciences will convene an ad hoc committee to assess the importance and impact of glycoscience and glycomics. Glycoscience is the confluence of scientific disciplines that study complex glycans and their relationships to other molecules. Glycans are involved in all phases of life, and an improved understanding could significantly impact diverse sectors of society, including health and energy. While genomics and proteomics have produced unparalleled discoveries that have advanced the understanding of biological processes, the picture these present is incomplete. Glycoscience and glycomics, the systematic analysis and characterization of the structure and function of glycans synthesized by a cell, tissue, or organism, could be a critical next step in building on genomics and proteomics, linking gene function to an observed phenotype, and decoding the molecular makeup of an organism.

In order to realize the potential of glycoscience and glycomics to build on genomics and proteomics and forge major new roads of discovery, the National Research Council of the National Academy of Sciences will convene an ad hoc committee to:

- Conduct an in-depth analysis of the current state of research in glycoscience and glycomics in the U.S.;
- Compare current U.S. and international research efforts in glycoscience;
- Discuss key challenges to the growth and development of the field of glycoscience and glycomics;
- Develop a roadmap with concrete research goals to significantly advance glycoscience and glycomics in the U.S., including the identification of metrics that may be used to help assess efforts to achieve these goals and objectives; and
- Articulate a unified vision for the field of glycoscience and glycomics.

The ad hoc committee will conduct workshops and other data-gathering activities to inform its findings and conclusions, which will be provided in the form of a consensus report.

report. In addition, the committee convened the Workshop on the Future of Glycoscience in January 2012, which brought together approximately 75 glycoscientists and scientific thought leaders with expertise in biology, chemistry, and materials science to discuss the field and its opportunities and needs. The workshop agenda and participant list are provided in Appendix C. The committee also solicited input from the broader scientific community through its public website, which included several questions to inform the study process. These questions are provided in Appendix D, along with further information on the feedback received and the individuals who shared their thoughts with the committee. This report does not focus on the roles of carbohydrates as food sources and nutritional supplements. Although these are important areas to be explored, they were outside the scope of the committee's study and outside the expertise of the committee's members.

1.6 ORGANIZATION OF THE REPORT

Chapter 2 discusses current glycoscience research efforts in the United States and worldwide. This general baseline helps inform the rest of the report, which lays out a vision for the future of the field. The chapter provides a brief overview of key messages arising from the committee's data gathering, with further details and examples included in Appendix B. In Chapter 3 the committee discusses how glycoscience is embedded in the key areas of health, energy, and materials science—areas that help illustrate the breadth and impact of glycoscience as a discipline. In Chapter 4 the committee poses a set of scientific questions and opportunities designed to illustrate more concretely how new glycoscience knowledge would contribute to answering relevant scientific questions in these fields. These questions are not meant to be comprehensive but rather to provide examples of scientific challenges that, if solved, would yield important basic and applied knowledge. Chapter 5 considers the toolkit for glycoscience in such areas as synthesis, analysis, and informatics. These tools are integral to studying glycoscience and will be needed to successfully address the types of challenges described previously. Finally, Chapter 6 presents the committee's conclusions and recommendations. In conjunction with each recommendation, the committee suggests several 5- and 10-year goals whose accomplishment would significantly advance the field. Together, these goals comprise a roadmap to help enable glycoscience to forge new roads of discovery.

The introductory and concluding chapters of this report are written with a general audience in mind. Chapters 3 and 4, which delve more deeply into the myriad ways that glycans contribute to the three focus areas of health, energy, and materials, presume a basic level of scientific

familiarity, although of necessity do not cover each topic in detail. Chapter 5, which describes the current scientific toolkit for studying glycans, is written largely for the scientific community and for those who have primary responsibility for shaping research programs and directions. The committee's assessment of this toolkit and of the needs and gaps remaining to advance the field is encapsulated in the report's concluding chapter, which lays out a glycoscience roadmap and research goals. Appendixes to the report contain committee member biographies (Appendix A) and additional information on the committee's data-gathering efforts (Appendixes B, C, and D). A glossary of terms also is included (Appendix E).

2

The Landscape of Current Research in Glycoscience

As a starting point to inform its deliberations, the committee sought to better understand the current landscape of major U.S. and international glycoscience efforts. This chapter presents a brief overview of the committee's findings in order to provide a baseline of current investments in the field and a sense of centers of research activity in the United States and abroad. Examples and further details on U.S. and international glycoscience programs are included in Appendix B.

Although it did not undertake an exhaustive survey to identify U.S. and international glycoscience efforts, the committee reviewed information provided to it by federal sponsors,[1] received community input through its website and through a workshop held in January 2012,[2] gained additional perspectives through further data-gathering efforts,[3]

---

[1] Representatives of the National Institutes of Health (NIH), Food and Drug Administration (FDA), U.S. Department of Energy (DOE), and National Science Foundation (NSF) briefed the committee on their motivations in sponsoring the study and their views on challenges and opportunities for glycoscience at the committee's first meeting on October 10, 2011.

[2] For the workshop's agenda and participants, see Appendix B. Information on the study's website (http://glyco.nas.edu) and the questions that members of the community were invited to address can be found in Appendix C.

[3] Committee members spoke with several additional scientists to gather information on current glycoscience research outside the United States; information can be found in Appendix C.

conducted a Web of Science review of published literature,[4] and drew on a background paper prepared by the National Research Council (NRC) that summarizes a range of federal agency and researcher viewpoints on the field (McGowan and Bowman 2010).[5] These materials provided an overview of the current landscape of glycoscience research efforts and informed development of the committee's roadmap.

2.1 AN OVERVIEW OF GLYCOSCIENCE WORLDWIDE

Glycoscience research is conducted worldwide in projects that cut across multiple disciplines. As can be seen from Figures 2-1 and 2-2, active glycoscience research is ongoing not only in North America (the United States and Canada) but also in Asia (People's Republic of China, Taiwan, Japan, South Korea, India—and Australia), in many countries in Europe, and in Latin America (Brazil).

A number of U.S. federal agencies support or conduct glycoscience research, including NIH (through multiple individual institutes), NSF, DOE, FDA, U.S. Department of Agriculture (USDA), and National Institute of Standards and Technology (NIST). These agencies have complementary interests in the field, including the application of glycoscience for human health and in support of therapeutic drug and vaccine development (NIH, FDA, NIST), the application of glycoscience to plant biology (DOE, USDA), and advancing basic science understanding and fundamental tool development for the field (NSF, NIST, NIH, and others). Additional details and further examples are provided in Appendix B, but one notable federally funded initiative is the Consortium for Functional Glycomics, which currently receives legacy funding from NIH and involves the participation of hundreds of researchers worldwide. The efforts of participating research groups have made available a range of resources for addressing questions in glycoscience and health, including

[4] A search of the Web of Science (WOS) Citation Index Expanded Database was conducted on May 15, 2012, using the following parameters: Topic: glycoscience* OR glycan* OR carbohydrate* OR *cellulos* OR glycobiolog* OR *saccharide*; years: 2005-2012; publication type: articles, meeting abstracts, and proceedings. The search produced 127,602 results.

[5] The background paper was prepared at the request of NIH, which asked the NRC to reach out to researchers and federal program managers for their views on the state of glycomics and glycoscience and challenges facing the field, in order to better understand how to frame the design of the current study. The paper summarizes information received during this outreach, in which NRC staff and a small group of glycoscience experts spoke with approximately 40 scientists and program managers from government, academia, and industry. The paper was not reviewed per the NRC's report review procedures and does not necessarily reflect the views of the NRC or its boards. The information it contained did help provide background material for the current study, particularly on the landscape of U.S. research efforts.

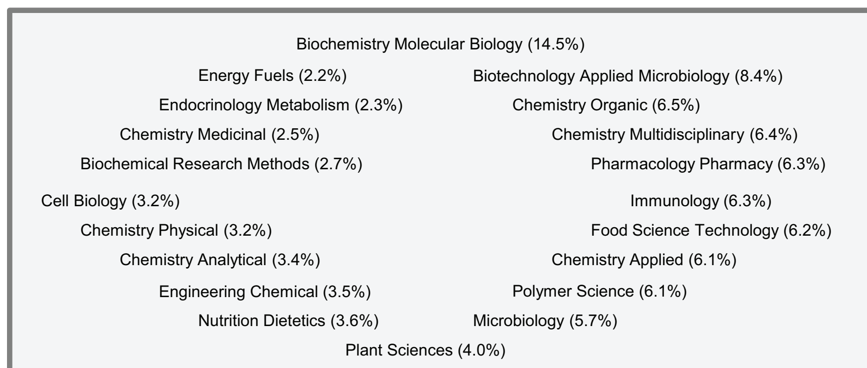


FIGURE 2-1 Glycoscience research spans a diversity of fields, as indicated by the Web of Science subject categories associated with published research. Results from the WOS citation search described above were sorted by WOS subject category, and the top 20 subject areas are depicted above.

resources for glycomics profiling, carbohydrate compounds and reagents, microarray analysis, mouse phenotyping, glycan array screening, and glycan databases.

Similarly, a number of U.S. research programs and clusters of research expertise were identified during the committee's data-gathering process. These span the country and may involve multiple researchers, providing a concentration of expertise across different aspects of glycoscience. Although many more examples are provided in Appendix B, one example of a center of excellence for glycoscience research in the United States is the Complex Carbohydrate Research Center (CCRC), located at the University of Georgia. The CCRC includes a cluster of centers that address plant, microbial, and human carbohydrates, along with research resources in areas such as nuclear magnetic resonance analysis and computational modeling.

Glycoscience research is also conducted across the globe in projects that cut across disciplines. Although not described here, examples of research activities and investments from Canada, the United Kingdom, Germany, Japan, China, Taiwan, Australia, New Zealand, and Brazil are provided in the appendix. Glycoscience also has significant relevance to companies invested in the development of protein-based biotherapeutics or vaccines containing carbohydrate antigens (discussed in Chapter 3). Other companies interested in glycoscience include those that use carbohydrate-based materials for drug delivery and tissue engineering applications, those interested in the development of cellulose-based biofuels and products derived from agricultural sources and by custom chemical producers. Representative examples are provided in Appendix B.

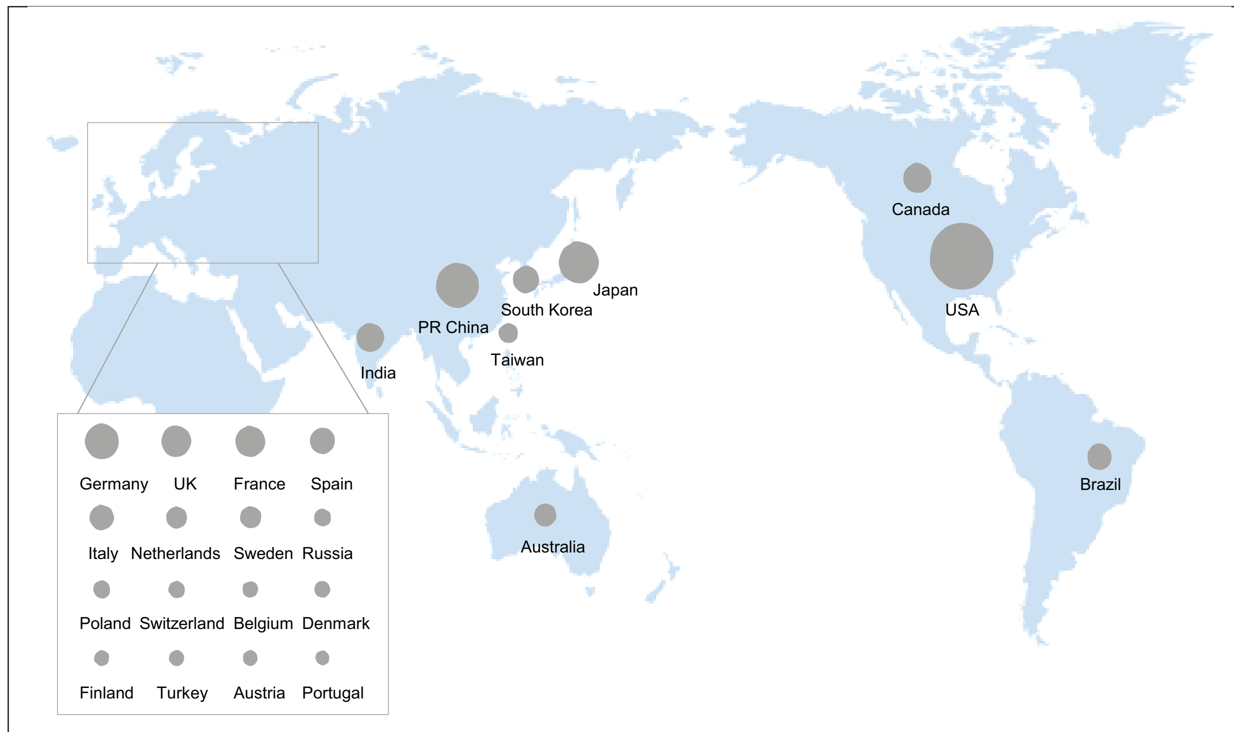


FIGURE 2-2 Glycoscience research occurs worldwide, with centers of activity around the globe. Results from the WOS citation search described above were sorted by country of author, with the top 25 countries depicted in the figure. The size of the bubbles represents the relative proportion of glycoscience papers published. Underlying world map adapted from David Niblack. Found at: http://imagebase.davidniblack.com.

In addition to the many programs and collaborations identified by the committee, several scientific forums bring together members of the glycoscience community both in the United States and internationally. These include scientific meetings such as the annual Society for Glycobiology conferences; biennial Gordon Research Conferences on Glycobiology, on Glycolipid and Sphingolipid Biology, and on Plant Cell Walls; the biennial international carbohydrate symposium organized by the International Carbohydrate Organization); and the biennial meeting by the International Glycoconjugate Organization (IGO). The long-standing Annual San Diego Glycobiology Symposium involves glycoscientists from throughout California and regularly attracts participants from all over the world. Other scientific forums include the biennial Charles Warren workshops on glycoscience characterization and analysis and the Beilstein Symposia on Glyco-Bioinformatics.

2.2 AN "OMICS" FIELD—GLYCOSCIENCE IN ITS INFANCY

The citation search in the WOS database undertaken by the committee was intended to be broad so as to include research on cellulose and other carbohydrate polymers as well as glycoconjugates. By this measure the number of papers published annually in the overarching field of glycoscience is similar to those published in genomics or proteomics alone[6] (see Figure 2-3). Glycomics,[7] however, clearly remains in its infancy, with annual publications several orders of magnitude lower. When the Human Genome Project was initiated in 1990, genomics publications also were substantially lower in number than they are today.[8] Over the ensuing decades a massive expansion of gene-sequencing capabilities and a decrease in costs have occurred (see Figure 2-4). Enabling genome analysis to advance to its current state required integrated efforts across the scientific community, including both international collaborations (such

[6] A WOS search was conducted on May 15, 2012, using the following parameters: (A) A "Glyco (all terms)" search was conducted as identified above, except that Years: 1995-2011, producing 253,658 total results. (B) Topic: genom*; Citation database: Citation Index Expanded; Years: 1995-2011; Publication type: article, meeting abstract, and proceeding, producing 285,067 total results. (C) Topic: glycom*; Citation database: Citation Index Expanded; Years: 1995-2011; Publication type: article, meeting abstract, and proceeding, producing 1,624 total results. (D) Topic: proteom*; Citation database: Citation Index Expanded; Years: 1995-2011; Publication type: article, meeting abstract, and proceeding, producing 45,370 total results.

[7] Analogous to genomics (the study of the full set of nucleic acid genetic material) and proteomics (the study of the full set of proteins), glycomics involves comprehensive study of the full set of glycans present in a cell or an organism.

[8] A WOS search for the term "genom*" similar to that described above produced 5,700 results for the year 1990.

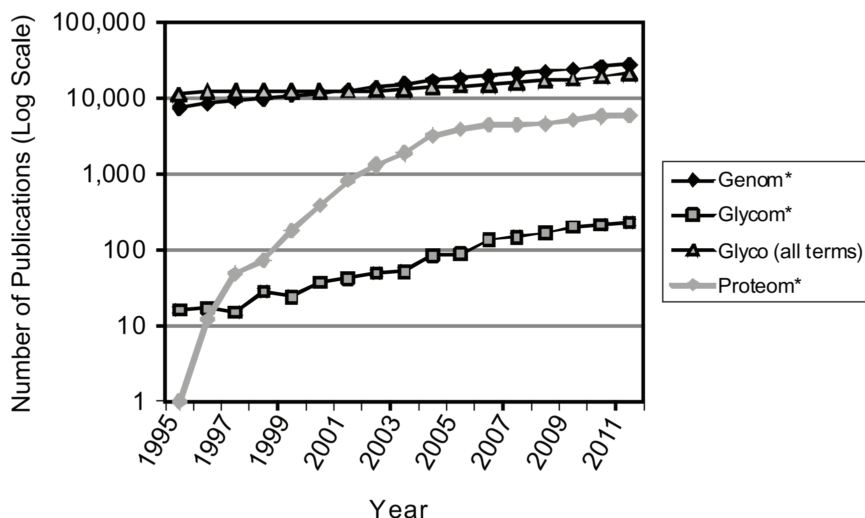


FIGURE 2-3 Comparison of WOS search results for annual genomics, proteomics, glycomics, and glycoscience research.

as the Human Genome Organization, HUGO), and the contributions of many individual scientists. It is beyond the committee's charge to propose a similar formal glycoscience initiative comparable to that of the Human Genome Project or the National Nanotechnology Initiative. Rather, the committee seeks to describe the current status of the field of glycoscience and explore its potential, while clearly recognizing that advancing the field in a similar dramatic fashion, as with genetics and genomics, would require engagement by and the efforts of multiple stakeholders beyond the current community of glycoscience specialists.

2.3 COMMON CONCERNS AMONG U.S. AND INTERNATIONAL GLYCOSCIENTISTS

During its data-gathering efforts, the committee did not observe significant differences among the viewpoints shared by U.S. and international researchers. Although individual scientists might vary in the challenges or opportunities they choose to highlight, several common themes emerged as being of fundamental importance to the field. Many of the technical challenges that make up the core toolkit to enable the next generation of glycoscience discoveries are addressed later in this report and are reflected in the committee's recommendations for a roadmap to advance the field. This section focuses on several other significant messages:

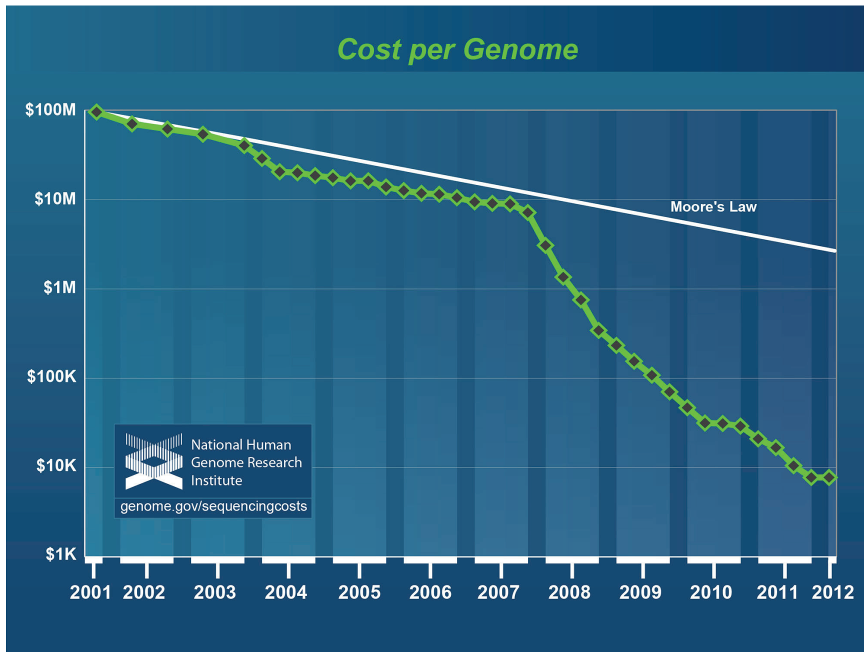


FIGURE 2-4 The rapid decline in the cost of sequencing a human genome.
SOURCE: Wetterstrand 2012; Image courtesy of The National Human Genome Project. Found at: genome.gov/sequencingcosts.

- *Visibility and vision of the field*: Both U.S. and international glycoscientists noted that relevant work taking place in a variety of disciplines may not be labeled with the term "glycobiology" and thus may not be well recognized as falling under the glycoscience umbrella. The field would benefit from having a clear picture to present to nonexperts, as well as compelling goals behind which the community could rally. Because glycoscience provides a level of data that can build from and complement genomic and proteomic information, inclusion of glycoscience components in international Human Genome Organization (HUGO) and Human Proteome Organization (HUPO) projects could help draw out these connections.[9]

[9] One example is the Human Disease Glycomics/Proteome Initiative (HGPI) through HUPO. HGPI investigates glycosylation changes in efforts to identify possible biomarkers relevant to the diagnosis or progression of disease (see http://www.hupo.org/research/hgpi/).

- *Education and awareness*: Both students and peer researchers who are not experts in glycoscience lack an understanding of why glycoscience is significant, they lack a comprehensive view of what the glycoscience field encompasses, and they do not see how glycoscience relates to their own interests. Researchers expressed concerns about the limited coverage of carbohydrates in academic programs when compared to classes of molecules such as nucleic acids and proteins and about how the field suffers from a perception that it is too complex to study effectively or is not exciting.
- *Critical role of collaborative approaches*: Many glycoscience challenges are likely to benefit from synergistic efforts that bring communities of people together to address problems from different perspectives. As the field seeks to advance to the next level of discoveries, there will be a need to foster collaboration and understanding between, for example, clinicians and laboratory researchers, between biologists and chemists, and between computational/informatics experts and experimental scientists.

2.4 CONCLUSION

The landscape of glycoscience research provides a picture of a global field with a range of ongoing research efforts, both academic and commercial, and one for which the community sees significant opportunities as well as common challenges. The following chapters of this report attempt to present a holistic view of glycoscience's contributions to critical areas such as human health, energy, and materials science; to bring new attention from both experts and nonexperts to the field; and to point the way toward a roadmap and a vision for the future of the field.

3

Glycoscience in Health, Energy, and Materials

Glycoscience contributes in fundamental ways to three key areas on which the committee focused: the understanding of human health and disease, the search for alternative sources of energy, and the development of new materials. The committee selected these areas because they illustrate the range and diversity of research encompassed by glycoscience as a field. They also help illustrate how glycoscience knowledge will be embedded in efforts to address fundamental challenges in health and sustainability.

The chapter begins with examples and questions related to human health because this has been a major focus of efforts in the field of glycoscience and glycomics, particularly in the United States. Indeed, many scientists may automatically think of health when they think of glycans and their functions. Although other researchers actively study carbohydrates and their uses (e.g., in polymer engineering), the terminology and techniques used by these fields may vary. As a result, the scientific community may not immediately think of the totality of glycan research as part of a unified field of glycoscience. One goal of this report is to provide a view of glycoscience that encompasses a broader range of topics. Indeed, although health care remains an important driver for research, increased attention is being paid to other drivers, including the environment and energy security (Johnson 2012), and glycoscience will be relevant in multiple contexts.

3.1 GLYCOSCIENCE AND HEALTH

Over the past several decades, research from many laboratories has established that glycans are directly involved in normal physiology and in the etiology of every major disease afflicting mankind (Varki et al. 2009). Deciphering the glycome creates an expanding frontier for knowledge and discovery about human health. The section begins with an explanation of the roles of glycans in fundamental biological processes, such as inflammation and immune system activation, and moves on to consider examples from infectious diseases and vaccine development. It then turns to chronic diseases such as diabetes and cardiovascular disease and to a discussion of cancer and congenital genetic disorders. Finally, the significance of glycans in the development of new pharmaceuticals is discussed. Examples of the diverse roles that glycans play in human health are provided to illustrate the breadth and importance of glycoscience to this field. The section does not attempt to comprehensively address all glycan functions. As it illustrates, however, the development of a more complete understanding of glycans can impact the diagnosis and treatment of infectious, chronic, and genetic diseases. (See Figure 3-1 for a partial summary of some of the roles played by glycans in biological systems.)

3.1.1 Glycans' Regulation of Inflammation

Inflammation, both chronic and acute, underlies the pathology of a broad range of diseases, including diabetes, cancer, arthritis, asthma, heart disease, and infectious disease (Barreiro and Sanchez-Madrid 2009; Celie et al. 2009; Kobayashi et al. 2009a; Korpos et al. 2009; Langer and Chavakis 2009; Schauer 2009; Sperandio et al. 2009; McEver 2010; McEver and Zhu 2010; Sorokin 2010; Zarbock et al. 2011). Glycans play a key role in inflammation at many levels.

Inflammation begins with the generation of multiple cytokines by various cell types that react to pathogen-associated molecular patterns (PAMPs) or danger-associated molecular patterns (DAMPs) arising from damaged tissues. Many of these molecular patterns are glycoconjugates, and many cytokines themselves bind to endogenous glycans. More recently recognized is the fact that glycans found in an individual host, such as a human, can serve as self-associated molecular patterns (SAMPs) that dampen inflammation and that SAMPs can be mimicked by microbes (mSAMPs; Varki 2011). The signaling associated with these molecular patterns is part of the multistep process that results in leukocyte homing into affected tissues, a process initiated when leukocytes adhere to activated endothelial cells lining blood vessel walls. This Velcro-like adhesion slows the rapidly flowing leukocytes, causing them to roll along the surfaces of the endothelial cells. Rolling leukocytes are

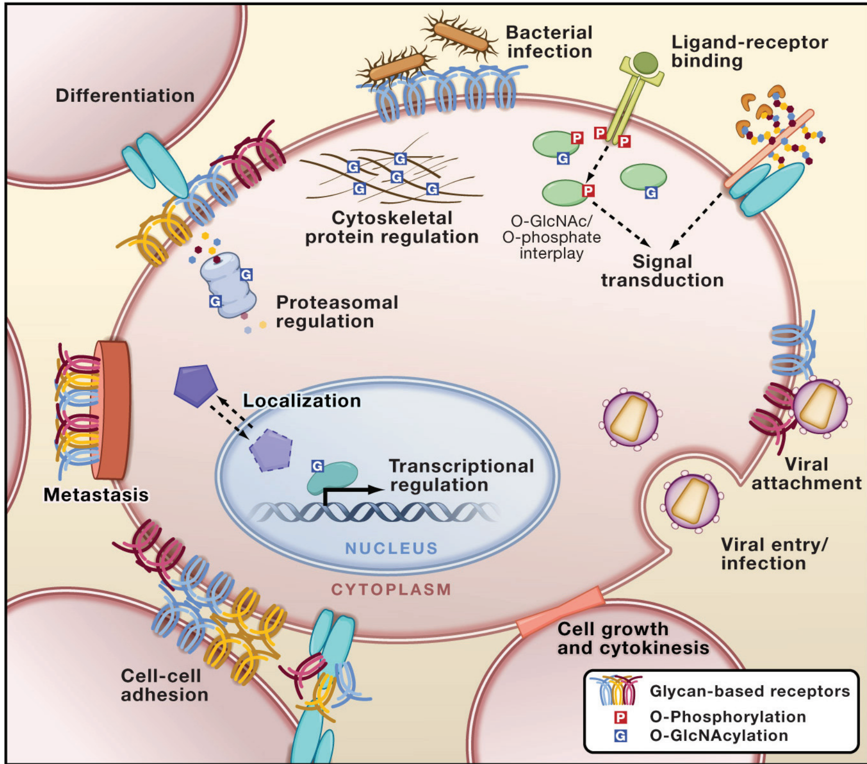


FIGURE 3-1 Glycans play diverse roles within biological systems.
SOURCE: Reprinted from Hart and Copeland 2010, with permission from Elsevier.

able to bind tightly to glycoprotein receptors, called integrins, which can lead to penetration of the endothelial cell monolayer and its basement membrane. The mechanisms for the initial binding and rolling of leukocytes have been studied extensively and involve transient expression of highly regulated glycan-binding proteins, called selectins, that decorate the surfaces of leukocytes, activated platelets, and activated endothelial cells. Selectins are exquisitely specific at binding to certain glycan structures. One of the primary determinants for their binding specificity is the tetrasaccharide known as sialyl Lewis x. The unique properties of the interactions between a selectin protein and its specific glycan ligand are critical to slowing leukocytes down so that they can bind to and then extravasate into inflamed tissue. Selectins and their glycan ligands also play a role in tumor metastasis (Dube and Bertozzi 2005; Laubli and Borsig 2010; St Hill 2011).

Glycoconjugates such as collagens, laminins, and sulfated proteoglycans that surround endothelial and other cells also play critical roles in immune cell infiltration of tissues. Leukocytes secrete hydrolases, enzymes that degrade extracellular matrix glycoconjugates and release bioactive glycan-containing fragments in the extracellular milieu. These glycan-containing fragments help perpetuate inflammation by affecting leukocyte chemotaxis, activation, and differentiation. For example, fragments of the polysaccharide hyaluronan are pro-inflammatory as a result of their ability to bind to a class of receptors called Toll-like receptors and serve as "danger signals" of acute cell injury or infection.

The steps in inflammation that are controlled by glycans and glycoconjugates represent potential novel targets for therapeutics that could improve on current treatments like broad-acting steroids. In addition, advances in medicinal chemistry now allow for the rational design of a class of drugs—glycomimetics—that are based on the bioactive conformations of functional glycans (Imberty et al. 2008; Ernst and Magnani 2009; Magnani and Ernst 2009; Garber et al. 2010; Chabre et al. 2011; Drozdova et al. 2011; Jandus et al. 2011). For example, cell adhesion mediated by selectins underlies the vascular occlusion crises that characterize sickle cell anemia. One company has recently designed a small-molecule antagonist that binds to selectins and is now in clinical trials as a treatment for this disease. Other companies are investigating the use of antibodies directed against a selectin and its ligand. Meanwhile, the already approved drug heparin is known to block inflammation by blocking selectin interactions. As a result of these advances, the study of functional glycans represents a source of leads for novel therapeutics for treating inflammation and a variety of other human diseases. The key roles played by glycans in inflammation and the trafficking of white blood cells to tissues also helped stimulate interest in understanding the physiological importance of glycans in cell adhesion and cell signaling more broadly.

3.1.2 Glycans' Essential Role in Regulation of the Immune System

Inflammation is one result of the immune system's response to danger. Immunity is the other. Antibodies themselves are glycosylated proteins, and gycans can also be the targets (antigens) for antibody binding and the generation of immune responses.

The general importance of glycans in immunity has been appreciated for many years, and the early discovery that the ABO blood groups derive from specific glycan structures is just one example (Morgan and Watkins 1969). While in most cases the glycans that form part of antibodies do not play a direct role in their binding to antigens, they do play a critical role in their effector functions to activate components of the immune

system (Raju 2008; Lux and Nimmerjahn 2011). Mammals have developed sophisticated glycan recognition systems to recognize pathogen-associated molecular patterns, including specific types of glycan-binding proteins such as lectins, and Toll-like receptors. In fact, the most common phenotype in mice that survive the selective deletion of glycosyltransferases is defective immune cell function (Marth and Grewal 2008). In humans, dendritic cells play a primary role in the presentation of foreign antigens to the immune system (Erbacher et al. 2009). The glycans on dendritic cell surfaces and multiple lectins are involved in dendritic cell functions, including antigen uptake, immune modulation, detection, processing of viral antigens, and trafficking (Silva et al. 2012). Meanwhile, host glycans such as sialic acids serve as self-associated molecular patterns. It is becoming evident that many of these glycan-protein interactions are involved in the critical balance between immune tolerance and the generation of a strong immune response.

Some human autoimmune diseases involve auto-antibodies that recognize self-glycans. The glycans on a pathogen may be sufficiently similar to a human glycan such that an immune response to the pathogen leads to the generation of cross-reactive antibodies. For example, this is thought to be the case in the development of Guillain-Barré syndrome and Miller Fisher syndrome (Kaida et al. 2009). The significance of antiglycan antibodies is also being explored in other autoimmune diseases, such as systemic lupus erythematosus and rheumatoid arthritis (Dai and Gao 2011; Fattal et al. 2010; Louthrenoo et al. 2010).

The specific branching pattern of *N*-glycans on T-cell antigen receptors regulates their threshold of activation, and a deficiency of a key *N*-glycan branching glycosyltransferase known as GnT5 may contribute to autoimmune disease (Lee et al. 2007). During their biosynthesis, B- and T-cell antigen receptors are glycosylated in a manner reflective of the cell's physiological state. This glycan plasticity alters their molecular interactions at the cell surface, associations with signaling complexes, and internalization via endocytosis. These glycans also appear to control the spatial organization of receptors laterally in the plasma membrane. Galectins, which are multivalent glycan-binding proteins, and their glycan ligands have numerous roles in varied immune processes, including pathogen recognition, regulation of inflammation, and modulation of the adaptive immune response (Rabinovich and Toscano 2009). There are currently more than 15 different galectins known in the human genome, each with varied glycan-binding specificity and distinct cellular distributions. Specialized galectins also modulate the threshold of T-cell receptor activation during T-cell development (Demotte et al. 2008). In addition to their role in regulating immune functions, galectins mediate cellular interactions with parasites, viruses, bacteria, and fungi. Recent studies

have indicated that glycan-galectin lattices on the surfaces of immune cells modulate receptor signaling and play a role in modulating effector functions (Rabinovich et al. 2007).

Another class of glycan-binding proteins—siglecs—also have critical functions in immunity. Siglecs are membrane-bound, sialic-acid-binding, immunoglobulin-like, glycan-binding proteins that play a key role in regulating immune cell adhesion, signaling, and endocytosis (Crocker et al. 2007; Crocker and Redelinghuys 2008). Sialic acids are negatively charged monosaccharides that often appear at the terminus of glycan structures. Siglec interactions with immune cell receptors at the cell surface help inhibit abnormal immune cell activation. Siglecs are important for preventing autoimmunity, and they influence the responses of almost every cell in the immune system. There are currently 17 known siglecs encoded in primate genomes, each with a different immune cell type distribution and function. The types of sialyl-oligosaccharides and the structures of the sialic acids on the surfaces of immune cells also play key roles in the activity of siglec regulation of immune cell functions. Given the very recent finding of sialic acids as self-associated molecular patterns that help regulate these immune reactions, it is reasonable to suggest that other self-glycan patterns might yet be discovered.

For high-affinity binding, siglecs require the glycans to be clustered in specific patterns that are not well understood. Critical to their functions is the ability to bind sialic acid glycans both on the same cell (*cis*) and on a different cell or microorganism (*trans*). The molecular mimicry of host sialo-glycans by a variety of pathogens and beneficial microorganisms takes advantage of siglecs. As is the case for the selectins, siglec-specific agonists and antagonists represent potentially powerful, but as yet untapped, targets for the development of therapeutics to treat autoimmune and inflammatory diseases. Meanwhile, siglecs already serve as targets for delivering chemotherapeutic agents to specific cell types. For example, CD33, also known as Siglec-3, is the target of an antibody approved for treatment of acute myeloid leukemia, and antibodies targeting CD22/Siglec-2 are in clinical trials for treatment of B-cell non-Hodgkin's lymphoma and autoimmune diseases (O'Reilly and Paulson 2009).

Glycans in the nucleus and cytoplasm also play a critical role in the regulation of immunity. Recent studies have shown that O-GlcNAcylation, a ubiquitous monosaccharide modification of nuclear and cytoplasmic proteins (Hart et al. 2011), plays a key role in both T- and B-lymphocyte activation (Golks et al. 2007). O-GlcNAc transferase is required for early activation of B-lymphocytes via the B-cell receptor. Data suggest that O-GlcNAcylation is required for nuclear translocation and functions of key transcription factors regulating B-lymphocyte activation and functions.

Although there is still much to be learned about the roles of glycans in immunity, new insights could yield important advances in treating a wide variety of human diseases (Kolarich et al. 2012).

3.1.3 Glycans' Key Role in Infectious Diseases and Vaccine Development

While glycans are important in regulating immunity, they are also key actors in the constant battle between our cells and invading pathogens, including viruses, bacteria, and parasites. Indeed, glycans are the dominant molecules at this interface. Glycans and glycan-binding proteins, in part because of their plasticity and rapid evolution, play a critical role on both sides of this battle in nearly every species of pathogen (Bardoel and van Strijp 2011). Not only are glycans commonly used by microbes and viruses to bind to and infect host cells but also nearly all of the vaccines for infectious diseases recognize glycans present on the disease-causing organism. The complex mucin-bound glycans lining the epithelial surfaces of the human body not only block invasion by pathogens but also provide binding sites essential to colonization by beneficial bacteria that reside in our bodies and are required for our survival.

Many, if not most, bacteria have adhesins on their surfaces that bind to cells via glycans. These protein-glycan interactions often determine the tissue selectivity of bacterial pathogens (Pieters 2011). For example, lung and airway pathogens, such as *Pseudomonas aeruginosa*, *Haemophilus influenza*, and *Staphylococcus aureus*, primarily recognize glycans terminating in GalNAcβ1-4Gal structures. Recent studies of the bacteria responsible for gastric ulcers (*Helicobacter pylori*) indicate that these bacteria bind sialylated glycans, such as those found on mucins and gangliosides in the stomach (Kobayashi et al. 2009b). Conversely, the innate immune system of humans, which is a major line of defense against pathogens, has evolved to deal with millions of species of bacteria, fungi, and viruses primarily by recognizing their foreign glycoconjugate structures (Bardoel and van Strijp 2011). For example, the lipid A component of bacterial lipopolysaccharide and the complex mannan structures on fungi are particularly potent elicitors of an immune response. The C-type lectins made by our cells are an important component of our innate immunity in that they recognize a wide variety of glycans on pathogens.

Research has confirmed only recently that the complex glycans present in human milk play a role in protecting newborns from infections and represent a major form of innate immunity (Newburg et al. 2005). The variety of different glycan structures in human milk is enormous, and recently developed glycomic methods are beginning to elucidate the human milk glycome (Chichlowski et al. 2011; Tao et al. 2011). Milk

glycans serve as soluble receptors for pathogens, in much the same way that glycans on epithelial mucins function to inhibit pathogens from binding to the mucosal surface of the gastrointestinal tract. These and other findings suggest that more detailed knowledge about milk glycans might lead to novel antimicrobial agents to prevent, rather than treat, infections.

Because all human cells are covered with a thick glycocalyx, nearly all pathogens, including viruses, must gain entry to their target cells by interacting with glycans. The importance of glycans in influenza infection has been known since the 1940s (Karlsson 1998). In recent years, however, the critical roles of specific glycans in viral infections have been highlighted by fears of a new influenza pandemic. The first step in flu virus infection is the binding of a viral coat glycoprotein—hemagglutinin (HA)—to glycan structures on the host cell. Small mutations in HA enable it to bind to differently shaped glycans on a cell (e.g., to a human cell glycan rather than to a bird cell glycan; see Box 3-1). Thus, a remarkably small change in the ability of a protein to bind to a specific linkage of a single monosaccharide on cell surfaces can have a huge effect on society.

BOX 3-1
Pandemic Influenza

Four major human pandemics—in 1918, 1957, 1968, and 2009—were due to influenza viruses from birds and swine crossing into the human population, causing widespread disease because of a lack of preexisting immunity. Concern about pandemics from new viruses, such as the highly pathogenic H5N1 avian flu virus, has drawn increased attention to the potential for influenza to cross species' barriers. The designation H5N1 is a classification based on two proteins on the surface of the virus that interact with host receptors. The H stands for hemagglutinin, which attaches the virus to sialic acid receptors on cells. The N stands for neuraminidase, which cleaves sialic acids to allow release of newly formed virus from the infected cell. The neuraminidase is the target for current antiinfluenza medicines such as Tamiflu and Relenza, which blocks the cycle of influenza replication.

Over 60 years ago influenza virus was found to bind to sialic acids on host cells (Karlsson 1998). It is now known that the hemagglutinin of avian flu viruses recognizes sialic acid receptors that differ from those recognized by human viruses. Avian viruses recognize 2-3 linked sialic acids found in susceptible cells in birds, while human influenza viruses recognize 2-6 linked sialic acids, which are found in human airway cells (Stevens et al. 2006; Viswanathan et al. 2010). This small difference is crucial for the transmission of influenza virus in humans. As a result, receptor specificity represents a barrier for transmission of new animal viruses to humans (Tumpey et al. 2007). Owing to the importance of receptor specificity, it is now tracked by the Centers for Disease Control and Prevention as a risk factor for the emergence of new human pandemics from animal influenza viruses.

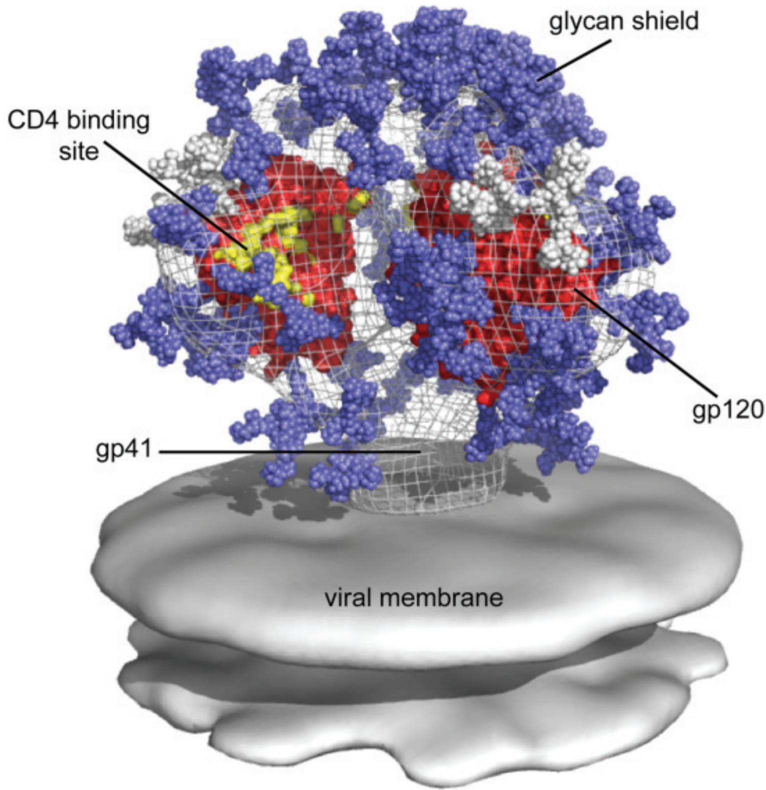


FIGURE 3-2 The glycan shield of HIV. Glycans (blue) cover the HIV gp120 protein, which is responsible for virus binding to the CD4 receptor on T cells. The binding and entry of HIV result in infection of the T cells and, ultimately, in immunodeficiency.
SOURCE: William Schief, The Scripps Research Institute, used with permission.

The human immunodeficiency virus (HIV) that causes AIDS has had an enormous impact on human health worldwide. The HIV coat protein (Env) is among the most heavily glycosylated proteins known. As with many viruses, the glycans are assembled using the host glycosylation machinery (Raska and Novak 2010). The HIV virus uses this "glycan shield" to prevent attack by the human immune system (see Figure 3-2). However, recent findings show that humans immune to HIV produce antibodies that bind to the glycan shield and neutralize infection by most HIV strains (McLellan at al. 2011; Pejchal et al. 2011; Walker et al. 2011). These insights are providing new hope for developing vaccine strategies to prevent the disease. The glycans covering pg120 obscure it from recog-

nition by the host immune system. The most potent antibodies identified to date that neutralize HIV infection are those that bind to the glycan shield.

Glycan-containing vaccines were first reported as early as 1929 (Tillett and Francis 1929). Some of the most effective vaccines against infectious organisms are directed toward glycans, and successful glycan-based vaccines include those against *Haemophilus* influenza and *Streptococcus pneumoniae*; others are in development. As of 2010, more than 30 glycan-based vaccines were in preclinical and clinical trials (Astronomo and Burton 2010). Recent advances in glycomics, glycan synthesis, glycan arrays, and methods for structural determination have resulted in a quantum leap in glycan vaccine development (Seeberger and Werz 2007; Huang and Wun 2010; Lepenies and Seeberger 2010). Advances in synthetic glycan chemistry are also allowing researchers to create fully synthetic vaccines, a development that may eliminate some safety concerns (Huang and Wun 2010). Still, many issues and challenges remain, such as the identification of epitopes on glycans as a function of a pathogen's life cycle, stimulation of both humoral and cellular immunity without triggering tolerance, and the design of antigen presentation to generate high-avidity neutralizing antibodies.

More than 15 million deaths occur annually because of parasitic diseases. Even though the immune responses to parasites are almost always directed against their unusual glycans (Nyame et al. 2004), there are no effective vaccines against major parasitic diseases such as malaria, trypanosomiasis, schistosomiasis, and amebiasis. Given the success of glycan-based vaccines against bacteria and the growing knowledge with respect to novel lineage-specific glycans in parasites, this area also represents a promising target for future vaccine development

While only a few examples are given here, it is clear that glycans play a central role in our battle against invading organisms of all types. As antibiotic resistance continues to rise, and the need for antivirals and antiparasitics becomes acute, more focused efforts to understand the central roles of glycans in infectious disease and vaccine development will be increasingly important.

3.1.4 Glycans' Multifaceted Role in Cardiovascular Disease

Cardiovascular disease is the leading cause of death worldwide (Nieuwdorp et al. 2005; Broekhuizen et al. 2009). It has recently become clear that the glycocalyx of vascular endothelial cells plays a critical role in the etiology of cardiovascular disease. This glycocalyx is comprised of membrane glycoproteins, proteoglycans such as syndecans, and associated glycosylated plasma proteins. The molecules hyaluronan and hepa-

ran sulfate glycosaminoglycans are major components of the endothelial glycocalyx. Normally, these glycans protect the vasculature from damage, but disruption or damage of the endothelial glycocalyx contributes directly to the onset of atherogenesis. The endothelial glycocalyx regulates important enzymes such as nitric oxide synthase and superoxide dismutase and serves as a barrier to macromolecules. In noninflammatory states it also prevents the adherence of platelets and leukocytes (Nieuwdorp et al. 2005).

Diabetes, another chronic disease of increasing prevalence and concern, is a major cause of atherosclerosis. One current model suggests that abnormal expression of proteoglycans or glycosaminoglycans in diabetics contributes to the binding of cholesterol-rich lipoprotein particles at sites in the vascular endothelium. Localized inflammation results in infiltration of macrophages mediated by selectins, which then take up the lipoprotein particles to become foam cells, leading to a plaque that eventually occludes the blood vessel (Tannock and King 2008). In contrast, heparan sulfate proteoglycans play an important role in the clearance of lipoprotein particles from circulation by the liver (Stanford et al. 2009). Nuclear and cytoplasmic protein glycosylation by the molecule O-GlcNAc also plays an important role in diabetic cardiomyopathy (Clark et al. 2003; Jones 2005; Fulop et al. 2007). Recent studies show that the contractile machinery in the heart is heavily modified by O-GlcNAc and that both the modification and association of O-GlcNAc cycling enzymes are strikingly increased in diabetes (Ramirez-Correa et al. 2008). Both the glycocalyx of endothelial cells and the modification of contractile machinery and transcription factors regulating the expression of key cardiac proteins represent novel targets for therapeutic discovery.

3.1.5 Glycans and the Molecular Mechanisms of Chronic Diseases

The prevalence of chronic diseases such as diabetes and Alzheimer's disease is on the rise, and glycans appear to play critical roles in the etiology of these and other chronic illnesses. Hyperglycemia and hyperlipidemia are the root cause of the biochemical events leading to the morbidity and mortality associated with diabetes, and glycans are involved in regulating a variety of cellular processes. Within the nucleus and cytoplasm, modification of proteins by glycans is surprisingly abundant (Hart et al. 2007; Copeland et al. 2008; Hart et al. 2011). Throughout the cell a system for glycosylating proteins with the monosaccharide O-GlcNAc has extensive cross talk with protein phosphorylation to regulate many cellular processes in response to nutrients and stress, including transcription, signaling, and cell division. Recent data from several laboratories indicate that hyperglycemia-induced increases in O-GlcNAc modification

of regulatory proteins affect insulin signaling (Teo et al. 2010) and are a major mechanism of glucose or lipotoxicity in diabetes. This cellular toxicity results from hyper-O-GlcNAcylation's effects on signaling as well as from its regulation of gene expression, in which extensively modified transcription factors do not function properly (Clark et al. 2003; Kudlow 2006; Solomon et al. 2008). There is also growing evidence that increased O-GlcNAcylation of mitochondrial proteins in diabetes plays a role in hyperglycemia-induced cellular toxicity (Hu et al. 2009).

Research has shown that elevated free fatty acids or a high-fat diet cause nuclear exclusion and loss of expression of key transcription factors regulating the expression of a glycosyltransferase, called GnT-4a, in the beta cells of the pancreas (Ohtsubo et al. 2011). This loss of transcription factor trafficking and expression is likely related to O-GlcNAcylation, but this possibility has not yet been investigated. Nonetheless, GnT-4a activity is critical for proper glycosylation and expression of the glucose transporters that are key to pancreatic beta cell function. These studies not only serve as another example of how subtle changes in glycosylation can have an impact on human disease in unexpected ways but also suggest a novel target for the possible prevention of beta cell destruction in type 2 diabetes.

Another glycan implicated in diabetes is hyaluronan, a large polysaccharide belonging to the family of glycosaminoglycans made by all cells (Wang et al. 2011). This glycan's synthesis is exquisitely sensitive to hyperglycemia via both the levels of the donor nucleotide sugars, such as UDP-GlcNAc, required for its synthesis and hyaluronan synthase's regulation by O-GlcNAcylation (Jokela et al. 2011). The abnormal expression of hyaluronan in diabetes likely plays a critical role in inflammation and in extracellular matrix abnormalities in diabetes, particularly in diabetic nephropathy. Hyaluronan and hyaluronan-binding proteins are directly involved in inflammation, tissue injury and repair, release of cytokines, and cell migration in lung diseases, kidney disease, brain injury, and heart disease (Jiang et al. 2011a).

Other studies have shown that dysregulation of O-GlcNAcylation plays a major role in neurodegenerative diseases of aging, such as Alzheimer's disease (Dias and Hart 2007; Lazarus et al. 2009). Recent investigations have shown that O-GlcNAcylation plays an important role in brain and neuronal functions, including learning and memory (Tallent et al. 2009; Skorobogatko et al. 2011; Rexach et al. 2012), and pharmaceutical companies are exploring drugs that increase O-GlcNAcylation as possible therapeutics for late-onset Alzheimer's disease (Yuzwa et al. 2012). The glycan polysialic acid also plays a role in the central nervous system. It is attached to the neural cell adhesion molecule and is involved in such functions as cell adhesion and signaling, which can affect neuron

growth, synaptic plasticity, and leaning. The roles of polysialic acid and neural adhesion molecules in such disorders as Alzheimer's disease and schizophrenia are also being explored (Kochlamazashvili et al. 2010).

3.1.6 Glycans' Roles in Cancer Progression and Early Detection

If heart disease is the leading cause of death in the world, cancer is number two and gaining ground. Again, glycans appear to play a central role in this complex set of diseases. If nuclear and cytoplasmic glycosylation is included, it is likely that greater than 80 to 90 percent of all proteins are covalently modified by glycans (Hart and Copeland 2010). Altered or abnormal protein glycosylation is universal in cancer cells of all types (Hakomori 2002; Varki et al. 2009), making such glycans potentially important cancer biomarkers (Reis et al. 2010). Indeed, many current clinical biomarkers in both tissue and serum assays are based on glycans. Much current research, such as that sponsored by the National Cancer Institute's Alliance of Glycobiologists for Detection of Cancer (http://glycomics.cancer.gov), has already recognized the value of glycans as biomarkers for the early detection of cancer. In addition, recent advances in glycan synthetic chemistry and chemical biology, as well as the development of novel analytical tools, have produced a convergence of proteomics and glycomics to develop improved biomarkers in this area (Drake et al. 2010).

Many of the oldest and most widely used clinical tests for cancer, such as those for carcinoembryonic antigen (colon cancer), prostate-specific antigen (prostate cancer), and cancer antigen 125 (ovarian cancer), rely on the detection of glycoproteins. All of these glycoproteins have both altered polypeptide expression and altered glycosylation in cancer. The concept of using specific glycoforms of glycoproteins to improve the sensitivity and specificity of biomarkers has already been validated by the use of core-fucosylated alpha-fetoprotein as a marker of hepatocellular carcinoma (HCC; Li et al. 2001). Elevation of this specific glycoform in HCC is due to increased activity of a specific fucosyltransferase by HCC cells. While today's Food and Drug Administration (FDA)–approved assays rely on detecting only a single biomolecule, it is likely that the true diagnostic power of cancer-specific glycoforms will result from multiplexing several glycoprotein biomarkers.

It has been suggested that abnormal expression of specific glycosyltransferases themselves might serve as useful biomarkers for cancer (Meany and Chan 2011). In fact, certain glycosyltransferases, such as GlcNAc-T5 and GALNT5, have been linked directly to the etiology of cancer and are potential targets for therapy. There are numerous examples in which simple overexpression of a glycosyltransferase leads to tumorigen-

esis (Dennis et al. 1999; Meany and Chan 2011). Thus, improvements to existing biomarker assays for nearly all types of cancers will likely involve the specific cancer-associated glycans attached to the polypeptide as well as the polypeptide expression. By combining data analyses from genomics, proteomics, metabolomics, and glycomics with pathway analyses, it should be possible to make great strides in cancer biomarker discovery as well as to reveal novel therapeutic targets (Adamczyk et al. 2011).

Many tumor-associated antigens were discovered by nonchemical methods and were later determined to be glycans (Heimburg-Molinaro et al. 2011). Adult tumor cells often display glycan structures that are normally found only on specific cells of embryos. There is even an instance in which a diet-derived nonhuman sialic acid Neu5Gc is incorporated into tumor cell glycans as a "xeno-autoantigen," driving chronic inflammation via interactions with "xenoautoantibodies." Over many years, glycan tumor-associated antigens have been studied and used as potential cancer vaccines (Li et al. 2010). It has been suggested that as more is learned about glycomics and about how the immune system recognizes cancer-associated glycans, effective development and use of glycan vaccines against cancers may come to fruition.

Altered glycosylation also underlies many of the properties of the cancer cell, including the propensity to metastasize. Tumor metastasis is a complex process requiring adhesive interactions, many of which are mediated by cell surface glycans and lectins, particularly the selectins (Rambaruth and Dwek 2011). *N*-glycan branching on the epithelial cell adhesion molecule E-cadherin (Pinho et al. 2011) and selectin ligands play a role in critical tumor cell interactions with platelet and endothelial selectins. Laminin and the hyaluronan-binding protein CD44 play a role in invasion and migration of tumor cells through connective tissues. Lectins and heparan sulfates, another class of glycans, play a role in both angiogenesis and in interactions with vascular endothelial cells. The glycocalyx of tumor cells plays a key role in masking surface antigens and evading immune surveillance mechanisms. While it is clear that cell surface glycans and glycan-binding proteins contribute to all stages of cancer progression and metastasis, and thus may provide novel targets for therapy, there is still much to be learned about the specific roles of glycans in cancer.

Nuclear and cytoplasmic glycosylation also plays a role in carcinogenesis by directly affecting signaling and transcription processes that contribute to cell growth and neoplastic transformation (Slawson et al. 2010; Slawson and Hart 2011). In fact, every nuclear oncogene or tumor suppressor protein examined to date is dynamically O-GlcNAcylated, and every cancer type examined to date has strikingly elevated O-GlcNAcylation on many different proteins. Cancer cells also have increased levels of the

enzymes that control O-GlcNAc cycling in cells. Nonetheless, the roles of O-GlcNAcylation in cancer remain a relatively unexplored area.

3.1.7 Critical Roles of Glycans in Human Development

Genetic diseases of both glycan biosynthesis and glycan degradation clearly establish the critical roles of glycans in human development. Currently, there are some 4,500 identified genetic disorders, with the biochemical cause unknown for 2,700. It is estimated that 2 percent of the genome encodes glycosylation-related genes. As a result, many genetic diseases whose etiology is unknown today may prove to be tied to glycosylation. In fact, research over the past 15 years has identified more than 65 glycosylation disorders (Hennet 2012), many of which did not have defects in known genes. These disorders encompass all known glycosylation pathways but saturate none of them, which suggests that there are likely more glycosylation disorders yet to be identified. The value of research aimed at identifying these disorders is shown by the development of enzyme replacement therapies for enzymes that act on glycans. Gaucher's disease, which results from defective degradation of glycosphingolipids, and mucopolysaccharide storage diseases that result from defects in degradation of glycosaminoglycans are examples of glycan-processing disorders now successfully treated with enzyme replacement therapy (Butters 2007).

Congenital diseases of glycosylation (CDGs) can result from mutations in almost all aspects of the glycosylation machinery, including the biosynthesis and transport of nucleotide sugars, proteins controlling vesicular trafficking in the endoplasmic reticulum and Golgi apparatus, and glycosyltransferases (Freeze and Ng 2011; Jaeken 2011). CDGs are rare in the human population because defects in glycosylation are almost always fatal in development, but when the fetus survives to birth the defects are often quite severe. History has shown that rare diseases have served as Rosetta stones for physiology. These rare genetic diseases of glycosylation have already revealed critical functions of glycans in human development. But to fully capitalize on the data that studies on CDGs will generate will require informatics and systems biology-based approaches to map the links between glycobiology and physiological function, which in turn could generate new therapies and new leads for finding other glycosylation-related genes.

Severe forms of muscular dystrophy result when a genetic mutation causes abnormal glycosylation of alpha-dystroglycan, a protein that links the extracellular matrix to the actin cytoskeleton in muscle cells (Muntoni et al. 2011). Mutations in any of the glycosyltransferases involved in synthesis of the glycan structure on alpha-dystroglycan result in various

forms of muscular dystrophy, clearly supporting an essential role for these glycans in muscle physiology (Godfrey et al. 2011). Another example of how specific glycosylation regulates a critical developmental signaling pathway comes from the study of Notch signaling, which is regulated by the glycans attached to the Notch receptor.

Many lysosomal storage diseases result from the mis-targeting of lysosomal glycohydrolase enzymes to the lysosome (Cox and Cachon-Gonzalez 2012). The targeting of lysosomal enzymes to the lysosome by glycans containing mannose-phosphate moieties has long been a model of glycan-specific functions (Kornfeld 2010). These storage diseases are characterized by severe developmental abnormalities, especially in the brain. Many glycan storage diseases involve proteoglycan or glycosaminoglycan accumulation (Coutinho et al. 2012) or result in the accumulation of gangliosides (Schulze and Sandhoff 2011), perhaps as a result of their abundance and their complex structures. There are seven known forms of glycosaminoglycan storage diseases, all sharing some clinical features, such as organomegaly, and central nervous system malfunctions (Coutinho et al. 2012).

3.1.8 Bioactivity and Pharmacokinetics of Drugs

With notable exceptions such as insulin and growth hormone, the great majority of biotherapeutics are produced as glycoproteins. Attempts to express such proteins in prokaryotic systems, such as *E. coli*, often result in misfolded, unstable, or inactive products. To circumvent these issues, multiple platforms have been developed for the manufacture of glycoproteins, including the recombinant expression of biotherapeutics in yeast, insect, plant, and mammalian cell systems. However, expression in such eukaryotic systems often results in incorporation of nonhuman glycosylation, which can alter efficacy, half-life, antigenicity, and other features of the biotherapeutic (see Box 3-2). Glycan engineering approaches to "humanize" the glycans attached to proteins produced in these systems are a strategy that is actively being explored (Gomord et al. 2005; Chiba and Jigami 2007; Zhang et al. 2011).

Despite the many roles of glycans in biology, there are still relatively few glycan-based drugs (Galan et al. 2011), with a recent review identifying only nine on the market (Shriver et al. 2004). This small number is a consequence of the difficulties in synthesizing glycans and their high polarity, which typically results in poor pharmacokinetics. Instead, most companies synthesize molecules to mimic glycans (Ernst and Magnani 2009; Magnani and Ernst 2009). However, with the right pharmacokinetic properties, glycan-based drugs can prove incredibly useful. For example, one of the oldest and most used set of drugs known is the polysaccharide

BOX 3-2
Influence of Glycosylation on Drugs: Erythropoietin

The anemia drug erythropoietin (EPO) provides an example of a recombinant biotherapeutic in which differences between normal human glycosylation and non-human glycosylation have functional significance and can be used as the basis for testing. The glycosylation of recombinant erythropoietin affects clearance by the liver and influences the pharmacokinetic interactions of EPO with its receptor, which affects uptake into tissues and clearance from the circulation. Nonhuman glycosylation, such as the presence of Galα1-3Gal in recombinant biotherapeutics, has been reported to cause antigenicity, while the capping of glycans with sialic acid is known to dramatically increase circulating half-life. The glycosylation difference between natural human erythropoietin (made by the kidneys) and the recombinant form (made in Chinese hamster ovary cells) is the primary basis for the World Anti-Doping Agency test to detect illicit erythropoietin use by elite athletes, such as cyclists in the Tour de France. Despite a potential illicit market in the billions of dollars, no one has yet succeeded in mimicking the natural human glycosylation in cultured cells.

The importance of glycosylation on biotherapeutic agents goes beyond these biological factors. First, because of differential activity, claims of unique intellectual property can sometimes be made based on differential glycosylation even considering the same underlining polypeptide. The glycosylation difference between erythropoietin made in different types of cells was partly the basis for the successful defense of intellectual property rights challenges against recombinant erythropoietin. This has also been the case with Fc effector functions of antibodies. Second, because of nontemplate addition of glycosylation, a biotherapeutic is actually a mixture of glycoforms rather than a single active ingredient. Biotherapeutic characterization and lot-to-lot comparison are made more difficult because of this microheterogeneity. As a consequence, regulatory agencies such as the FDA must develop definable and "acceptable" ranges for variations of this microheterogeneity, to define process control and comparability across different manufacturing systems. Finally, changes in culture conditions can result in significant changes to glycosylation. Thus, glycosylation is an independent way of monitoring process control. Our ability to understand, analyze, control, and modify glycosylation has and is likely to continue to have a huge impact on the quality and application of biotherapeutic agents. The market for these drugs is in the billions of dollars, and this is an aspect of glycoscience's impacts on not only human health but also economics and world regulatory standards.

heparin and fragments of heparin, which are used in the clinic as an anticoagulant and as an anti-inflammatory agent.

It has been estimated that as few as 20 thousand glycan structures might represent most of the binding specificities of known human glycan-binding proteins (Cummings 2009; Smith et al. 2010). Indeed, it soon will be in the realm of possibility to construct a glycan array with this number

of different glycan structures. These arrays can impact the field of drug discovery by allowing the specificities of lectins, growth factors, cytokines, antibodies, and toxins to be analyzed rapidly (Liang et al. 2008; Wu et al. 2009). This would be the first step in the design and development of glyco-mimetic drugs.

Antibodies, which are glycoproteins, have been used as therapeutics for more than 100 years and the development of monoclonal antibodies in the early 1970s revolutionized their therapeutic potential (Yamada 2011). In recent years, the pharmaceutical industry has put significant effort into the development of antibody therapeutics. However, only recently have they realized the importance of understanding the types of glycans attached to antibodies, which play roles in their efficacy and safety. For example, antibodies made in cultured cells or animals often have glycan structures attached to them that differ from those made by humans, and these changes can elicit deleterious immune responses and affect therapeutic effectiveness. For example, antibodies synthesized in nonhuman cells often are modified by nonhuman sialic acids known as Neu5Gc (Ghaderi et al. 2010) or nonhuman alpha-galactose residues, both of which can elicit an immune response (Lux and Nimmerjahn 2011). Recent advances in glycoengineering of therapeutic monoclonal antibodies should greatly advance the development of useful therapeutic antibodies, and this area represents a substantial investment in the field of drug development. A large number of other therapeutic agents also are glycoproteins, and the glycosylation of these biological drugs plays various roles, ranging from initial protein folding, to stability in circulation, to the criteria used by the FDA to monitor process control.

3.1.9 Key Messages on Glycoscience and Health

As the preceding discussion has demonstrated, glycans play critical roles across all aspects of human health. The importance of glycans to so many areas of biology and the diversity of roles that glycans play are not surprising, because they are one of the four fundamental classes of biological molecules and are abundant on cell surfaces, as well as being major modifiers of proteins and lipids. Significant advances in understanding the functions of glycans in human health and disease, and unlocking ways in which they may contribute to the development of new medical treatments, have already been made. In many cases, glycans provide an additional layer of biological information that builds on what can be learned through genomics and proteomics. Investments in understanding the human genome and proteome have already yielded insights. Because it is so entwined in biology, however, incorporating an improved understand-

ing of glycoscience will be required to fully realize the goals of genomic and proteomic research to improve human health.

As a result, the committee finds that:

- Glycans are directly involved in the pathophysiology of every major disease.
- Additional knowledge from glycoscience will be needed to realize the goals of personalized medicine and to take advantage of the substantial investments in human genome and proteome research and its impact on human health.
- Glycans are increasingly important in pharmaceutical development.

3.2 GLYCOSCIENCE AND ENERGY

The need for increased energy sources to power modern society is a major area of research and discovery. The U.S. Department of Energy has estimated that the United States derives 37.13 quads of energy from petroleum sources (Greene 2011) and that in 2010 U.S. oil consumption was more than 19 million barrels per day (CIA World FactBook; https://www.cia.gov/library/publications/the-world-factbook/). In 2009, fossil fuels (petroleum, coal, and natural gas) were used to meet approximately 83 percent of this country's energy needs, whereas only 8 percent came from a range of renewable energy sources (Greene 2011).

Although the global economic recession and continued development of new oil fields and new oil extraction techniques influence the cost and accessibility of petroleum resources, global energy consumption is expected to continue to increase, and there is ongoing interest in exploring a range of energy solutions. Lynd (2010) has written that twice in human history—the Neolithic Revolution and the Industrial Revolution—major changes in resource use by humans have transformed day-to-day life and societal organization. Today, there are indications that a third revolution, a Sustainability Revolution, may be required. Biomass, which currently provides only an estimated 3.88 quads of U.S. energy (Greene 2011), is one option as an energy source, particularly for the production of liquid fuels that will always supply as much as two-thirds of the demands from the transportation sector. A major underlying theme for the development of biomass resources is the role of sugars in the formation of the complex glycans that make up the bulk of plant cell walls. In principle, these complex glycans, if processed efficiently, can serve as a renewable source of high-value biofuels and bioproducts in much the same way that oil serves as the source of high-value fuels and petroleum products. However, achieving this sustainability revolution will be a challenge because

it must be global in scale, occur over a shorter period of time than the previous two revolutions, and will require overcoming the recalcitrance to degradation of plant cell walls that make up the bulk of what is collectively known as biomass (Himmel 2008; Lynd 2010; Lynd and Cruz 2010).

3.2.1 Biomass—Plant Cell Walls

Plant cells walls represent Earth's dominant biological carbon sequestration system. However, although plant cell walls are a terminal carbon sink, they are not merely a storage depot. Rather, they are a highly complex set of polysaccharides with structural proteins and lignin. It is important to recognize that there is no such thing as a "cell wall." The cell wall structures of plant species, grasses, and trees, for example, have different levels of structural heterogeneity (Avci et al. 2011). In addition, there can be 35 to 40 tissues in a single plant, each with a cell wall containing structural differences. Even in a given tissue or cell, cell wall layers and domains may have very different structures or polysaccharide distributions (Albersheim et al. 2010).

During photosynthesis, plants use sunlight to "fix" carbon from carbon dioxide into sugars, a form of chemical energy. In this process the photons in sunlight are used to produce molecules that contain high-energy chemical bonds, such as ATP and NADPH. These molecules in turn are involved in a series of reactions that create carbon-containing sugars. It is these sugars in biomass that are considered a source of energy for the production of biofuels. However, little is known about how the cell converts those sugars into complex polysaccharides and assembles the completed polysaccharides into networks and then into the overall architecture of the cell wall. Cellulose, hemicellulose, and pectins—all complex polysaccharides—plus lignin, and often structural proteins, all contribute to the overall architecture of the cell wall (see Figure 3-3). Once the networks are formed, there are further processing steps, such as hydrolase processing, about which little is known but that are major contributors to polysaccharide heterogeneity. How this whole system is regulated is still largely a mystery (Albersheim et al. 2010) and remains a major challenge in glycoscience.

How this whole system is regulated is still largely a mystery (Albersheim et al. 2010) and remains a major challenge in glycoscience. Although the process of how plants make sugars from carbon dioxide and water is fairly well understood, little is known about how the cell partitions that carbon into different sugars and then into the 14 different nucleotide diphosphate sugars (NDP sugars) needed for the biosynthesis of cell wall polysaccharides. In most cases the enzymes involved in these processes are known, but it is not known how they are regulated. Also,

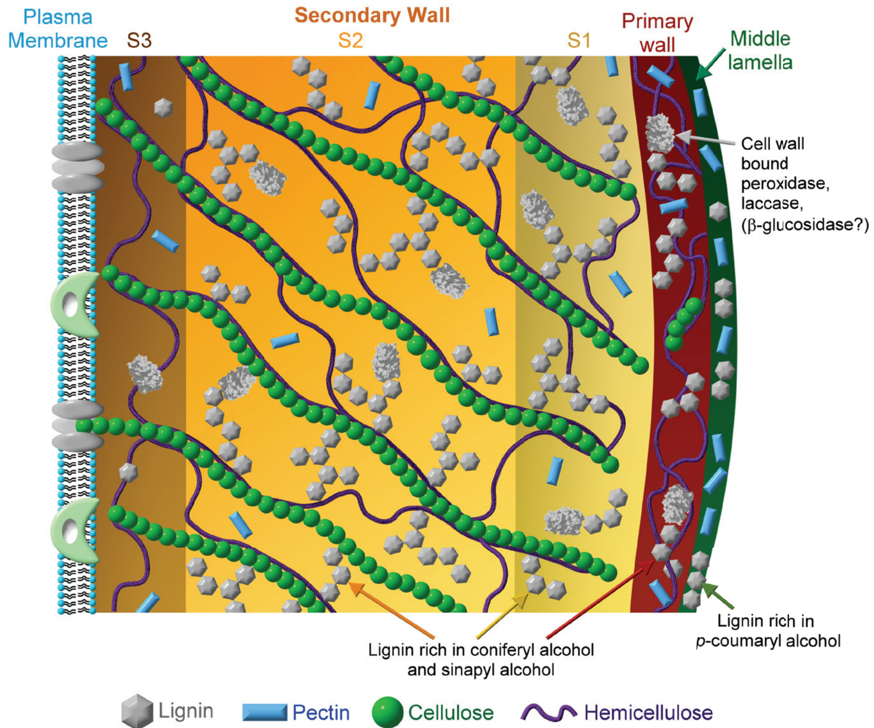


FIGURE 3-3 Schematic of a plant cell wall.
SOURCE: Achyuthan et al. 2010, used with permission.

yet to be clarified is the role that Golgi-localized NDP-sugar transporters play in the subsequent synthesis of polysaccharides and the identity of all of the glycan syntheses and glycosyltransferases involved in the production of the component polymers of plant cell walls (Bar-Peled and O'Neill 2011). Even when the identity of these enzymes is known, it is still unclear how they choose their substrates, whether they need a primer to start the synthesis of a polymer chain, whether they work in protein complexes, and what determines the ultimate length of a synthesized polysaccharide. To date, based on the glycosidic linkages found in any given plant cell wall, a good estimate is that about 120 different glycosyltranferases are needed to synthesize the plant wall, of which only about seven are known. Another important unknown regarding cell wall structure is how a plant actually assembles the completed polysaccharides into networks and then into the overall architecture of the insoluble cell wall. Once the

networks are formed, there are more processing steps that are major contributors to polysaccharide heterogeneity (Albersheim et al. 2010).

3.2.2 Recalcitrance to Degradation of Biomass Feedstock

The major obstacle to realizing a future in which liquid fuels and other high-value products and materials are produced from biomass as a supplement to, or in place of, petroleum is the recalcitrance of the cell wall to degradation and release of sugars for fermentation (Himmel 2008). Cellulosic biomass's recalcitrance to degradation is related to the need for a plant's cell walls to support its morphology and biological functions, including upright growth habit, resistance to internal turgor pressure, and water and nutrient transport. It also arises from the plant's need to evolve resistance to attack by the elements and from microbes and their enzymes.

There are several approaches to overcoming this recalcitrance, including:

- creation of a modified cell wall architecture that would be more susceptible to deconstruction by microbes or enzymes,
- reduction of the inhibitors in plant cell walls that reduce the efficiency of microbial and enzymatic deconstruction,
- reduction or modification of the cell wall lignin content, and
- increase in the amorphous cellulose and hemicellulose content of the cell wall and increase in susceptibility to breakdown.

This is an area where glycoscience understanding will play an essential role in bioenergy science, given the large number of unknowns that exist concerning how a plant makes a cell wall and in understanding the enzymes and microbial interactions that can aid in overcoming this cell wall recalcitrance to degradation (Himmel et al. 2007).

Because of the wide variation in cell wall composition and structure across the plant kingdom, the type of feedstock used as a source of sugars for biofuel production has a pervasive impact on biofuel performance metrics. Perennial biomass cellulosics (e.g., switchgrass) score highly except in terms of cost-efficient conversion of biomass to sugars as a result of cell wall recalcitrance. A number of companies that are currently developing commercial conversion facilities to produce liquid fuels from lignocellulose believe that it will be possible to sell such fuels at the same price as petroleum without any subsidies (DOE 2006; Lynd et al. 2008). As noted above, however, the grand challenge to achieving that kind of price drop is overcoming the recalcitrance of cellulosic biomass to conversion into the sugars that are needed to produce liquid fuel via fermentation. This is the most costly processing step or set of steps in producing liquid

fuel from biomass, but there is potential for research and development–driven cost reductions. Such advances are not only necessary but also would be sufficient to create a cellulosics biofuels industry (DOE 2006).

Understanding cell wall recalcitrance at this basic level would further the achievement of several goals. Efforts are under way to develop improved enzymes and other catalysts for decomposing cell walls, and an understanding of plant recalcitrance helps in both the design of new molecules and investigating how and why different cell walls respond differently to them (Himmel et al. 2007; Rubin 2008; Carroll and Somerville 2009; Van de Vyver et al. 2011). Genomic investigations have identified a large number of potential carbohydrate-active enzymes, and the CAZy database for carbohydrate-active enzymes (http://www.cazy.org/) is an important resource in this area. For example, more than 27,000 putative carbohydrate-active enzymes were identified in a recent study of a cow rumen (Hess et al. 2011). It remains largely unknown whether or how this range of potential genetic and structural enzyme variations translates to a range of useful functional variations, and research will be required to decipher the functions and possible applications of these potential enzymes. As this area demonstrates, glycoscience will indeed continue to build on new discoveries and developments in fields such as genomics.

Basic knowledge about cell walls can also be used to reconstitute or, more likely, modify the cell wall to make it less recalcitrant to processing for energy production. This may happen by altering existing pathways in plants or introducing new ones that lead to the creation of less recalcitrant polysaccharides or by altering lignin-polysaccharide interactions (Albersheim et al. 2010). The creation or modification of cell walls with altered chemical compositions or properties will draw on such techniques as genetic engineering. Although the scientific and economic drivers for these developments are clear, the potential ecological impacts and social perceptions of genetic engineering will need to be assessed as part of these and other biofuels research efforts.

Cell wall recalcitrance manifests itself at surfaces where microbial enzymes and the plant cell wall interact (Himmel et al. 2007). Given that both the plant cell wall and microbe surfaces are composed primarily of glycans, glycoscience will play a central role in understanding these interactions. While understanding how isolated enzymes or other catalysts interact with the plant cell wall as part of deconstruction is an important avenue of investigation, investigations are also under way to determine whether microbes can be more effective than enzymes at deconstructing the plant cell wall for lignocellulose feedstocks (Lu et al. 2006; Olson et al. 2011). Theories being explored include the suggestion that microbes alter the structured layers of water that form along the cell wall surface. This water layer may present a barrier to the approach of free enzymes and

inhibit the escape of soluble products, which in turn would slow hydrolysis rates. Thanks to 1 billion years of evolution, microbes have surfaces that may enable them to disrupt this water layer, effectively solubilize the plant cell wall, and capture solubilization products. Molecular mechanics simulations appear to support this hypothesis (Himmel et al. 2007; Ding et al. 2008; Lynd 2010). Manipulating or modifying such microbes may lead to the generation of more effective microorganisms for consolidated bioprocessing (Olson et al. 2011).

Another challenge in creating biofuels from plant biomass is to increase production of biomass per hectare and to increase the yield of fermentable sugars from each ton of biomass. Increased production can be done by choosing the right crop for a given set of environmental conditions and by increasing the resource efficiency of these plants by decreasing water and fertilizer needs (Tilman et al. 2009; Lynd and Cruz 2010). From a biological perspective, one possibility is delaying the time when plants flower. Keeping plants in a juvenile stage will also increase biomass production. At least one research team has found that introducing a micro-RNA into a plant that caused it to remain in its juvenile stage increased biomass production but also caused the plant to reduce lignin production, leaving the resulting biomass more susceptible to degradation and conversion (Chuck et al. 2011).

3.2.3 Key Messages on Glycoscience and Energy

There is increasing interest in how to make use of the carbon in plant biomass as an alternative source of energy to petroleum. However, plants have evolved to resist destruction by the environment and by microbes, and so efficient deconstruction of cell walls into carbohydrates and sugar intermediates that can be fermented into biofuels is a challenge. Understanding how plant cell walls are synthesized and constructed and, conversely, how they can be effectively broken down remains a major goal for improving biomass-derived energy. Addressing this challenge will require continued research into the cell wall formation and breakdown process and forms an important part of the glycoscience field.

As a result, the committee finds that:

- Plant cell walls, made mostly of glycans, represent the planet's dominant source of biological carbon sequestration, or biomass, and are a potentially sustainable and economical source of non-petroleum-based energy.
- Understanding cell wall structure and biosynthesis and overcoming the recalcitrance of plant cell walls to conversion into feedstocks that can be transformed into liquid fuels and other energy

sources will be important to achieving a sustainable energy revolution. Glycoscience research will be necessary to advance this area.

- Glycoscience can also contribute significantly to bioenergy development by advancing the understanding of how to increase biomass production per hectare and how to increase the yield of fermentable sugar per ton of biomass.

3.3 GLYCOSCIENCE AND MATERIALS

As discussed above, plant cells walls represent the planet's dominant biological carbon sequestration system. They are estimated to account for 120 billion to 170 billion tons per year. In comparison, the annual global production of chemicals, including fertilizers but not pharmaceuticals, was 1.2 billion tons in 2010. In other words, it takes plants three days to produce biomass in the form of cell walls equal to the total annual output of the world's chemical industry (DOE 2011).

Many of today's most widely used materials are petroleum based. With the world's population growing, consumer-based economies expanding, and petroleum resources ultimately being stretched in many directions, there is a global need for the development of renewable and sustainable resources that can serve as the source of a new generation of materials. This is a surmountable task, one that could be at least partly accomplished using polysaccharides produced by a wide variety of living species, including plants, algae, fungi, and even insects and arthropods. This diverse collection of organisms produces sufficient quantities of harvestable polysaccharides to potentially meet the world's demand both for energy and synthetic materials (Perlack et al. 2005). Investigating glycan-based materials further opens the possibility of designing new materials with tailored properties that can expand on the range of materials currently available and may find new applications.

Materials based on polysaccharides, particularly cellulose-based polysaccharides, have a long history as a source of functional materials used by human society. Wood, cotton, linen, hemp, and other cellulose-based polysaccharides have been used as engineering materials for thousands of years, and their use continues today, as evidenced by the enormity of the worldwide industries in forest products, paper, and textiles, among others. As a chemical raw material, cellulose has been used for about 150 years (Klemm et al. 2005). Regenerated cellulose has been used in the processing of synthetic cellulose films (cellophane) and fibers (rayon), while cellulose derivatives, made, for example, by replacing hydroxyl groups with other functional groups, have been used to produce a wide variety of cellulosic polymers, including cellulose esters and cellulose ethers, which

have many industrial and pharmaceutical applications. Commercial products made from cellulose derivatives include coatings, inks, binders, thickening and gelling agents, and controlled-release drug tablets (Klemm et al. 2005). Recently, the availability of cellulose-based nanoparticles has begun changing the paradigm of what is achievable with natural materials, including the production of biodegradable transparent films that are stronger than steel (Moon et al. 2011). For two examples, see Box 3-3.

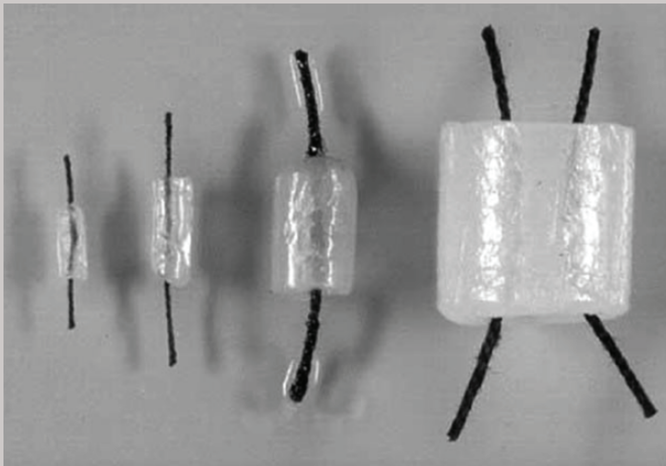
Further advances in glycoscience could drive the discovery of a wide range of new sustainably produced polysaccharide-based materials in three general categories: fine chemicals and feedstocks, polymeric materials, and nanomaterials. Glycoscience has a real opportunity to positively impact the progress of the baseline technologies needed to develop cost-effective materials from polysaccharides that can compete, in terms of cost and performance, with petroleum-based chemicals and polymers. This vision can be achieved by providing mechanisms to develop an increased understanding of polysaccharide biosynthesis in plants and trees, new characterization tools and methods for understanding polysaccharide structure, new methods for polysaccharide isolation, synthetic process and chemical modification, and improved predictive modeling capabilities.

3.3.1 Fine Chemicals and Feedstocks

A variety of polysaccharides are being investigated in the production of functional chemical precursors that are then subsequently used to make industrially relevant chemicals and engineering polymers (Bozell and Petersen 2010). Examples of functional chemical precursors include alcohols, such as ethanol, propanol, butanol, xylitol, and sorbitol; furans, such as furfural and hydroxymethylfurfural; biohydrocarbons, such as isoprene and long-chained hydrocarbons; and organic acids, such as lactic acid, succinic acid, and levulinic acid. Carbohydrates are one of a handful of natural products that can be used for production of many chiral compounds with defined stereochemistries. Research has focused on optimization of the bioconversion of polysaccharides in terms of yield, rate, separation, titer, and product specificity. Much of this work has focused on identifying and engineering improved fermentation organisms, fermentation processes, and catalysts for converting sugars into chemical precursors (Bozell and Petersen 2010). Significant advances will require further development of methods to improve our ability to engineer specific enzymes and organisms, such as yeast and fungi to produce large yields of single materials. Advances in materials development will also require a better understanding of how to achieve specific chemical

BOX 3-3
Examples of Carbohydrate-Based Materials: Flexible Displays and Artificial Blood Vessels

Polysaccharides can be used in the development of a wide range of new materials with very diverse properties. For example, cellulose nanocomposites from wood can be used as the basis for flexible, optically transparent materials that can be used as substrates in the creation of luminescent organic light-emitting diodes. Such materials have the potential for new applications in flexible electronics and displays. Cellulose, in this case derived from bacteria, can also be used to make thin, flexible tubes for use as implanted blood vessels.



Glycan-based materials have multiple uses. SOURCES: *Top*, Reprinted from Okahisa et al. 2009, with permission from Elsevier; *Botttom*, Klemm et al. 2011, used with permission.

reactions with organisms and catalysts in a way that enables selective reductions, conversions, and chemical bond formation or bond breaking.

3.3.2 Polymeric Materials

As discussed earlier, polysaccharides represent a broad range of natural polymers consisting of repeating sugar monomer units joined together by glycosidic bonds. They can be linear or highly branched and can have a wide variety of side groups. As a result of the variety of polysaccharides produced in living species through their biosynthesis processes, the diversity of polysaccharide materials available in nature creates an almost limitless range of possibilities for creating useful new materials. For example, trees produce cellulose and hemicelluloses, whereas nonwoody plants produce cellulose, pectins, and starches, and various bacteria can synthesize polysaccharides that include glycogen, alginate, xanthan, dextran, curdlan, gellan, colanic acid, K30 antigen, hyaluronic acid, and cellulose (Rehm 2010). Each of these polysaccharides has a unique chemical structure comprising different combinations of sugars linked together in different configurations, all of which influence the properties of the given polysaccharide in terms of its structural configuration, thermal stability, reactivity, rheology, and mechanical properties.

With these natural materials as starting points, extensive research has gone into the development of new reaction pathways that modify the existing polysaccharide backbone structure or the side groups that branch off the backbone. The goal of these efforts is to create new polymeric materials with novel properties and functions (Klemm et al. 2005; Roy et al. 2009). For example, cellulose can be dissolved and the glucan chains reassembled to produce regenerated cellulose, or cellulose II, which has been used to produce cellulose films, known as cellophane, and fibers including rayon. Research has also yielded reaction pathways that create cellulose derivatives such as cellulose acetate, cellulose acetate propionate, cellulose acetate butyrates, carboxymethyl cellulose, cellulose butyrate succinate, and cellulose acetate propionate (Klemm et al. 2005). Each of these cellulose-based polymers was created by replacing accessible hydroxyl groups with other chemical groups to produce a material with novel performance characteristics. Given the wide range of polysaccharides produced by nature, these examples represent only a fraction of the potentially novel and useful materials that could be produced. But the complexity and variability of polysaccharides represent a challenge that requires fundamental research to develop:

- faster and more accurate methods for structural characterization of polysaccharides,

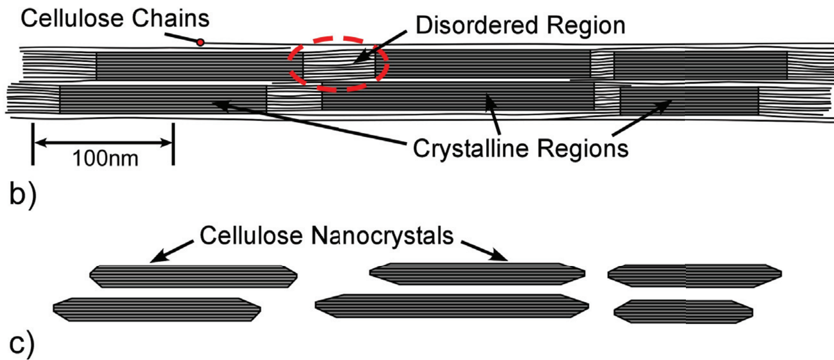
- improved technology for separating and isolating polysaccharides from their natural sources,
- novel chemical reactions to target chemical modifications to specific locations on the polysaccharides in order to create regioselective functionalization, and
- synthesis pathways for long-chain polysaccharides.

3.3.3 Nanomaterials

Polysaccharides with linear or minimally branched backbone structures can self-assemble into ordered bundles in which the polymer chains stack in parallel with each other along the chain axis. Such parallel-stacked chains from a crystalline structure that can be characterized by x-ray, neutron, and electron scattering techniques, among others. Two primary examples of polysaccharides that show this crystalline structural behavior are cellulose and chitin, a polysaccharide isolated from fungi and from the exoskeleton of crustaceans and insects. The structures of cellulose and chitin are similar, the difference being that chitin has one hydroxyl group on each sugar replaced with an acetyl amine group. However, the two polysaccharides have considerably different physical and mechanical properties. Nonetheless, during the biosynthesis process, both of these linear polysaccharides form fibril structures containing both crystalline and amorphous arrangements of polymer chains. These fibrils serve as the base reinforcement unit that provides the high mechanical strength, strength-to-weight ratio, and toughness of plants, trees, crustaceans, and insects.

Using specialized chemical-mechanical extraction methods, these fibril structures and their crystalline regions can be isolated and used to develop the next generation of plastics. To date, most work has been completed with cellulose, largely because of its availability and the extensive scientific and technological expertise developed by the pulp and paper industry. With this in mind, the next sections focus on cellulose to illustrate some of the opportunities for creating such nanomaterials, along with the associated challenges. The issues discussed are relevant to other ordered polysaccharides as well.

For cellulose the particles that are isolated after chemical and mechanical extraction have dimensions on the nanoscale and are generically called cellulose nanomaterials (CNs) (see Figure 3.4). CNs can have either a rod and whisker–like structure or a fibril particle morphology, with the dimensions varying depending on cellulose source, extraction methods, and extraction conditions (Habibi et al. 2010; Moon et al. 2011). Typically, CNs range from 3 to 30 nanometers (nm) in diameter and from 50 nm to several microns (μm) in length, with length-to-width (aspect) ratios of 20



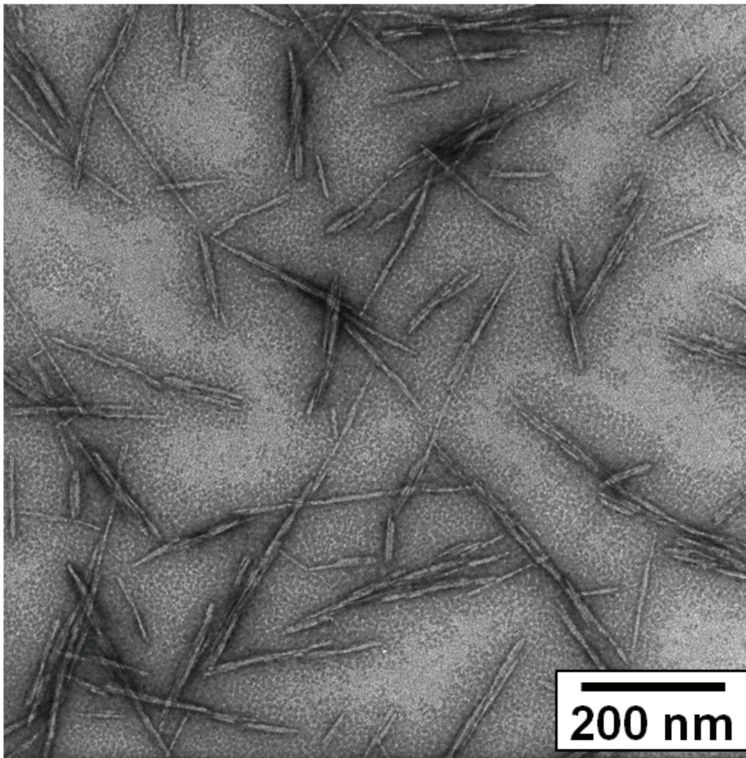


FIGURE 3-4 Cellulose nanocrystals. *Top*, Stacking of cellulose chains showing areas of "order" and "disorder." During one type of cellulose nanomaterial extraction process that uses acid hydrolysis, the amorphous regions are preferentially dissolved and only the crystalline regions are left. *Bottom*, Transmission electron micrograph of cellulose nanocrystals produced by acid hydrolysis.
SOURCE: Moon et al. 2011, used with permission.

to 100. CNs have high stiffness, low density, low thermal expansion, and thermal stability up to about 200 to 300°C, and their surfaces can be readily modified using a variety of chemical methods.

The sources from which CNs are extracted are themselves sustainable, biodegradable, and carbon neutral, and they generally have low environmental, health, and safety risks. CNs have the potential to be processed at industrial-scale quantities at low cost, although reaction conditions, feedstock crystallinity, and other factors influence achievable yield (Habibi et al. 2010; Qua et al. 2011). Preliminary testing shows minimal environmental, health, and safety risks for CNs (Vartiainen et al. 2011), although investigations into the environmental and health effects of all types of nanoparticles continue to be an area of research and discussion.

With these advantages in mind, CNs are being considered for use in the development of a variety of new plastics and composite structures, including films, fibers, aerogels, and hydrogels. These materials have potential applications in barrier films, separation membranes, antimicrobial films, transparent films, flexible displays, cardiovascular implants, wound and burn dressings, tissue regeneration scaffolding, drug delivery vehicles, fibers and textiles, templates for electronic components, batteries, supercapacitors, electroactive polymers, and body armor.

Glycoscience could have a transformative impact on the CN bioplastics industry by advancing the technologies necessary to control the crystalline structure, properties, surface chemistry, particle morphology, and particle size distribution of CNs. Such technologies could then be used to produce "tailored" CNs to meet specific performance metrics at low cost. This vision can be achieved with advances in the following areas, each of which are described in the sections that follow:

- understanding cellulose nanomaterial extraction processes,
- CN characterization,
- atomistic modeling of cellulose, and
- cellulose synthesis.

3.3.3.1 Understanding cellulose nanomaterial extraction processes

Unlike the extractions generally needed to produce fine chemicals, feedstocks, and new polymers, the techniques needed to extract completely ordered polysaccharide particles from natural cellulose source materials demand particular delicacy, and developing suitable processing technologies requires new techniques and methodologies. Extracting CNs from cellulose requires several steps, each of which influences particle morphology or shape, particle size and size distribution, and interfacial properties (Moon et al. 2011). Current processing technologies afford min-

imal control of each of these physical properties, which in turn affects the performance traits of the resulting nanoparticles. Because they are derived from cellulose, a better understanding of the mechanisms involved in plant cell wall construction and destruction can help provide insight into new mechanisms that could be exploited to improve CN extraction methods to decrease the internal damage in CNs, narrow the particle size range for a given CN processing methodology, improve extraction efficiencies, increase CN yields, and scale production to industrial quantities.

3.3.3.2 Cellulose nanomaterial characterization

In addition to the structural and chemical characterization needed to produce fine chemicals, feedstocks, and new polymers, novel structural, chemical, and mechanical characterization techniques are needed for ordered polysaccharides. In particular, there is a need for techniques that can characterize the configuration of the parallel-stacked polysaccharide chains found in CNs. The properties of any given CN depend on the arrangement of the cellulose chains, defects in the ordering of chains, local changes in the chemistry internal to the CN, and changes in the chemistry on the external CN surface. Characterizing these aspects is important for understanding CN properties and for developing the means to rationally design new CNs with specific properties. In addition, improved characterization of structure, nanomechanical properties, and surface chemistry will provide the opportunity to better understand processing-structure-property relationships as they relate to the CN particles themselves and to CN-CN and CN-water interactions, all of which are important for the design of composite materials with improved performance.

While the cellulose polymorph structures of CNs are generally known, characterization of individual CNs is currently lacking in such areas as the percentage of crystallinity, the location of amorphous regions on the CN surface or throughout the CN core, the fraction of a given cellulose polymorph structure and its location in the particle, the identification of defects such as missing cellulose chains, and the hydrogen bonding networks both within and between cellulose chains in the CN. Characterizing the elastic and tensile strength and other nanomechanical properties of CNs remains a huge challenge, because their small size pushes the limits of sensitivity of current methodologies such as atomic force microscopy (Wagner et al. 2011). While the reported properties for CNs are on par with atomistic model predictions, these models are too variable for conducting fundamental research on structure-property relationships, let alone for assessing the influence of different extraction processes or cellulose source on the quality of the resulting nanomaterial. Also lacking is the ability to pinpoint the particular location of a given chemically func-

tionalized side group along the CN surface. This information is important to assess the accessibility of the given side group and how these groups can interact with neighboring CNs or with the surrounding liquid in CN-liquid suspensions.

3.3.3.3 Atomistic models of cellulose

New methods for atomistic modeling can contribute to better understanding of CN properties by providing insights into how to characterize and describe the configuration of the parallel stacking of the polysaccharide chains and how that configuration affects assembly properties. When linked with experiments for direct comparison and validation, atomistic modeling is a useful tool to help develop a better understanding of the structure, mechanical properties, and surface chemistry of nanomaterials such as CNs. Atomistic modeling can also provide information on the interactions of CNs with neighboring nanoparticles, which affects the composite properties of a CN-based material, and the interactions within a particle suspension, which affects rheology. In addition, atomistic models that could accurately represent surface interactions and preferential bonding site location would provide insights into composite properties and aid efforts to tailor CN properties.

3.3.3.4 Cellulose synthesis

Today, it is particularly challenging to synthesize ordered polysaccharides with the desired surface chemistry that self-assembles with controlled crystalline properties. Through detailed understanding of the polysaccharide biosynthesis process, it will ultimately be possible to develop processing routes that facilitate tailoring both the side-group chemistries of the polysaccharide backbone structure and the assembly of individual chains into an ordered chain structure.

3.3.4 Key Messages on Glycoscience and Materials

Plastics and many other materials are currently derived from petroleum resources. Carbohydrates derived from plants, microorganisms, insects, and other biological sources can provide a wealth of new options for materials with diverse chemical and physical properties. Glycoscience knowledge is helping unlock the development of these new materials and the control of their design. The example of cellulose nanoparticles was described in greater detail previously in this section to help illustrate some of the research questions involved in characterizing and designing new materials derived from carbohydrates. However, the need to under-

stand chemical structures and interactions, to characterize and modify materials, and to control production processes is important to developing functional and economically feasible new materials and products more broadly. The application of glycoscience to materials science and engineering represents an expanding area for the field and one for which continued research and development will be needed.

As a result, the committee finds that:

- By fostering a greater understanding of the properties of glycans and of plant cell wall construction and deconstruction, glycoscience can play an important role in the development of nonpetroleum-based sustainable new materials.
- Glycan-based materials have wide-ranging uses in such areas as fine chemicals and feedstocks, polymeric materials, and nanomaterials.
- There are many pathways to create a variety of functionalities on a glycan, creating a wide range of options for tailoring material properties.

3.4 SUMMARY

This chapter explores three significant ways in which carbohydrates contribute to society—in human health, in energy, and in materials. Glycans are closely linked to both normal physiological function and to the genesis and development of disease. They play promising roles in the discovery of new diagnostic biomarkers for such diseases as cancer and for new therapeutic targets. But glycans and glycoscience also have important roles to play in the improved conversion of biofuels and the design and creation of new carbohydrate-based materials. As a result, glycoscience knowledge will contribute to the development of new energy and materials science solutions that can replace some of the roles currently played by petroleum-based products. This chapter attempts to provide a sense of some of the unanswered questions in glycoscience as well as the exciting potential that may be realized through future research. The next chapter continues this discussion by exploring a set of questions that we may not yet have the necessary tools to fully address but the pursuit of whose answers remains an exciting challenge.

4

Examples of Outstanding Questions in Glycoscience

As this report highlights, glycans are critical players in virtually every biochemical process that makes life possible on Earth, yet compared to the study of nucleic acids and proteins, glycoscience is in its infancy. The reasons that this leg of the four-legged stool of life (nucleic acids, proteins, lipids, and carbohydrates) is underdeveloped are many, but they largely come down to four factors: the astonishing complexity of glycan structures and biochemistry dwarfs the complexity of both nucleic acids and proteins; the relative lack of tools to probe and understand glycan structure and biochemistry in light of that complexity; the lack of adequate data resources and informatics tools for their analysis; and the lack of education in glycoscience.

Today, glycoscience sits at a crossroads. It can continue to be a specialty practiced by a smattering of investigators studying specific problems. If the field stays on that path, it will surely continue to generate important insights but not at a rate that will allow it to live up to its enormous potential and, as this report has already discussed, to address major problems in human health, energy, and materials development. Or glycoscience can take advantage of advances in genomics and proteomics, as well as such fields as chemical synthesis, microbiology, microfluidics, biochemistry, and nanotechnology, to not only enable a more aggressive and comprehensive effort involving a large cadre of researchers but also to provide a blueprint for realizing the transformative leaps in technology and methods that make such an effort feasible and likely to generate enormous benefits to society.

The previous chapter discusses the importance of glycoscience as it relates to the areas of health, energy, and materials science. This chapter describes a series of scientific questions in which glycoscience plays a central role. As in previous chapters, this list is not meant to be exhaustive but is meant to stimulate discussion and highlight the scale of progress that could be made in addressing a broad set of questions embraced not only by glycoscientists but also by the larger science and technology community. Although many of the sections pose questions immediately related to human health, addressing the core issues they raise can have relevance to other areas. Included are examples from the fields of energy and materials, and the committee invites the broader community to develop and embrace other possible challenges important to those subjects.

4.1 WHAT ARE THE MECHANISMS AND ROLES OF GLYCAN DIVERSIFICATION IN EVOLUTION?

It is well accepted that "nothing in biology makes sense, except in the light of evolution." But when it comes to glycans, very little is known about glycan diversification during evolution. Over three billion years of evolution has failed to generate any kind of living cell that is not covered with a dense and complex array of glycans. Why and how has evolution led to this diverse array of glycans, and what are some of their roles, for example, in determinants of host-pathogen recognition? Glycans on host cells may be targets for pathogen recognition, and glycans can undergo subtle changes that may allow evasion from a pathogen, even while preserving sufficient intrinsic function. Glycans appear to remain a preferred class of molecules for the cell surface given their tolerance of such subtle changes, while proteins appear to be somewhat less tolerant of sequence changes, which more frequently result in loss of structure or function. In some cases these changes can result in glycan polymorphisms in the population or even wholesale elimination of specific types of glycans from certain taxa. Once a type of glycan diversification has occurred in a particular species, it could then be recruited for other intrinsic functions, some of which may remain noncritical and some of which may become essential. Thus, glycan diversity may involve continuous evolutionary adaptation and diversification for the generation of intrinsic functions as well as co-evolution through interactions with pathogen and symbionts. Remarkably little is known about the evolutionary diversity of glycans in nature.

4.2 HOW CAN SINGLE GLYCOFORMS AND POLYSACCHARIDES BE SYNTHESIZED AND HOW CAN SPECIFIC GLYCANS AT SPECIFIC SITES ON GLYCOPROTEINS BE MODIFIED?

One of the confounding features of glycoproteins is the incredible diversity of specific molecular species that can exist even in a single cell. For each site on a protein that can be glycosylated, and often there are multiple sites, the number of glycans that can be attached can be large. Indeed, research suggests that cells create different glycoforms for a given protein as an important means of modulating the properties of that protein and its interactions with other biomolecules (Varki 1993). Glycan diversity as expressed in glycoforms may, in fact, help explain human complexity (Varki 2006; Bishop and Gagneux 2007).

The factors that govern glycan diversity pose significant challenges to the isolation, structural characterization, and synthesis of single glycoforms. The structural diversity of glycans arises from the various linkage combinations between the monosaccharides that make up glycans, and those linkages are determined by an array of more than 250 enzymes in the human secretory pathway that support glycan synthesis and processing, including a suite of glycosyltransferases that add sugars using activated sugar donors and glycosidases that cleave them (Ohtsubo and Marth 2006; Varki 2006). In this network, competing or overlapping substrate and donor specificities, substrate availability, and varying levels of enzyme expression, activation, and localization along the secretory pathway all contribute to functionally significant glycan heterogeneity (Lowe and Marth 2003). While such heterogeneous mixtures serve a purpose in biological systems, they are inadequate for the important task of establishing how the molecular architectures of glycans convey specific biological properties.

Access to homogeneous glycoproteins is necessary to first determine the molecular details of a glycan's function and then to produce those glycoforms with the desired properties. Today, it is possible to isolate relatively simple homogeneous glycoforms using enzymatic trimming of glycoproteins (Schmaltz et al. 2011), but synthesizing more complex homogeneous glycoforms requires the development of more sophisticated methods for the chemoenzymatic manipulation of glycoproteins as well as new techniques for synthesizing them de novo. Establishment of single glycoform synthesis has been critical to the identification of glycoforms with distinct properties, such as enhanced glycoprotein stability (Hanson et al. 2009; Price et al. 2010; Culyba et al. 2011), altered binding or immunogenic properties (Dwek 1996), and increased therapeutic efficacy (Arnold et al. 2007; Jefferis 2009)—to name a few. The identification of desired glycoforms will also spur large-scale production needs. For example, glycoproteins with increased α-2-3 sialylation of *N*-glycans

have longer serum half-lives (Stockert 1995; Walsh and Jefferis 2006), and monoclonal antibodies lacking core α-1,6-fucosylation on an *N*-glycan in a conserved Fc region have higher therapeutic antibody-dependent cellular cytotoxicity (Satoh et al. 2006). Unfortunately, recombinant expression in cells is the only practical method for large-scale production of glycoproteins, and this process results in heterogeneity that does not maximize glycoform efficacy (Sethuraman and Stadheim 2006). Viable routes to targeted human glycoforms by recombinant means will require advances in pathway engineering, a field that is quickly emerging. There have been some recent advances in the use of recombinant glycoprotein synthesis in industry, but widespread use has yet to be achieved.

A challenge, then, is to develop new synthetic, chemoenzymatic, whole-cell routes to create single glycoforms in order to facilitate the ongoing process of better understanding glycan function and producing active glycoforms (Koeller and Wong 2001). In particular, the development of new chemical technologies for de novo glycoprotein synthesis will have the advantage of affording precise chemical control of many aspects of glycoprotein structures, including site-specific incorporation of the glycan into the protein component. At the same time, novel enzymatic and recombinant methods would augment chemical approaches and perhaps be useful for creating homogeneous glycoforms by remodeling heterogeneous glycoprotein populations from natural sources.

4.3 HOW DOES GLYCAN MICROHETEROGENEITY OCCUR, WHAT DOES IT DO, AND WHAT IS ITS IMPACT?

Not long after the discovery that most cell surface and secreted proteins are modified by covalently linked glycan moieties, it became obvious that these glycan structures on glycoproteins are extremely diverse. Structural heterogeneity is the general rule even when characterizing the glycans at a single, well-defined glycosylation site on an identified mature protein produced by a uniform population of cells. Similarly, the same protein expressed in two different cell types is often modified by different ensembles of glycans. A central hypothesis of glycobiology is that cellular control over glycoprotein microheterogeneity allows for precise regulation and diversification of function, not unlike the manner in which splice variants of a transcript impart greater flexibility to gene function. However, unlike the characterization of splice diversity, the experimental techniques that would allow assignment of specific functions to identified glycans on individual proteins are still in their infancy. Robust technologies are needed to be able to determine site-specific glycosylation of proteins in complex mixtures. Furthermore, integrated analytic and biological technologies will be required that can make correlations between

glycoprotein glycoforms, recognition by glycan-binding proteins, regulatory factors, and biological functions. Comparing the resulting glycointeractomes in specific physiological or disease contexts will provide an understanding of the molecular mechanisms that control the formation of glycoprotein glycan microheterogeneity, offer insight into the ways in which glycans mediate specific functions, and present unique opportunities for rational development of therapeutic strategies for a wide range of diseases.

4.4 WHAT ARE THE THREE-DIMENSIONAL STRUCTURES OF INTACT GLYCOPROTEINS?

The majority of proteins in mammals are glycosylated, and carbohydrate components play essential roles in development, in the immune response, in intercellular communications that may be defective (e.g., in cancer cells), in inflammatory responses, and in many other biological functions. Yet currently, proteins are frequently expressed without carbohydrates, and the effects of glycans on the structure and function of glycoproteins are avoided for expediency. What are the three-dimensional structures of *intact* glycoproteins? How does the carbohydrate affect the three-dimensional structure? Because the role of carbohydrates is often essential to their function, what techniques need to be developed to determine their three-dimensional structures, inclusive of the carbohydrate components themselves? For all glycans their overall structures ultimately determine their biological functions. When they bind to a specific protein, the three-dimensional conformation of structures known as determinants that have three to six monosaccharide units is involved, and sometimes the multivalent presentation of determinants is important in the strength and specificity of the glycan-protein interaction. It is also clear that the population of glycoforms produced by a cell is key to understanding their overall biological roles in modulating the function of the peptide. Thus, an ongoing challenge is to develop the tools necessary to enable robust and accurate three-dimensional structures of different defined glycoforms of glycoproteins to be determined at the atomic level.

4.5 HOW DOES NUCLEAR AND CYTOPLASMIC PROTEIN GLYCOSYLATION REGULATE CELLULAR PHYSIOLOGY?

In the early 1980s, researchers made the surprise discovery that nuclear and cytoplasmic proteins are dynamically modified at their serine and threonine residues by an *N*-acetylated amino sugar (O-GlcNAc) derived from glucose via the hexosamine biosynthetic pathway. While it is becoming clear that O-GlcNAc is essential to life and plays an impor-

tant role in diabetes and neurodegenerative diseases, little is known about how it functions at the molecular or metabolic level. Fundamental knowledge about O-GlcNAcylation is not only essential to understanding chronic diseases, such as diabetes and Alzheimer's disease, but also without an intimate understanding of O-GlcNAcylation, fundamental cellular physiology cannot be understood. O-GlcNAc glycosylation is nearly as abundant as protein phosphorylation, and the two have extensive cross talk between them to regulate signaling and transcription in response to nutrients and stress. O-GlcNAc also modifies cytoskeleton proteins and the contractile machinery in cells and muscles. Key unanswered questions include:

- Why is nearly every protein involved in transcription extensively O-GlcNAcylated?
- How does O-GlcNAcylation regulate gene expression?
- How is cellular physiology regulated by the interplay between phosphorylation and O-GlcNAcylation by the cross talk with other protein modifications, such as ubiquitylation?
- Current systems biology approaches vastly underestimate the complexity of signaling. To what extent are the human kinome expression and activity regulated by O-GlcNAcylation?
- How are ribosome biogenesis and protein translation regulated by O-GlcNAcylation?
- How does O-GlcNAcylation regulate cytoskeletal and contractile protein functions?
- How are the O-GlcNAc cycling enzymes regulated?
- Currently, the lack of facile tools to study O-GlcNAc is the greatest impediment to understanding the roles of O-GlcNAcylation in cellular physiology and disease. For example, the development of a large number of O-GlcNAc-dependent site-specific antibodies (such as are now available for phosphosites), the production of better inhibitors of the enzymes that control O-GlcNAc cycling, and the invention of imaging methods to visualized O-GlcNAc dynamics in a living cell would dramatically propel this field forward.

4.6 HOW DOES THE GLYCOCALYX AFFECT THE ORGANIZATION OF MOLECULES ON THE CELL SURFACE?

The surface of all cells in nature is decorated with a dense, complex, cell-type and tissue-specific array of glycan structures, known as the glycocalyx or, in the case of plants or fungi, the cell wall. Many advances have been made in understanding the structure of cell surface glycans by

releasing, purifying, and fractionating them and studying their structure in detail, an approach often called glycomics. While yielding extremely valuable and important information, this approach amounts to cutting down all the trees in the Amazon jungle and separating and identifying the trees by individual structure and type. This approach does limited justice to the intact forest that is the glycocalyx or cell wall.

Evidence to date indicates that the glycans present on cell surfaces are not always available for recognition by glycan-binding proteins in the same manner as they might be when isolated from one another in a glycan array. Rather, these glycans are present in complex mixtures, interacting with each other in various ways, involving glycan-glycan and glycan-protein interactions. In some instances the nature of these interactions has been recognized as functionally important. In other cases the recognition of a defined structure on a cell surface can vary enormously depending on the nature of the other glycans on the same cell surface. For example, three different glycan-binding proteins that recognize α-2-6-linked sialic acids and display identical binding patterns on glycan arrays show remarkably different patterns when binding to the surface of red blood cells. Moreover, these patterns are influenced by the ABO blood group status of the cell, which is determined by a different neutral glycan that does not even have sialic acid on it (Cohen and Varki 2010). Evidence such as this indicates that glycans present on the surface of cells can form *clustered saccharide patches* that are unique entities distinct from the individual glycans themselves. Current evidence also suggests that cell surface glycans regulate the lateral arrangements and associations of receptors on cell surfaces. The glycocalyx even regulates the local concentrations of cations and other small molecules. Today, evidence of such cell surface organization is mostly inferential, but in the future methods might be modified or new methods developed to try to probe such cell surface structures.

4.7 HOW CAN THE GLYCANS AND GLYCOPROTEINS ON A SINGLE CELL BE DETERMINED?

There are many cases in which the specific structures of carbohydrates on a single differentiated cell are desired, often because that cell resides among composite groups of cells in tissues made up of many cell types. Such is the case in many cancers in which tumor cells are mixed with normal cells, or in type 1 diabetes where islet beta cells are only one cell type among the islets, which are only a small component among the pancreatic exocrine cells. How can the glycans, glycoproteins, and other glycoconjugates be determined on a single cell? If techniques were available for investigators to rapidly determine the glycomes and

glycoproteomes of single cells rapidly and confidently, how would this revolutionize our understanding of their roles in various developmental, immunological, pathological, and structural processes? How dramatically would this deepen our understanding of important medical conditions, some currently thought of as incurable or involving difficult treatments, such as diabetes, cancer, autoimmune diseases, or drug-resistant infections? Currently available techniques to solve the structures of glycans largely arose out of analytical chemistry, yet they fall far short of providing the sensitivity required to analyze a single cell. A key challenge is to be able to rapidly catalog all of the glycans and glycoproteins on a single cell, whether it is microbial, plant, or mammalian. Many novel technologies will be important to develop in order to sensitively separate the many different glycans and glycoproteins in a single cell, to assess the purity and isomeric heterogeneity of glycans and glycoproteins in the resulting sample, and to confidently determine the structures of the molecules in that sample.

4.8 WHAT ARE THE FUNCTIONS OF MICROBIAL AND HOST INTERACTIONS INVOLVING GLYCANS?

Microbial and host interactions include microbial (pathogen, commensal) recognition of host (animal, plant) glycans, host recognition of microbial glycans, molecular mimicry of host glycans by pathogens, and microbial community interactions involving glycans.

Most pathogens gain initial access to hosts via recognition of host glycans. Conversely, glycosylation of bacteria and viruses plays multiple roles in host-pathogen interactions, including, but not limited to, shielding of immunodominant epitopes, stabilizing viral/bacterial proteins, and acting as sites of recognition for the innate immune system. Notable findings highlight the dramatic interplay between pathogen glycosylation, receptor recognition, and host immune response:

- There is now increased understanding and appreciation of the importance of glycan binding and recognition by lectins, such as the galectins, DC-SIGN, and MBL.
- Many of the pathogen-associated molecular patterns recognized by Toll-like receptors are glycoconjugates.
- There is increasing evidence that glycans themselves can be recognized by the adaptive immune response, including both B cells and killer T cells.
- Glycosylation plays a key role in receptor specificity and recognition by pathogens.

- Molecular mimicry of host glycans by pathogens provides novel virulence mechanisms.

Although there is a general appreciation that glycans play important roles in both pathogen escape and, conversely, activation of immune surveillance, a challenge is to provide a more detailed, mechanistic, and holistic understanding of the interplay between the glycosylation of pathogens and their virulence, evolution, and recognition by both innate and adaptive immune responses. This challenge can be answered by building on previous work in such fields as virology and microbiology, immunology, and structural biology and by utilizing new synthesis, sequencing, and enzymatic tools. Additionally, answering this challenge will require detailed descriptions, through a combination of analytical, structural, and data-mining approaches, of protein-glycan interactions, including those involving antibodies and lectins. Such knowledge will enable engineering of specific binding and recognition by tailored antibodies and lectins. Answering this challenge would not only dramatically increase our understanding of the evolution of infectious agents, including drug-resistant forms, but also would facilitate development of the next generation of vaccines and therapeutics.

4.9 HOW DO GLYCAN BINDING PROTEINS DECODE THE GLYCOME?

Major roles of glycans in health and disease are mediated by binding proteins (GBPs) that decode the information content of the glycome through their recognition of glycans as ligands (Sharon and Lis 2004; Bishop and Gagneux 2007; Crocker et al. 2007; Taylor and Drickamer 2007; Varki 2007; Imberty et al. 2008; van Kooyk and Rabinovich 2008; Taylor and Drickamer 2009; Rillahan et al. 2011). Thus, key to understanding the functions of glycans is elucidation of the functions of GBPs that recognize them. For example, mammalian GBPs mediate diverse biology, including trafficking of white blood cells to sites of inflammation, regulation of cell-signaling receptors, and aiding the immune system to distinguishing between self and nonself. Plant GBPs mediate defense against pathogens while also facilitating critical symbiotic relationships with bacteria required for essential processes, such as nitrogen fixation. GBPs of pathogenic and commensal microorganisms mediate attachment to glycan ligands as their receptors on host cells. Although there has been enormous progress to define the roles of exemplary GBPs over the past two decades, the sum of the knowledge to date represents the tip of the iceberg. With the exception of humans and other mammals, no systematic effort to identify GBPs has been made, particularly for microorganisms,

including the estimated 6,000 species of commensal bacteria that comprise the gut metabiome.

Progress in elucidating the roles of known GBPs and their mechanisms of action is hampered by a lack of robust tools to establish even the most essential information. Although glycan arrays have demonstrated their utility in defining the ligand specificity of a GBP, the number of glycans elaborated on the largest arrays, comprising approximately 600 glycans, represents a tiny fraction of the human glycome. Also lacking are arrays of microorganism glycans needed to define the specificity of GBPs from animals, plants, and other microorganisms that bind them. Similarly, although analytical methods exist to profile glycans of complex biological systems, these methods provide limited information, and methods to routinely determine the complete structures of glycans from biological materials are lacking. Also needed are glycan reagents to probe the functions of GBP-ligand interactions and to produce glycan-specific antibodies. However, there is an extremely limited supply of synthetic glycan reagents, and current methods of synthesis produce them in small amounts with great effort and resources.

To address these needs, better methods to synthesize glycans, to build libraries of natural glycans for the expansion of glycan arrays, and to produce a reagent bank or service to produce glycans as reagents for biological studies are needed. There should also be a concerted effort to develop analytical methodologies that will permit determination of the complete structures of glycans from biological samples.

4.10 HOW CAN PLANT RECALCITRANCE TO DEGRADATION BE UNDERSTOOD AND OVERCOME?

The major obstacle to perennial cellulosic materials serving as an economical sustainable source of liquid fuels is an inability to overcome the recalcitrance of cellulosic biomass to conversion into the sugar intermediates needed to produce liquid fuel (see Sections 3.2.1 and 3.2.2). In addition, the economic and efficient release of oligosaccharides and polysaccharides for use as replacements for petroleum-based materials is of considerable importance. An outstanding question in the area of biomass should be how to modify the cell wall, the major component of biomass, to make it less recalcitrant to processing for energy and biomaterial production. Such advances would be sufficient to create a cellulosic biofuels and biomaterials industry.

Lignocellulosic plant cell walls give shape and protection to plant cells, tissues, and organs. These cell walls have evolved resistance to microbial and enzymatic deconstruction, and it is this recalcitrance that is largely responsible for the high cost and slow kinetics of lignocellulose

conversion to energy or in the extraction of cellulose or other cell wall polymers. Current technologies to overcome recalcitrance have primarily been developed empirically with little knowledge of the biological and chemical properties of biomass. Increased understanding of plant cell wall biosynthesis and detailed structural information on cell wall components now provide unique insight for modifying cell wall biosynthesis and targeting the deconstruction of cell wall components for the production of bioenergy, bioproducts, and cellulose nanomaterials (see Section 3.2.1). However, the detailed architecture of the cell wall is far from complete. Currently, partial structures of pectin and hemicellulose matrix polysaccharides, cellulose, and lignin can be described, yet the description provides little insight into how these polymers interact to form an insoluble lignocellulosic cell wall.

Glycoscience can address this issue by developing more powerful characterization and modeling techniques, which could be used to advance research into cell wall biosynthesis, architecture, and deconstruction, as well as new methods for manipulating the genetics of biomass species and the microbes and enzymes that deconstruct biomass (see Sections 5.2, 5.3, and 5.4). Key questions to be addressed are:

- What is the detailed structure and what are the interconnection points in the wall between cellulose-hemicellulose-lignin-pectin-protein?
- What are the linkages or interactions that can be modified during synthesis or broken to gain access to the cellulose or matrix polysaccharides so that they can be extracted or broken down to sugars?
- What are the control points for cellulose synthesis and the cellulose microfibril structure, particularly as they pertain to the production of cellulose nanoparticles?
- What are the morphology, the crystal structure, and the surface chemical characteristics of the extracted cellulose nanoparticles and how can these be modified?

4.11 HOW CAN SUGARS BE REASSEMBLED TO DEVELOP MATERIALS WITH TAILORED PROPERTIES AND FUNCTIONALITY?

The ability to reassemble sugar units on demand to make new polymer structures will provide limitless opportunities to design materials that have tailored properties or that are functionality specific for a given application. The foundation of what gives a polymer its material properties is its chemical structure, and having the ability to strategically design the repeat sugar unit(s), the combination of different sugars, the

type of linkages between the sugars, the specific side groups that branch off the backbone structure, and properties that control self-assembly into domains or overall polymer structures gives tremendous flexibility to make new polysaccharide-based materials. Such a capability would allow scientists and engineers to develop the next generation of sugar-based polymers and nanomaterials with the desired mechanical properties, high-temperature stability, and on/off-switchable biodegradation mechanisms that the plastics industry needs to expand the numbers and types of sustainable and biodegradable products that society is increasingly demanding. One can imagine, for example, the significance of the development of a transparent plastic water bottle produced from renewable materials, with a long shelf life and the ability to switch "on" a biodegradation mechanism, when necessary, so that the bottle degrades by mechanisms abundant in nature.

Polysaccharide synthesis is currently limited to relatively small quantities of structures with relatively low degrees of polymerization and limited side-group functionalization. There are also difficulties in isolating specific polysaccharide structures, which complicates the structural characterization of these materials. These factors greatly impede the development of new polysaccharide-based materials. To achieve rapid prototyping of these materials, it will be necessary to expand the capabilities and interconnections between synthesis, characterization, and modeling tools. Areas for which further developments would be useful include the synthesis of polysaccharide polymers with higher degrees of polymerization, tailored backbone structure, and targeted side-group functionalization and the synthesis of greater than milligram-scale quantities. The ability to produce up to gram-size quantities of a homogeneous polymer structure is necessary to evaluate the role of a given derivation of the resulting material properties, such as structural configuration, thermal stability, reactivity, rheology, and mechanical characteristics. Being able to produce gram-size quantities of materials also greatly expands the number of available property characterization tools to test and evaluate it. Improvements in polysaccharide isolation procedures to produce homogenous samples, in characterization methods that have increased structural sensitivity and speed, and in predictive modeling to give additional insight into synthesis pathways and properties of the resulting materials, will be important as well.

4.12 SUMMARY

The questions posed in this chapter address several overarching themes, many of which reflect gaps in knowledge about fundamental biological and biochemical processes:

- understanding glycan diversity—how it arose, what it does, and how to study it;
- understanding the roles of glycans as modifiers of other biological molecules, such as proteins (What are the functions of glycan modifications on glycoconjugates, and how is this process controlled?);
- understanding the role of glycans inside cells—for example, in nuclear and cytoplasmic glycosylation;
- understanding the roles of glycans on cell surfaces—for example, as the glycome; and
- understanding how to control or manipulate glycans for desired ends—for example, to reduce plant cell recalcitrance to degradation or to design polymeric materials with new properties.

Although these areas address fundamental aspects of glycans, they also form a base from which to apply glycoscience knowledge to solve practical problems, such as how to better diagnose and treat diseases, how to create new fuels, and how to design improved products. Addressing the questions posed in this chapter has relevance to glycoscientists and nonglycoscientists alike. Chapter 5 turns to a discussion of the tools and technologies needed to help address these questions.

5

The Toolkit of Glycoscience

The incredible opportunities for glycoscience in health, energy, and materials science described in the previous chapters and in the questions posed in Chapter 4 can only be realized with a set of new analytical tools. Today, glycoscience is practiced by a relatively small community of biologists, chemists, and materials scientists. This community must expand if glycoscience is to extend its impact and become pervasive. The broader scientific community must participate in the development of the necessary tools that will transform the field and empower researchers in both the glycoscience field and the larger scientific community to incorporate glycoscience into their research pursuits as a matter of course. To this end, glycoscience needs new analytical tools, including methods development for separation, purification, characterization, localization, and structure identification. The tools used today are limited in their capabilities and will not enable realization of glycoscience's full potential. The best analytical chemists and other measurement scientists, including tools developers, need to turn their attention to glycoscience and bring their creativity to the field. They need to apply existing tools and methods that have not yet been applied to glycoscience, and they need to develop new tools to solve analytical problems that existing tools cannot address.

Similarly, the synthesis community needs to begin to embrace glycochemistry as an essential field of organic chemistry. Glycochemistry needs to be brought into the mainstream of synthetic organic chemistry rather than kept as a specialized field practiced by only a handful of glycochemists. New synthetic methods need to be brought to bear

on glycan synthesis, and the creativity of the entire synthesis community needs to be leveraged to solve the long-standing and vexing problems of stereoselective, regioselective syntheses with simple, high-yielding reactions. The biochemistry and genetics communities need to participate in identifying all enzymes and characterizing all pathways involved in glycan metabolism. Finally, computer scientists, modelers, and bioinformaticists need to be fully engaged. The community needs to set up glycoscience databases and integrate glycoscience into existing biological databases. Glycan and proteoglycan structure prediction and modeling tools need to be developed. Full interaction pathways must be developed to incorporate all aspects of glycobiology into systems biology. Details of these opportunities are described in the remainder of this chapter, but the main message is clear: Glycoscience needs the full participation of the broader scientific community to help develop tools that can solve some of the most vexing problems in glycoscience and to catalyze its integration into the scientific mainstream. By helping develop tools for glycoscience, it is expected that these tools will have follow-on benefits to all fields of science.

5.1 SYNTHESIS

5.1.1 General Aspects

The development of routine procedures for automated chemical synthesis of oligonucleotide fragments (DNA and RNA) and peptides has brought significant change to modern biology. Unfortunately, no general methods are available for the preparation of complex carbohydrates (Boltje et al. 2009; Kiessling and Splain 2010). As a result, the synthesis of a target is often a research project unto itself, which may take many months and in some cases years to complete. This problem is compounded by the fact that glycoconjugates in biological samples are often found in low concentrations and in microheterogeneous forms, greatly complicating their isolation and characterization. Glycomes of eukaryotic organisms are extremely diverse; for example, it has been estimated that the human glycome contains 10,000 to 20,000 minimal epitopes for glycan-binding proteins (Cummings 2009). Thus, robust synthetic technologies are urgently needed that can readily provide large collections of complex oligosaccharides. Furthermore, biological and analytical studies often require glycans to be modified by a tag, immobilized to surfaces, presented at a multivalent scaffold, or attached to a lipid, peptide, or protein (Seeberger and Werz 2007; Rich and Withers 2009). As a result, additional technologies are required that can readily provide such conjugates.

Current approaches for obtaining well-defined oligosaccharides and

glycoconjugates include chemical synthesis, enzymatic and chemoenzymatic synthesis, and microbial production (Boltje et al. 2009; Kiessling and Splain 2010; Hsu et al. 2011; Schmaltz et al. 2011). The next sections cover the scope and limitations of these methodologies. Despite the shortcoming of these technologies, they have been instrumental in addressing a number of important problems in glycobiology research and for the discovery of vaccines and therapeutics. In particular, the Consortium for Functional Glycomics, funded by the National Institute of General Medical Sciences, has employed a chemoenzymatic approach for the preparation of a collection of approximately 600 glycans derived from *N*- and *O*-linked glycoproteins and glycolipids (Stevens et al. 2006; Rillahan and Paulson 2011). These compounds are modified with an artificial aminopropyl linker, which allows covalent attachment to *N*-hydroxysuccinimide-activated glass slides. The resulting microarrays have found wide utility for integrating binding specificities of a diverse range of glycan-binding proteins, determining dissociation constants and dissecting binding energies, and analyzing multivalent and hetero-ligand binding. The species-specific nature of the interaction between virus and host glycans and determination of ligand specificities of monoclonal antibodies have allowed use of glycan arrays in rapid assessment of influenza virus receptor specificity. A significant barrier to widespread use of glycan arrays, however, is the limited availability of well-defined oligosaccharides, and current arrays contain only a fraction of naturally occurring oligosaccharides. Also, very similar arrays displaying very similar glycans can, nevertheless, provide significantly different results with regard to GBP binding. There are exciting challenges ahead before glycan arrays can become a standardized method of analysis.

Development of a fully synthetic heparin fragment for treatment of deep vein thrombosis exemplifies the importance of the organic synthesis of glycans. Heparin and heparan sulfate are naturally occurring linear polysaccharides that are modified by sulfate esters. A heparin-derived pentasaccharide that can bind to antithrombin III (AT II) and that exhibits anticoagulant activity has been identified. A fully synthetic analog (fondaparinux) of this domain has been developed, which is being produced on a multikilogram scale to treat deep vein thrombosis (Petitou and van Boeckel 2004). In contrast to porcine mucosal tissue-derived heparin, the synthetic compound is easy to characterize and has a much-improved subcutaneous bioavailability. The importance of synthetic oligosaccharides for anticoagulation therapy was highlighted by the recent discovery of batches of heparin that caused hypotension and resulted in nearly 100 deaths. These reactions resulted from contamination with oversulfated chondroitin sulfate, which is a popular shellfish-derived oral supplement for the treatment of arthritis (Guerrini et al. 2008). The ability to synthe-

size pure, well-characterized glycans would eliminate the need to rely on poorly characterized and highly variable glycans obtained from natural sources.

Heparin and heparan sulfate are examples of glycosylaminoglycans (GAGs), which have been implicated in many other biological processes and can have pronounced physiological effects on lipid transport and adsorption, cell growth and migration, and development (Bishop et al. 2007). Significant changes in the structure of GAGs have been observed in the stroma surrounding tumors, which is noteworthy when considering tumor growth and invasion. GAGs are also involved in neurobiological processes and, for example, have been implicated in neuroepithelial growth and differentiation, neurite outgrowth, nerve regeneration, axonal guidance and branching, deposition of amyloidotic plaques in Alzheimer's disease, and astrocyte proliferation. Large arrays of well-defined heparan sulfate oligosaccharides are needed to identify compounds that mediate or inhibit these processes. It is possible that synthetic analogs of heparin may find application in the treatment of several neurological diseases, cancer, and infection.

Synthetic oligosaccharides have also been used in the development of vaccines for such diseases as *Haemophilus influenzae* type b, HIV, *Plasmodium falciparum*, *Vibrio cholerae*, *Cryptococcus neoformans*, *Streptococcus pneumoniae*, *Shigella dysenteriae*, *Neisseria meningitides, Bacillus anthraces,* and *Candida albicans* (Costantino et al. 2011; Morelli et al. 2011). Polysaccharides isolated from natural sources, which are conjugated to a carrier protein, are used in prevention of life-threatening bacterial infectious diseases such as meningitis and pneumonia. However, the wide utility of this approach is limited by such problems as the destruction of vital immunodominant features during the chemical conjugation to a carrier protein. Furthermore, isolated polysaccharides often display structural heterogeneity, which may complicate reproducible production. These compounds may also contain toxic components or immunosuppressive domains that may be difficult to remove. These problems can be addressed by using chemically or enzymatically synthesized glycan epitopes. In such an approach a synthetic oligosaccharide is equipped with an artificial spacer to facilitate selective conjugation to a carrier protein. In general, antibodies recognize epitopes no larger than a hexasaccharide, and compounds of this complexity can be readily obtained by organic synthesis. The recent approval of *Haemophilus influenzae* vaccine based on a synthetic glycan epitope highlights the potential use of organic synthesis for the development of glycoconjugate vaccines. The expansion of this capability to other vaccines would have a tremendous impact on both safety and efficacy and could potentially compress the timeframe for developing new vaccines, especially for new threats.

Synthetic oligosaccharides and glycopeptides are also being used in the development of cancer vaccines (Buskas et al. 2009). Oncogenic-transformed cells often overexpress oligosaccharides such as Globo-H, Lewis^{Y}, and Tn antigen, and numerous preclinical and clinical studies have demonstrated that naturally acquired, passively administered, or actively induced antibodies that recognize glycan-associated tumor antigens are able to eliminate circulating tumor cells and micrometastases. The development of tumor-associated polysaccharides and glycopeptides as cancer vaccines has been complicated by the fact that they are self-antigens and therefore are tolerated by the immune system. The problem of self-tolerance is being addressed by the design, chemical synthesis, and immunological evaluation of fully synthetic vaccine candidates.

Synthetic oligosaccharides are also used for the preparation of glycopolymers, glycodendrimers, and glyconanoparticles (Garcia et al. 2010). These materials are receiving considerable attention because monovalent polysaccharides often exhibit weak affinities for their protein receptors. However, glycan-binding proteins often exist as higher-order oligomeric structures presenting multiple binding sites, acting as "polydentate" donors, and thereby circumventing the intrinsic weak binding interactions of monovalent ligands. Also, gold nanoparticles, quantum dots, and magnetic nanoparticles provide additional functionality as they allow detection by SPR or fluorescence or make it possible for convenient isolation by using a strong magnetic field.

An increasing number of drugs contain glycans or glycomimetics as a major component. Examples include many antibiotics, antiviral drugs such as Relenza and Tamiflu, hyaluronic acids, and selectin antagonists (Gantt et al. 2011b). Synthetic oligosaccharides are also being used in the preparation of well-defined glycoproteins. Approximately one-quarter of new approvals are protein-based drugs, with a majority being glycoproteins. The glycan moiety of glycoproteins plays an important role for its pharmacokinetic properties. Hence, it is critical to control the exact chemical composition of the oligosaccharide moieties of glycoproteins. Protein glycosylation is, however, not under direct genetic control and results in the formation of a heterogeneous range of glycoforms that possess the same peptide backbone but differ in the nature and site of glycosylation. In general, it is difficult to control glycoform formation in cell culture, which is a major obstacle for the development of therapeutic glycoproteins.

In summary, a diverse set of glycan structures can be used to discover specific inhibitors of glycosyltransferases with pharmaceutical applications, including diagnostics. Large arrays representing the diversity of glycans or focused on a specific set of glycan structures could be used for drug screening and discovery. Access to diverse glycans via synthesis

for preparing such arrays would lead to new uses that have yet to be imagined.

5.1.2 Synthetic Tools

5.1.2.1 Chemical glycan synthesis

The realization that complex oligosaccharides and glycoconjugates are involved in numerous biological processes has stimulated development of chemical and enzymatic methods for the preparation of oligosaccharides. Unlike oligonucleotide and peptide synthesis, there are no general protocols for the preparation of glycans. As a result, the synthesis of specific targets is often a demanding and time-consuming task (Boltje et al. 2009; Kiessling and Splain 2010). However, recent technological advances are making it possible to streamline the process of oligosaccharide assembly and are providing opportunities to prepare collections of oligosaccharides and glycoconjugates.

The chemical synthesis of glycans involves coupling a fully protected glycosyl donor, which bears a leaving group at its anomeric center, with a suitably protected glycosyl acceptor that often contains only one free hydroxyl (Zhu and Schmidt 2009). The result of this chemical reaction is a glycoside product. The process of sequentially generating glycosyl donors and acceptors can be repeated until a complex target has been obtained. Preparation of monosaccharide building blocks requires extensive protecting-group manipulations, and typically 6 to 10 chemical steps are needed to create each building block. Because preparation of the monomer building blocks consumes the majority of effort invested in chemical glycan synthesis, rapid and inexpensive access to building blocks should greatly accelerate chemical glycan synthesis. One approach to speed up monosaccharide synthesis involves parallel combinatorial sequential one-pot multistep procedures for the selective protection of monosaccharides (Wang et al. 2007). This approach can incorporate as many as seven chemical steps, obviating the need to carry out intermittent tedious work-up and time-consuming purifications. A complementary approach involves identification of monosaccharide building blocks that can be used repeatedly for the synthesis of a wide range of target structures. For example, it has been proposed that 88 percent of glycoside motifs found in mammalian glycoconjugates could be constructed from only 20 different monosaccharide building blocks, which could then be prepared in bulk for widespread use in automated glycan synthesis technology (Werz et al. 2007). In addition, disaccharide building blocks have been identified that resemble the saccharide motifs found in heparan sulfate and that can be used repeatedly for rapid assembly of libraries of heparan sulfate

oligosaccharides. This building block approach could be extended to many other classes of oligosaccharides, although because of the enormous structural diversity of natural glycans, it will be important to focus on glycomes of particular interest.

The success of the unified building block approach relies on the premise that each glycosylation will be high yielding. In practice, this has been shown to not be the case, and additional saccharide building blocks are required to address possible synthetic difficulties. Sets of orthogonal protecting groups are being developed that provide additional synthetic flexibility in that they offer the possibility to change the order of glycosylation. Stereoselective installation of glycosidic bonds in high yield is a major challenge in complex oligosaccharide synthesis. In recent years this aspect of oligosaccharide synthesis has progressed considerably, and a wide range of stable yet highly reactive anomeric leaving groups have become available, making it possible to examine several glycosylation protocols to achieve optimal results (Zhu and Schmidt 2009). By exploiting neighboring-group participation, steric and conformational effects, or direct displacement of leaving groups, glycosides can be obtained with high anomeric selectivity, even for the more challenging α-sialyl and β-mannosyl linkages. However, glycoside products are often contaminated by unwanted anomeric products, making it necessary to include time-consuming purification protocols.

Minimizing purification steps has been the focus of efforts to streamline chemical synthesis of oligosaccharides. Approaches that are being pioneered include one-pot multistep solution-phase glycan synthesis, solid-phase glycan synthesis, and fluorous tagging. One-pot multistep procedures are based on the sequential addition of glycosyl donors with well-defined anomeric reactivity to a reaction flask to provide an oligosaccharide without the need to purify synthetic intermediates (Kaeothip and Demchenko 2011). Although many variations of the one-pot strategy have been developed, there are three major concepts: chemoselective, orthogonal, and preactivation glycosylation strategies. In chemoselective glycosylation strategies, glycosyl donors with decreasing anomeric reactivity are allowed to react sequentially. Orthogonal glycosylations use glycosyl donors and acceptors that have different anomeric groups that can be activated without affecting each other. A flexible approach that exploits the advantages of the aforementioned strategies utilizes preactivation of a glycosyl donor to generate a reactive intermediate in the absence of the acceptor. After addition of the glycosyl acceptor, a glycoside product can be formed that has an identical leaving group at the reducing end. In the same reaction flask the process of anomeric activation and glycosylation can be repeated to construct complex oligosaccharides. Successful implementation of this strategy requires that the promoter be completely con-

sumed to prevent activation of a subsequent saccharide building block. Furthermore, the reactive intermediate should be sufficiently long lived to permit addition of a glycosyl acceptor yet sufficiently reactive for a high-yielding glycosylation. It has been difficult to design glycosylations that meet these requirements.

Encouraged by successes with polymer-supported peptide synthesis, the first attempts at solid-phase oligosaccharide synthesis were reported in the 1970s. These efforts were not successful, largely because of a lack of efficient glycosylation methods. The past decade has seen a renewed interest in polymer-supported glycan synthesis, and different polymer support materials, linkers, synthetic strategies, and glycosylating agents have been explored (Seeberger 2008). However, a general solution for routine and automated oligosaccharide synthesis remains to be established, in large part because of the need for large excesses of glycosyl donors, the lack of anomeric control when 1,2-*cis*-glycosides need to be installed, the unpredictability of glycosylations, and the additional steps required for linker functionalization and protecting-group removal. However, progress is being made in these areas, bringing the promise of routine automated oligosaccharide synthesis closer to fruition. In particular, it has been shown that automated synthesis can provide complex oligosaccharides such as a branched β-glucan dodecasaccharide, blood-group oligosaccharides, and tumor-associated glycan antigens.

Solution-based strategies have been developed in which the growing oligosaccharide chain is modified by a tag that allows selective precipitation, extraction, or absorption for convenient purification. In particular, light fluorous tagging technology is attractive because it makes possible protecting-group manipulations and glycosylations under conditions typically used in solution-phase chemistry (Jaipuri and Pohl 2008). In this approach, tagged products can be selectively captured by a fluorous solid-phase extraction column and then released by elution with methanol. Fluorous-tagged glycans can also be directly printed on fluorocarbon surfaces, providing interesting opportunities for glycan array development. Efforts are under way to automate fluorous-supported synthesis of oligosaccharides with liquid handling devices.

Despite considerable progress, chemical synthesis of glycans remains a challenging endeavor that is practiced only by expert laboratories. The lack of large collections of universal building blocks and the need to optimize glycosylation conditions complicate routine synthesis of this class of compounds. There is an urgent need for the development of more reliable glycosylation protocols, which might be accomplished with a better understanding of the mechanistic aspect of glycosylations. Furthermore, new approaches need to be developed for controlling anomeric selectivities of glycosylations. In particular, reliance on exten-

sive protection-deprotection schemes adds to the number of steps and reduces the yields of complex glycans. High-throughput protocols for the rapid evaluation of many reaction conditions—for example, by employing microfluidics devices—may provide more reliable glycosylation protocols. It is also to be expected that searchable databases of reported glycosylations can accelerate the optimization process and may lead to standardized protocols.

5.1.2.2 Enzymatic synthesis of glycans

Enzyme-catalyzed glycosylations offer an approach that is complementary to chemical synthesis for obtaining structurally well-defined oligosaccharides, polysaccharides, and glycoconjugates. Current enzymatic and chemoenzymatic approaches apply glycosyltransferases, glycosidases, glycosynthases, and *trans*-glycosidases to construct glycosidic linkages, whereas enzymes such as sulfotransferases, epimerases, and acetyltransferases have been used for postglycosylation modifications (Schmaltz et al. 2011). (See also Section 5.4.2.1, which discusses glycan synthesis in the context of glycoenzyme applications.)

Glycans are synthesized in an assembly-line manner by enzymes such as glycosyltransferases. In this process the product of one enzyme becomes the substrate of the next. The glycosyltransferases form a connection, termed the glycosidic linkage, between a growing glycan chain and another sugar building block. The most common building blocks for glycosyltransferases are called nucleotide sugar donors, and the structures to which those building blocks are added are generally referred to as glycosyl acceptors. In general, there is a unique glycosyltransferase for nearly every type of glycosidic linkage formed, and these enzymes are among the most specific enzymes known. They are able to distinguish the spatial orientation of a single atom, even on very large structures, within both their glycosyl donors and acceptors.

As a result, glycosyl transferases are essential enzymes for oligosaccharide biosyntheses. Their ability to transfer a sugar residue from a sugar-nucleotide mono- and di-phosphate to a maturing oligosaccharide chain (Lairson et al. 2008) and their highly regio- and stereoselective nature make them ideally suited for the preparation of complex oligosaccharides. Currently, more than 50,000 genes encoding potential glycosyltransferases have been identified, although only a very small number have been characterized. Many of these enzymes are membrane bound and glycosylated, making their isolation and utilization difficult. The number of glycosyltransferases with good catalytic activity and defined substrate specificity that can be expressed as a soluble form in large amounts is currently small, which hampers efforts to develop enzymatic

glycoconjugate synthetic schemes. However, several activities are under way to address these problems. High-throughput assays are being developed to identify activities and substrate specificities of glycosyltransferases. Furthermore, it has been found that glycosyltransferases from bacterial sources often exhibit considerable substrate promiscuity, thereby offering unique opportunities for chemoenzymatic synthesis of glycan libraries and their derivatives. The substrate specificity of glycosyltransferases can be altered by protein crystal structure-based rational design or directed evolution. However, glycan-modifying enzymes are exceptionally underrepresented in structural databases, particularly as enzyme substrate complexes, and the resulting incomplete and fragmentary collection of biochemical and structural data on this class of enzymes has led to an incomplete understanding of the molecular mechanisms that control oligosaccharide biosynthesis. Recently, the Repository of Glyco-enzyme Expression Constructs was created to focus on generating expression vectors that encode all human glycosyltransferases and glycoside hydrolases as well as a limited set of glycan-modifying enzymes for production in bacteria, insect cells (baculovirus), and mammalian cells. The goal of the repository is to facilitate the production of soluble forms of enzymes as catalytic domains, when possible, for use in biochemical, enzymatic, and structural studies. Many of the constructs have been designed as truncated forms devoid of the transmembrane protein domain and linked to affinity tags or other larger fusion proteins to facilitate affinity purification and quantification.

Several convenient approaches have been developed for the preparation of sugar nucleotides, key substrates for all glycosyltransferases, including in situ sugar nucleotide regeneration, fusion protein strategies, one-pot multienzyme systems, and superbead technologies. Progress is being made in identifying and characterizing many sugar nucleotide biosynthetic enzymes, including those involved in salvage pathways. However, many uncommon sugar nucleotides for glycosyltransferase-catalyzed synthesis of glycosylated natural products are less accessible because of their much more complicated biosynthetic pathways and instability.

The combined use of glycosyltransferases, sulfotransferases, and epimerase has been successfully implemented for the synthesis of structurally defined heparin and heparan sulfate oligosaccharides, as well as for polysaccharides with specific sulfation patterns. In particular, these developments have led to the production of ultra low molecular weight heparins in 10 to 12 steps, with an overall yield of 40 to 50 percent (Xu et al. 2011). This compares well with the process now used to make the anticoagulant drug Arixtra, which involves some 50 steps and has an overall yield of less than 1 percent.

Currently, microfluidics and microarray formats are being explored to conduct enzymatic syntheses. These types of approaches can be used to create a wide range of products that can be analyzed using mass spectrometry and then assayed for biological activity. At the other extreme, work is being conducted to develop macroscale enzyme-assisted syntheses of heparin, although a major obstacle in this effort is the cost of a critical cofactor—3′-phosphoadenyl-5′-phosphosulfate (PAPS)—which donates the high-energy sulfate groups that are covalently attached to the heparin backbone to make bioactive heparan sulfate. One solution is to regenerate PAPs in situ from the byproduct 3′-phosphoadenyl-5′-phosphate (PAP) enzymatically, a process similar to that used in large-scale oligosaccharide synthesis with sugar nucleotide regeneration.

Unlike chemical synthesis, the synthesis of unnatural saccharide sequences is a challenge for enzyme-based methods as a result of the strict substrate specificities of most glycan-synthesizing enzymes. To overcome this limitation, additional studies of heparan sulfate biosynthetic enzymes are necessary, especially to advance our understanding of the substrate specificities and to conduct mechanistically based mutagenesis to engineer the specificities.

Glycosyl hydrolases are a class of enzymes that degrade oligosaccharides by cleavage of glycosidic linkages. The reverse hydrolytic activity of this class of enzymes can be exploited for glycosidic bond synthesis. This approach suffers, however, from relatively low yields because of the challenge of driving reactions in a thermodynamically unfavorable direction. This problem has been addressed by the introduction of "glycosynthases," which are glycosidases rendered hydrolytically incompetent by replacement of a nucleophilic aspartic or glutamic acid with an alternative unreactive amino acid. Glycosynthases can, however, transfer activated glycosyl substrates that have the opposite anomeric configuration of the natural substrate. For example, β(1,4)-mannans, which are major plant cell wall polysaccharides, have been prepared by using a mutant endo-b-mannanase and an α-mannobiosyl fluoride as glycosyl donor. In addition, glycosynthases have been used for the preparation of β-linked glucuronic and galacturonic acid conjugates. Currently, only a small number of glycosynthases have been developed, which limits the scope of this technology. Recent studies have shown that the catalytic activities and substrate promiscuities of glycosynthases can be improved by directed evolution.

5.1.3 Manipulating Glycans by Pathway Engineering

All cells have endogenous machinery for synthesizing glycans and glycoconjugates. The glycan-modifying infrastructure in cells includes glycosidases, glycosyltransferases, mechanisms for activated sugar syn-

thesis and transport, and supporting functions, which are under coordinated control to orchestrate the formation of precise glycan structures. The vast majority of glycans and glycoconjugates pose significant challenges to synthetic endeavors. As such the goal of pathway engineering approaches is to manipulate the biosynthetic power of cells to transform them into synthetic tools for the production of specific glycan products. Over the past 10 years, glycoengineering approaches have provided new routes to producing human glycoproteins with defined glycoforms by manipulating mammalian and other cellular glycosylation pathways, by creating glycoprocessing enzymes with novel properties, and by modifying nonmammalian systems. Successes include humanizing the *N*-glycosylation pathways of yeast, insect, and plant cell systems and introducing protein glycosylation pathways into *E. coli*. Glycoengineering efforts have also transformed bacteria and other microbes into factories for glycan production and natural product modification and into tools to study and thwart glycan-related pathogenic mechanisms. Key to these strategies is the manipulation and augmentation of the endogenous cellular glycoprocessing infrastructure with novel enzyme activity. This involves the engineering of several genes for biosynthesis, activation, transport, and transfer of mono- and oligosaccharides.

Mutation, knockout, and inhibition present possibilities for controlling or simplifying *N*- and/or *O*-linked glycosylation profiles in mammalian cell lines. Within this category, cell lines with genetic mutations in glycosylation pathways have been a valuable resource for studying glycobiology and continue to represent powerful tools for producing glycoconjugates with more tailored glycans. Although not technically glycoengineering, RNA interference and small-molecule inhibitors of glycosylation also can be used to knock down the activity of glycoprocessing enzymes and produce simplified glycan structures for further elaboration.

Nonmammalian systems demonstrate the power of domain engineering to provide novel catalytic activity in the secretory pathway. In yeast and plant systems, domain engineering has been critical to introducing the necessary mammalian-type glycoprocessing activity to generate mammalian *N*- and *O*-linked glycans, a process called "humanizing." Metabolic glycoengineering of *E. coli* has been a successful method of producing human complex glycan structures, too. These schemes focus on generating the glycan component in the bacterial cytosol by using microbial glycoprocessing enzymes. Key steps include engineering an appropriate glycan acceptor scaffold that cannot be metabolically diverted and engineering appropriate glycoprocessing infrastructure to build on acceptor scaffolds, including glycosyltransfer enzymes and supporting nucleotide sugar synthesis and transport.

Over the past 10 years, two primary applications of metabolically glycoengineered *E. coli* have been demonstrated. In one application, *E. coli* are transformed into factories for the production of high-value glycans. In another, glycan structures are displayed on chimeric lipooligosaccharides at the bacterial cell surface, which can be used to study bacterial glycomimicry, to develop probiotics, and for drug screening. Both approaches demonstrate that biologically relevant glycan epitopes can be efficiently generated using very different engineered acceptor molecules, suggesting that further applications for metabolic glycoengineering can be developed through the engineering of appropriate acceptor molecules, perhaps even glycoproteins. Metabolically glycoengineered *E. coli* and cells are compatible with large-scale fermentation and can provide access to significant quantities of complex glycans, making them of interest for both research and clinical applications.

5.1.4 Synthesis of Standards for Mass Spectrometry

Quantification and characterization of complex mixtures of glycans are underdeveloped, complicating glycome and glycoproteome analysis of biological samples. The current practice for quantification is that individual components of glycan profiles generated by mass spectrometry are quantified relative to each other. While this normalized parameter has been broadly useful for comparing major changes in profiles across samples or biological conditions, its value for characterizing minor glycans is frequently subject to undue influence by the major components, as teasing numerator and denominator effects apart can be difficult for low-abundance but biologically significant glycans. Generating methods for absolute quantification by mass spectrometry would allow profile changes to be assessed glycan by glycan, independent of variations in the whole profile. A set of well-characterized oligosaccharide standards is essential to realize this potential. The availability of glycan standards that contain isotopic labels would add an additional level of utility and would make it possible to accurately identify and quantify glycans in complex biological samples.

In addition to facilitating glycan quantification, well-characterized oligosaccharide standards would provide substrates for elucidating fragmentation rules for multiple mass spectrometry schemes. The assignment of glycan structures is challenging because of the isobaric nature of glycans—that is, different glycan structures can have identical molecular weights. For instance, at the monosaccharide level, all hexoses have the same molecular weights, as do the corresponding HexNAc derivatives. For extended structures the sequences are often identical but with different branching patterns. Also, glycosidic linkages can have two anomeric

configurations, which cannot be assigned by simple mass determination. It is, however, expected that this problem can be addressed by detailed mass spectrometry analysis of well-defined glycan standards. In this respect, isomers of oligosaccharides should produce unique fragmentation patterns that can be used for compound identification. A comprehensive analysis of a wide range of oligosaccharide standards will also create invaluable information for the bioinformatics community to build programs similar to the high-throughput software, such as the MASCOT and Sequest programs, now used to identify peptide sequences from proteomics data.

5.1.5 Key Messages on Glycan Synthesis

Well-defined complex oligosaccharides can be obtained by chemical synthesis, enzyme catalyzed reactions, and fermentation. Over the past 30 years, tremendous advances have been made in chemical and enzymatic synthesis of glycans. However, no routine synthesis processes exist for glycans. As a result, synthetic glycans remain relegated to small quantities and specialized laboratories, and current methods are not sufficiently robust to permit preparation of large collections of compounds for analytical and structure-activity relationship studies. Technology for the preparation of well-defined glycoconjugates, such as glycoproteins, is still in its infancy. For glycoscience to move ahead, significant further progress in synthesis will be needed. This progress is likely to include better methods for specific and selective glycosylation with less reliance on both protecting groups and linear strategies, the ability to obtain and characterize more enzymes with the requisite specificities for making any glycosidic linkage, and improved understanding of the genes involved in glycan synthesis, along with tools for engineering new pathways to make any desired glycan structure.

5.2 ANALYSIS

The "analysis" tools discussed in this report cover a broad range of chemical, physical, biological, computational, mathematical, and engineering techniques, albeit necessarily focused on applications to areas of glycoscience. However, many advances in analytical tools to dramatically improve current techniques are anticipated to draw on a broad base of expertise and talents, potentially from a wide-ranging group of scientists with markedly different backgrounds from those of current glycoscientists. Many unanticipated developments and novel ideas relevant to the areas of analysis detailed below may spring from scientists devoted to a variety of analytical methods outside what might now typically be used

or even considered by glycoscientists. This report therefore stresses that involvement of the broader analytical community should be welcomed in the development of novel technologies to address current problems and shortcomings, as outlined below, and that participatory collaborations be fostered between the broader analytical community and glycoscientists.

Because many truly transformative new technologies may be difficult to envision or even imagine, it can be anticipated that the requirements for individuals with skills from varied backgrounds, the degree of overlap in their expertise, their numbers, the collaborative modes, and the resources required to implement interesting new techniques may be equally difficult to foresee. Thus, the field will need to be flexible about the ideas and the nature, size, and types of collaborations necessary to meet specific challenges, with the consideration that at least some may be potentially high-risk yet high-payoff developments.

The topic of analysis can be subdivided into components that focus on several key aspects:

1. analysis of glycan molecules themselves, including techniques for their disassembly, separation, analysis of purity, and analysis of primary and three-dimensional structures;
2. analysis of glycoconjugates, which includes analysis of the molecular components of glycans conjugated to other molecules such as peptides, proteins, and lipids, particularly for mucins, proteoglycans, peptidoglycans, and lipopolysaccharides;
3. for many types of glycans, analysis of their interactions with proteins or higher-level structural interactions that occur in many types of cell walls;
4. for some glycans, analysis of their roles in metabolic functions related to a cell or an organism's pathways of energy utilization;
5. analysis of the relationship of glycan structures to the genome, which includes analysis of the enzymes that synthesize and degrade glycans—the glycosyltransferases and glycosidases, the precursor nucleotide sugars—in addition to understanding the phenotypic effects observed in cells/organisms as a result of their mutation and/or cell-specific genetic ablation; and
6. analysis of the locations of specific glycan structures in cells or tissues of organisms through molecule-specific imaging techniques.

Ultimately, the goal of improving structural techniques, which is related to the first two subtopics, is to understand the roles of glycans in various biological processes, including their interactions (subtopic 3), how their synthesis and degradation are controlled (subtopic 4), for some glycans their roles in metabolism (subtopic 5), and where specific structures

are expressed in organisms (subtopic 6). Each of these items is addressed in more detail in the following sections.

5.2.1 Analysis of Primary Glycan Structures

Perhaps the greatest advancement in accelerating the fields of genomics and proteomics was the development of accurate, sensitive, and rapid methods for determining the primary structures of these biopolymers. Indeed, the development of automated, enzyme-based sequencing technologies launched both genomics and proteomics as viable and productive fields of research.

The field of glycan sciences offers a different challenge because of the diversity of monomers across organisms and the inherent variations in the way these monomers are connected to one another. To begin with, sugar building blocks are not the same for many divergent groups of organisms. The monomers are isomeric and can include different ring sizes (five- or six-membered), anomeric configuration (α or β), absolute configuration (D or L), and aldoses versus ketoses (a carbonyl group at C-1 as opposed to C-2 or other positions). The monomers can also contain many variants of the basic sugar molecule, including many variants of deoxysugars (at different positions) and other substitutions of hydroxyl groups, with, for example, amino, sulfate, phosphate, acyl or alkyl functional groups (Schaffer 1972; Horton and Wander 1980; Williams and Wander 1980; Kremer and Gallo-Rodriguez 2004). In addition, sugars link to each other via glycosidic bonds in which the anomeric carbon (usually C-1 or C-2) of one sugar can potentially attach to any of the hydroxyl groups of another sugar. Moreover, each monomer can be linked to more than one other sugar, leading to the formation of branched polysaccharides.

Nucleic acid and protein analysis made major advances with the discovery of enzymes that selectively cleave large molecules into smaller ones, which can be more easily sequenced. Analogous enzymes, called "endoglycanases," exist in nature, as do enzymes that selectively cleave glycans from proteins or lipids. Isolating these enzymes and harnessing their ability to selectively cleave glycans at specific linkages could provide a critical set of tools for glycan sequencing efforts. It may also be possible to develop selective chemical reactions, perhaps with site-specific catalysts, to facilitate the deconstruction of glycans into smaller polysaccharides or monosaccharides that could be more easily sequenced. In this regard the development of nanopore sequencing, in which linear DNA strands are translocated through a pore as the bases are read, portends a similar approach for linear glycans.

Once glycans are broken down into smaller pieces, it will be necessary to separate the resulting sets of polysaccharides and monosaccharides for

analysis. These mixtures will contain not only molecules with different numbers of sugars but also many sets of isomeric molecules that might contain not only the same sugars with different linkages between them but also individual sugars that vary at a single stereochemical site. While current technologies, including high-performance liquid chromatography, are adept at separating polysaccharides that differ in the number or type of sugars, current technologies are not as successful at separating isomeric structures. Many other separation tools, such as immobilized glycan-binding protein columns, electrophoresis, and ion mobility separation of glycans as ions, among others, are being investigated, but today the separation and assessment of isomeric heterogeneity are the most time-consuming parts of structural determination.

The types of techniques likely to be needed in glycoscience include those that are "purely analytical" and those that are "analytical with the intent of providing structural proof of individual molecules." Purely analytical techniques include highly sensitive, high-resolution, and multidimensional methods for rapid assessment of isomeric heterogeneity, as well as tools capable of separating molecules or providing high-resolution spectroscopic information to provide evidence of purity. Frequently, molecules may co-migrate using one separation system, and therefore co-migration in a single system is not a valid criterion for purity or proof of structure. Possible options might include systems that concatenate multiple separation modes and other tools that can discriminate isomeric glycan molecules.

Separation methods coupled with methods that provide structural information include automated multidimensional methods paired with tools for primary structural determination. Whether samples are analyzed online or fractions are collected and analyzed afterward, rapid, automated, multidimensional separations that can generate pure samples with high probability and that can be applied as broadly as possible to structures from animal, plant, bacterial, and fungal sources would be useful for the field. Current methods, such as collection of multiple fractions from one chromatographic or electrophoretic system, concentrating them, then applying the molecular components in these fractions to another chromatographic separation, followed by another concentration of many more fractions, applying to another column, and so forth, proves very laborious and time consuming. Such samples can then be subjected to tools that assess purity and establish primary structures.

Owing to the great diversity of natural structures in different taxonomic kingdoms, primary structural determination or "sequencing" takes on a completely different nature when quite possibly none of the same monomers might be found between two organisms from different kingdoms or even within the same kingdom (such as prokaryotes). Further-

more, there are probably many biological systems, unexamined as yet, in which novel monomers will be discovered. Glycan structures contain a large number of stereocenters, and detailed assignment of their structures is currently best achieved by a combination of methods that provide orthogonal structural information.

Many developments in techniques that can prove structure will be needed, which may include but need not be restricted to the following:

5.2.1.1 Methods for complete disassembly of larger glycans to their monomers and better tools for determination of monosaccharide structures

This includes all aspects of their structures: (1) confident assignment of their stereochemistry; (2) confident assignment of modifications at different hydroxyl positions, such as acyl or sulfate groups; (3) determination of their enantiomers (D- or L-forms); and (4) confident determination of whether sugar components are aldoses or ketoses. Better techniques to quantitate monosaccharides, with a wide availability of standards, are needed for a wide variety of monosaccharide compositional analyses in the future.

5.2.1.2 Methods for determination of sugar ring forms (five- or six-membered) in oligo- and polysaccharides

In nature, both furanoside (five-membered) and pyranoside (six-membered) ring forms are present as glycosides in a wide variety of larger glycan molecules, and both types are frequently observed in the same molecule. Currently, the only techniques to maintain ring structures after depolymerization involve permethylation (making methyl ethers at every free hydroxyl group), followed by reductive depolymerization—a very old technique (Gray 1990). More sensitive and rapid methods are needed to assign ring forms with confidence and to determine their locations in larger unknown glycan molecules.

5.2.1.3 Methods for determination of anomeric configuration

In larger molecules, individual sugars cyclize and generate an asymmetric center (the anomeric position). There is currently no hydrolytic/solvolytic method to generate monomers that maintains this asymmetry, and only three existing tools can even address this on intact oligosaccharide/glycoconjugate molecules. Nuclear magnetic resonance (NMR) is the most general approach, and various NMR experiments can be used to assign anomeric configuration generally, to discriminate between aldoses and ketoses, and for five- and six-membered sugar rings (Duus et al. 2000;

Bendiak et al. 2002; Coxon 2009). A current drawback is sensitivity. Glycosidases can distinguish anomeric configuration much more sensitively through release of a sugar from a larger molecule having a sensitive fluorescent tag. Current drawbacks to their general use are (1) in many new systems being studied, many of the glycosidases are either not known or not commercially available; (2) contiguous regions of the same sugar having the same anomeric configuration, or branched structures decorated with the same sugar having the same anomeric configuration, can result in the release of several monosaccharides, whereby determination of the specific linkages between them is not possible to establish; (3) in many systems the absolute specificity of glycosidases for underlying structures is not known, so they may fail to act on their substrate, even though the anomeric configuration is correct; (4) in many systems the complete specificity for all isomeric sugar variants is unknown or has not been tested; and (5) glycosidases need to be pure. Any contamination with other glycosidases, even in small amounts, can lead to erroneous conclusions.

Mass spectrometry of medium-sized oligosaccharides has the potential for general determination of anomeric configuration, through higher stages of ion isolation/dissociation, whereby smaller fragments of the larger molecules can be compared to a much more limited number of possible isomeric variants, such as, for example, the glycosyl-glycolaldehydes (Fang and Bendiak 2007). Other potential fragments include small glycoside fragments derived from derivatized molecules, such as permethylated or other peralkylated derivatives (Mendonca et al. 2003; Zhang et al. 2005). Use of these or similar small fragments to differentiate anomeric configuration generally will require new developments in mass spectral capabilities and more widely available standards for spectral comparisons.

Perhaps more importantly, any other analytical techniques that could address the general issue of confident assignment of anomeric configurations of monosaccharides in larger glycan structures would be highly desirable. While improved existing technologies that can solve this problem are desirable, completely novel approaches also are needed, particularly at very high (i.e., single-cell) sensitivities.

5.2.1.4 Methods for assigning linkages between sugars

Currently, permethylation analysis (Hakomori 1964; Lindberg and Lonngren 1978), NMR through peracetylation with ^{13}C labeled isotags (Bendiak et al. 2002), and mass spectrometry (Zaia 2004; Morelle and Michalski 2005; Park and Lebrilla 2005) can provide information about linkage sites between sugars in larger molecules, each with their own limitations. Permethylation analysis, which cleaves all glycosidic bonds after

making methyl ethers at every free hydroxyl group, cannot determine which sugars are linked to which in a larger molecule, although it can provide much important information about stereochemistries, ring forms of sugars, and substitution positions of individual monomers. NMR can provide the greatest amount of information about linkages, particularly after derivatization with ^{13}C-labeled derivatives at all hydroxyl groups or through direct ^{13}C isotopic enrichment via three-bond ^{1}H-^{13}C couplings directly across glycosidic linkages. The main limitation in many systems is sensitivity. Mass spectrometry usually yields a great deal of linkage information based on multiple cleavage methods, either before or after derivatization, and is currently far more sensitive than NMR (Dell et al. 1994; Dell and Morris 2001; Jang-Lee et al. 2006). Detailed studies of dissociation of model glycan standards is needed, particularly with specific isotope labels (such as ^{2}H, ^{13}C, and/or ^{18}O), to firmly establish and further understand dissociation pathways and mechanisms. Current limitations are ease of synthesis of many of the isotopically labeled standards. These techniques may not be directly applicable to very small (single-cell) quantities in the future, so additional methods of greater sensitivity that can address linkage positions between sugars in oligo- and polysaccharides would be important to develop.

5.2.1.5 D- versus L-enantiomers

In nature, both D- and L-enantiomers of sugars are found, and depending on the organism, either or both types can be present (Kremer and Gallo-Rodriguez 2004). Currently, the only methods that can address this involve depolymerization of larger oligosaccharides or polysaccharides to their monomers, followed by either derivatization (often with chiral reagents) and/or separations to compare the monosaccharides or their derivatives to known enantiomeric standards. It is not currently possible to prove the location of D- or L-sugars in a general way in an intact molecule without first depolymerizing the molecule to its monomers and establishing their enantiomers. New methods are needed.

5.2.1.6 Aldoses versus ketoses

Both types of these monomers are found in many living systems, and current methods to determine their structures and locations in a larger molecule involve either depolymerization or NMR, each having specific limitations. Confident assignment of these structural variants, particularly determining their specific locations in larger molecules, is needed at sensitive levels, and any novel procedures that could acquire this information would be useful and important.

It should be emphasized that, while the aforementioned tools are those currently available for primary structure determination, development of new structural techniques that could confidently determine glycan structures, particularly techniques that could provide much higher sensitivities, would be desirable. Long-term developments that might use completely novel approaches to primary structural determination would be of value.

5.2.1.7 Tools for determining three-dimensional structures of glycans and higher-order superstructures

Almost invariably, glycans and glycoconjugates reside in an aqueous environment in nature and exist in either a soluble state with individual sugars extending into the solvent or, particularly for polysaccharides, an insoluble state having unique interactions in a higher-order three-dimensional structure. For all glycan molecules, their overall structures ultimately determine their biological roles.

It is important to understand the conformations and three-dimensional solution structures of glycans. In their interactions with proteins, usually structures having from three to six monosaccharide units, referred to as "determinants," define a three-dimensional structure that interacts with a unique protein pocket or binding site. Sometimes, multivalency, in either the proteins or multivalent presentation of determinants, plays an important role in the strength and specificity of their interactions. Understanding these interactions and how the glycan or protein conformation may be modified on binding is essential for a detailed understanding of their biological roles. Similarly, with larger cell wall polysaccharides or glycosaminoglycans, their three-dimensional structures or "superstructures" are often arranged in unique ways, a function of the primary structures themselves. These three-dimensional arrangements render physical properties that are essential for organisms to function. For instance, cartilages come in several variants that differ in their glycosaminoglycan types and relative contents. Hyaline cartilage is essential in part for its compression characteristics and is found on the articulating ends of long bones. Fibrocartilage, primarily in ligaments, is very resistant to shear forces. Chitin is an abundant polysaccharide comprised of *N*-acetylglucosamine and is used to provide structural strength in many invertebrates. Cellulose is especially suited to provide structural strength, with some flexibility, for plants. Bacteria and many colonial marine algae contain glycans of a widely differing nature that play important roles in their survival. Some higher-order structures provide unique physical states ranging from gel-like to resin-like to stiff, cross-linked, and yet hydrated solids. Determin-

ing their three-dimensional structures is essential in understanding both their physical properties and their biological functions.

The two major tools for atomic-level characterization of the structures of glycans are NMR spectroscopy and crystallography. Several other types of analyses are useful in studying their physical properties, such as rheology, and other instruments that study physical/mechanical properties of the structures such as their shear or compression. In addition, several other types of spectroscopies can be used to evaluate differences in physical properties as having changed in different bulk states of polysaccharide polymers, although not usually providing information about detailed atomic-level structures of those states.

NMR spectroscopy can provide detailed three-dimensional information about oligosaccharides and polysaccharides at the atomic level. While some structural information can be obtained through solution NMR using naturally abundant magnetically susceptible nuclear isotopes (nearly 100 percent for ^{1}H and about 1 percent for ^{13}C), much more information can be obtained with isotopic enrichment (Bose-Basu et al. 2000; Martin-Pastor et al. 2003; Yu and Prestegard 2006), which provides a number of important advantages for acquiring information about solution structures (conformations) and dynamics (internal motions) of the molecules. What is currently difficult to do is easily introduce isotopes of interest at any desired location in any glycan molecule. In most cases this is a "synthetic project." Hence, methods for rapid introduction of isotopes selectively at any location in glycan molecules would be of great importance. This involves not only enrichment with ^{13}C but also in amino sugars with ^{15}N, and at any position the introduction of a deuterium atom, ^{2}H, would be of considerable value in studies of molecular dynamics. Simple isotopic enrichment of specific desired atoms in an oligosaccharide is currently the major barrier in NMR determination of larger glycan molecules. If this impediment were lifted, a great deal more information about glycan three-dimensional structures could be achieved rapidly. Advances in NMR techniques themselves, both for liquid (dissolved) and solid samples, are needed.

Crystallographic techniques are also important to provide three-dimensional structural and "superstructural" information about glycans. Many small sugar structures have been crystallized, frequently in the presence of ions (Jeffrey and Sundaralingam 1985). Crystallography has already provided valuable information about their structures that NMR cannot ascertain. For example, crystallographic studies often indicate the participation of water molecules in hydrogen-bonded networks not observed by NMR. Also, crystallography has provided detailed information about bond lengths and bond angles, which can differ slightly among six-membered sugar rings depending on their stereochemistry.

Crystallography has also provided evidence for interesting intermolecular interactions, often through hydrogen-bonded networks, for higher-order structures, including helical structures for repeating polysaccharides and interactions between helices. Without doubt, development of crystallographic methods tailored to analyses of complex glycans will be important, and future advances in crystallographic techniques will provide valuable information about three-dimensional structure.

"Superstructural" information about polysaccharides in large structural complexes and composites will be important to advance. Additional techniques, spectral or otherwise, can provide information about higher-order structures, although sometimes not to atomic-level structures. Nonetheless, all such techniques and the development of new techniques are useful for the field. For example, for plant cell wall characterization, advances in dual-axis electron tomography are needed to investigate both the three-dimensional organization of cellulose microfibrils in the cell wall and the configuration and linkages between the different cell wall components. For cellulose nanomaterials, three-dimensional characterization is needed to better understand their structures because there are several factors that may cause deviations from the idealized crystalline structures: statistical variety of the crystallite formation during biosynthesis, effects of the extraction process, a large surface area–to–volume ratio, and the coexistence of crystalline polymorphs and amorphous cellulose. Understanding how all of these variations alter the mechanical properties of cellulosics will be important in understanding how to break them down effectively as a usable energy source. Percent crystallinity and cellulose polymorphs are typically measured with x-ray diffraction, Raman spectroscopy, and solid-phase NMR spectroscopy; however, the calculated percent (crystallinity or polymorph fraction) from each technique is different, a result based on the measurement method used and the assumptions used to calculate crystallinity. Used in combination with atomistic modeling, additional insight can be gained by showing that the hydrogen-bonding configuration is different for surface chains than interior chains, for instance.

Advances in techniques that provide information about higher-order organization of a large number of polysaccharides will be of importance in a wide variety of industrial, medical, materials science, and energy-sector applications. A great deal is currently unknown and remains to be learned about glycan structures and their molecular interactions with proteins and other molecules. Therefore, advances in these techniques toward higher sensitivities and faster and more confident assessments of their three-dimensional structures are needed.

5.2.2 Analysis of Glycoconjugates

Glycans are often conjugated to other molecules, and the most abundant and widespread of these are glycoproteins and glycolipids. Glycoproteins can be broken down readily to glycopeptides using peptidases. This increases the complexity of analyses because a single glycan can now be associated with a number of peptides, or a number of glycan structures (including isomeric species) can be associated with each peptide. This further complicates the separation problem because more individual molecular species are present and each is present in less abundance. To date, analysis has rarely been performed for a single glycoprotein for which all glycan structures associated at different glycan-peptide linkage positions have been determined unambiguously. Having detailed structural information on glycoproteins is an important part of studies to address their functions in vivo. Similarly, glycolipids have glycan structures linked to lipid moieties at their reducing end glycosides, and these lipid groups can be variable in structure. Both the nature of the glycan structures and the nature of the variations in these lipid groups are likely to be essential for understanding glycolipid functions. To routinely achieve a level of rapid and sophisticated separations and structural analysis with many glycoproteins and glycolipids will require improved technologies.

In addition to primary structural analysis, three-dimensional structural analysis of molecules such as glycoproteins is used. X-ray crystallography is one method that has been widely used to determine the structures of proteins. Some glycoproteins, depending on the nature of crystal contacts, have been crystallized, with the caveat that their glycan structures (or, sometimes, lack of observable structures) might result from their packing into unusual (or multiple) conformers compatible with crystal formation. Thus, electron density that is either observed or not observed for the glycan portions of glycoproteins can be prone to interpretive errors as compared to their true solution structures.

NMR spectroscopy has the potential to determine the solution-state glycan conformations of glycoproteins under physiological conditions. However, this requires isotopic labeling (^{13}C, ^{15}N) of the sugar component and currently carries the caveat that glycoproteins are less than ~50,000 molecular weight. As mentioned earlier with three-dimensional structures of the oligosaccharides themselves, similar techniques may be applied to isotopically labeled structures attached to proteins, and isotopic labeling of the protein portion is also desirable, either together, or independently, from labeling the glycan moiety. Currently, there are two major hurdles to performing these studies: (1) isotopic labeling of the glycan structures is far from straightforward and (2) isolating a glycoprotein structure having single glycan moieties (sometimes at multiple sites) is currently difficult to achieve. In the long term these developments will be essential. The

current approach has largely been to ignore the glycans either through expression of the protein in prokaryotic systems or in some cases through site-specific deletion of the glycan–amino acid linkage site. However, this approach will be met in the future with the inevitable question of whether this protein structure is correct or simply convenient to do, in which case all glycosylated proteins or proteins having other posttranslational modifications may need to be solved again to yield the correct structures. Because even a simple phosphorylation can have dramatic effects on a protein's activity or function, glycosylation would be expected to have equal or even more dramatic effects.

How might specific glycan structures be generated at specific protein sites? This is a current challenge to the scientific community but will be important in the future in understanding the specific roles played by unique glycans linked at unique protein sites.

In addition to the many high-resolution tools and techniques described above, valuable information about glycan structure can be obtained through the use of molecules that bind glycans with known specificities. These include glycan-binding proteins, lectins, and antibodies. The structural information such methods provide is of lower resolution than techniques such as NMR, but one advantage of these molecules is that they permit interrogation of intact glycoconjugates and even of intact cells and tissues. Such approaches are complementary to those involving glycan release and chemical and physical analysis. Having additional tools such as glycan-binding molecules with fully characterized binding specificities can also be a powerful approach to better understanding glycan structure function.

5.2.3 Analysis of Glycan-Protein Interactions

5.2.3.1 Glycan-protein interactions—the search for endogenous cognate ligands

For many glycoproteins the glycans contain crucial biological information, and that information is decoded, often at the cell surface, by glycan binding proteins. A very important goal is to develop methods that can assess the glycan binding specificities of various proteins and that can rapidly isolate the glycan component(s) having the highest binding affinities.

Frequently, the binding pockets in proteins have a specificity for glycan structures from about four to six sugars (their determinant), but multivalency of binding can mean that more than one of these determinants must be displayed on a branched glycan molecule for highest binding affinity. There are two general approaches: (1) Separate a natural series

of oligosaccharides isolated from some source into a number of fractions, many of which may not be pure. Immobilize the fractions (i.e., at their reducing ends) in an array. Then test for binding of a protein to all of the immobilized fractions in the array. Then, for positive fractions, further fractionate the sample until individual natural glycan molecules can be identified that interact with a specific glycan-binding protein. A variant of approaches also can be used via immobilization of the glycan-binding protein on a column, with passage of glycans over the column, followed by elution with a hapten (usually simple sugars like methyl glycosides), followed by structural determination of the glycans that preferentially bind to the immobilized protein. (2) A second approach is to synthesize a number of glycan substructural determinants and immobilize the synthetic molecules on a binding array surface. This approach has the advantage of knowing the precise structure at each array position. However, it has the disadvantage that a natural structure might have a much higher affinity for the protein, either in having a previously unknown structure with higher affinity than a synthetic array or a multivalency of determinants on a single glycan molecule could result in higher affinity. Both approaches are valid, though, for making the biological connection between specific glycans and the proteins that bind them.

5.2.3.2 Detailed studies of atomic-level glycan-protein interactions

Current techniques to evaluate atomic-level protein-ligand interactions are primarily crystallographic or involve NMR. But difficulties can ensue using either technique. Frequently, the most time-consuming aspects are the availability and/or preparation of the glycan structures involved in the interactions, as well as variants of those structures to assess binding efficacy and multivalency. In the case of NMR, preparation of specific isotopically labeled glycans is frequently rate limiting. Either technique, depending on the specific protein and glycan structure interacting, may not yield information: This may be due to an inability to obtain crystal structures or to ones that fail to diffract to yield atomic-level assignment or, in the case of NMR, the size or overall flexibility of the protein-glycan complex. This information is vitally important for a detailed understanding of protein-glycan interactions. Further development of these or any other techniques that can provide information about such binding interactions will be of value.

5.2.4 Analysis of the Roles of Some Glycans in Metabolic Pathways Related to Energy Metabolism

A vital role of glycans centers on the involvement of glucose in the glycolytic pathway, in utilization of its carbons in the Kreb's citric acid cycle, and in subsequent production of adenosine triphosphate, the currency of cellular energy, through oxidative phosphorylation. Likewise, storage of glucose as a retrievable polymeric energy source as glycogen or (in plants) starch forms the basis of the energy humans use to survive on a daily basis.

Each tissue has energy needs that differ, and in many organisms alternate sugar sources to glucose are possible to use through metabolic conversion. Much is still unknown about the use of alternative energy sources by many microorganisms, yet understanding these metabolic pathways may provide important keys to biomass conversion. A great many potential alternatives to the use of cellulose as an energy source may be feasible through future studies of additional bacterial or algal metabolic interconversion systems, and their efficiencies, rates of carbon dioxide fixation, and potential for large-scale conversion to biofuels need to be carefully examined. Each system will require analytical capabilities for studies of accumulated glycan structures and/or metabolic intermediates en route to potential molecules usable as biofuels.

In humans a number of metabolic diseases form the basis of a relatively recent area of science—metabolomics (Kaddurah-Daouk et al. 2008). Development of analytical tools to understand the balance of metabolic intermediates in healthy individuals, deficiencies in patients, and whole-body imaging of some metabolites (see below) are important needs and challenges related to our use of sugars as energy sources. A number of glycosylation deficiency disorders (>65) have been described elsewhere (Freeze and Sharma 2010) that also affect synthesis and degradation of glycoprotein, glycolipid, and glycosaminoglycan glycan structures, and techniques to diagnose and further detail metabolic intermediates (partial glycan structures) in these disorders form a starting point for a deeper understanding of the roles of glycans in a number of disease states.

5.2.5 Analysis Techniques That Relate Glycan Structures and Their Synthetic Enzymes

Ultimately, glycan structures in organisms are related to expression of the enzymes that synthesize and degrade them—glycosyltransferases, glycosidases, and other proteins such as nucleotide-sugar transporters, sulfotransferases, and a number of other proteins involved in their compartmentalization in cells. These are encoded by genes that are differentially expressed in many cells, and a detailed understanding of this

expression as well as mechanisms that control their breakdown will be valuable in understanding the biological functions of the glycans. Transgenic knockouts are one important tool to investigate the functions of these genes and have already proven valuable in determining the requirements of a number of glycosylation enzymes in developmental processes and the roles they play in some human diseases (Hennet and Ellies 1999; Haltiwanger and Lowe 2004).

While these genetic ablations have determined the crucial roles of glycans through dramatic phenotypic abnormalities, development of tools for a more detailed understanding of their functions in vivo is important. Tools for organ-specific, cell-lineage specific, and cell–type–specific ablation of glycosyltransferases and glycosidases will be needed in order to examine in more detail the specific effects they have in the functioning of specific differentiated cells. Furthermore, the informational role of glycans in intercellular recognition events will require an understanding of (1) the location of expression of specific glycan structures to the level of individual cell types and (2) the location of expression of the proteins that bind to them (endogenous mammalian lectins). Imaging techniques (see below) will be important for establishing localization of both binding partners between cells where suspected interactions occur. Detailed in vivo analysis of these interactions will require a number of technical developments, in both cell-specific knockout techniques and glycan imaging techniques.

5.2.6 Analysis of Locations of Specific Glycan Structures in Organisms Through Various Imaging Techniques

In studies of RNA expression, in situ hybridization has proven invaluable in localizing specific RNA transcripts in cells and in identifying their expression in tissue patterns in organisms. Similarly, in protein expression, detailed localization of proteins has been made possible through carefully studied monoclonal antibodies having low cross reactivity and high specificity for individual protein molecules or through expression of chimeric proteins tagged with fluorescent proteins. These tools alone have been extremely important in characterizing cellular and subcellular locations of these molecules and in providing essential insights into their functions in higher organisms.

Monoclonal antibodies unique to specific glycan determinants are well known, but arrays of many more that are specific to unique glycan structures are needed as biological probes. A large group of monoclonal antibodies specific for plant glycan determinants are being used for localization and glycome profiling (Avci et al. 2011). Potentially, a large array may be developed through syntheses of a number of glycan determinants

and selection of good monoclonals having high specificity and binding affinity to unique glycan structures, including those of glycoproteins, glycosaminoglycans, and glycolipids, in addition to glycan substructures unique to pathogenic organisms that might be valuable for their diagnosis or therapy. Antiglycan antibodies can include both IgM and IgG classes, although there is a general tendency for IgM to bind glycans over IgG, a factor to be considered in array development.

More recently, RNA and DNA molecules (aptamers) have been shown to be selectable for unique binding properties to many molecules, and they also have the potential to provide important selective binding probes for many unique glycan structures. It seems feasible that aptamers might provide another battery of arrayed probes useful for imaging glycans that have been explored little as yet, but these and any other methods for uniquely identifying single glycan species at specific cellular locations will be of great value.

Recent developments in other imaging techniques have occurred as well. These include mass spectral imaging of peptides and lipids in tissue slices and could conceivably be extended to a number of glycosylated peptides and glycolipids assessed through localization of selected precursor ion masses and unique glycan neutral losses. Mass spectrometry imaging techniques may enable analysis of a broader range of glycan structures simultaneously in tissue sections, and further developments in mass spectral imaging techniques as applied to glycopeptides or glycolipids are needed. While this comes with the caveat that more than one glycan isomer could contribute to a select glycopeptide or glycolipid precursor ion mass, it would still be of considerable value because many precursor masses could be independently studied, some attributable to unique peptide-glycan or lipid-glycan combinations.

Whole-body imaging techniques also might hold enormous promise in nondestructive molecular imaging in living animals, including humans. Magnetic resonance imaging with pulse sequences uniquely tailored to imaging specific metabolites or unique individual molecules such as ATP and specific oligosaccharides could be feasible in the future, studied initially through isotopically enriched molecular species in animal studies. The use of 2-^{18}F-2-deoxyglucose is already being routinely used in positron emission tomography to study glucose uptake in various tissues under different metabolic states and for investigation of different neurological conditions (Zijlstra et al. 2006). These and other whole-body imaging techniques need to be developed initially in animal studies, followed by potential applications in human diagnostics. There is great potential for imaging techniques in nondestructively determining the locations of glycan molecules in whole-body imaging protocols. Similar approaches can be used for in vitro analyses.

In summary, any novel imaging techniques that could address histological specimens, tissue slices, or whole-body living organisms and provide information about the locations and/or superstructural organization of specific glycan molecules will be of value in the future.

5.2.7 Key Messages on Glycan Analysis

Although analysis of glycans is challenging because of the number and diversity of glycan structures, a wide range of helpful tools exist. Perhaps not surprisingly, these techniques have different advantages and limitations, and no single analytical technique will be able to provide all desired information. The most suitable technique will depend on the nature of the question at hand and the essential information required. The different types of techniques also provide information at different levels of analytical specificity; as discussed further in Section 5.6, specification of the level of analysis as part of bioinformatics and other efforts is necessary. Many techniques are currently available for polymer and nanoparticle synthesis, including rheology and light-scattering techniques, and are currently used heavily in many fields, including materials engineering. There is a large array of possible data that can be obtained with current analysis tools, depending on the necessary level of analysis. However, further developments in such areas as separation techniques, structural analysis, and the use of glycan-binding proteins and antibodies of known binding specificities may all expand the analytical toolkit to advance glycoscience knowledge.

5.3 COMPUTATIONAL MODELING

Computational modeling in biology has been an active area of research for many years, particularly in terms of protein structure modeling, molecular dynamics simulations, protein docking, and virtual drug screening. However, for any of these methods to be deemed useful, modeling must be able to accurately predict structures that are experimentally verifiable. Developing such models for glycan structures is particularly challenging but is also an important component in the glycoscience toolkit.

5.3.1 Computational Modeling of Oligo- and Polysaccharides

SWEET-II is one of the first Web-based tools for computational modeling of glycan structures. It is provided as a part of the glycosciences.de website, where a number of tools for analyzing glycans in three-dimensional space are provided. Molecular dynamics simulations of oligosaccharides in explicit solvent have also been performed that provide an analysis of

the trajectories of a complex oligosaccharide or glycoprotein structures, and quantum mechanics, or ab initio, and quantum electrodynamics methods are also used for modeling of glycan conformation. However, these methods are computationally demanding and cannot be used routinely to study or predict the three-dimensional structure of complex glycans. The development of computationally efficient and accurate simulations of complex glycans is therefore a major challenge that needs to be addressed.

A number of glycan force fields have been developed, including GLYCAM, AMBER, CHARMM, OPLS-AA, GROMOS, MM4, and SPASIBA (Hancock et al. 2006; Fadda and Woods 2010). These force fields vary according to which atomic interactions are modeled, the mathematical form of those interactions, and the fit of the parameters to the resulting mathematical expressions. In addition, the details of a given force field may vary significantly from one version or release to the next, and the force fields are often subject to simplifications or approximations to improve computational efficiency. As a result, each force field provides a unique parameter set resulting from the various techniques invoked to improve performance or to ease implementation, transferability, or generality to expand beyond glycans. Different protocols also have different levels of compatibility with other biomolecules and solvent models, indicating the difficulty in reenacting real-world situations computationally.

For any modeling software or algorithm to be useful, it must be able to accurately predict known structures. The entropic components of solvent order and disorder may be extremely difficult to estimate now, but this is a necessary challenge to be resolved in glycan modeling. Moreover, another challenge lies in modeling even small molecules, such as simple methyl glycosides, that can serve as gold standards for modeling measurable spectral parameters. Although computational performance will likely improve with technological progress, new algorithms and software for such modeling challenges are greatly needed.

5.3.2 Protein-Glycan Interactions

One of the major challenges in molecular modeling is the development of efficient and accurate methods to estimate the binding affinity between proteins and glycans. Many protein-docking methods have been applied to study protein-glycan interactions. However, there are factors involved in these interactions, such as bridging water molecules and CH-pi interactions, that are often ignored in these simulations for a variety of reasons. Further research in this area is of great interest and would aid in our understanding of glycan function (Frank and Schloissnig 2010). Another difficulty with predicting these interactions involves the fact that binding affinity may not always necessarily increase with larger oligosac-

charide size. Moreover, glycan-protein interactions are intrinsically more dynamic than other protein-ligand interactions because their affinity for one another arises from several relatively weak interactions. As a result, accurate prediction of binding affinity is a challenging task that remains to be solved.

5.3.3 Atomistic Modeling of Crystalline Cellulose

Since the 1980s, atomic-scale modeling of cellulose has been used to complement experimental measurements of individual cellulose crystals. Such studies have advanced our understanding of cellulose structure, energetics, mechanical characteristics, and interfacial properties involving liquids, a given chemical species, cellulose chains or surfaces, and enzymes. Atomistic modeling also provides a fundamental understanding of the atomic-scale origins of these characteristics. However, the predictions made by these models are limited, because they are highly dependent on the accuracy of the force fields that describe atom-atom interactions and on the accuracy of the structural information at the atomic level of cellulose nanomaterials. To improve understanding of the construction and deconstruction of cellulose for biofuels, and to better tailor the properties of cellulose nanomaterials extracted from biomass, three key challenges need to be addressed:

- development of more accurate force fields for cellulose with experimental validation of input parameters;
- improved structural-property characterization and linkages; and
- improved interaction simulations involving glycans with liquids, a given chemical species, cellulose chains or surfaces, and enzymes.

5.3.3.1 Force fields for cellulose

For cellulose a force field must accurately describe the stretching, bending, and torsion of covalent bonds, electrostatic interactions, van der Waals forces, and hydrogen bonding. The force fields most commonly used for cellulose modeling are MM2/MM3, GROMOS, CHARMM, CVFF/PCFF/COMPASS, AMBER, Dreiding, and COSMOS. Among the numerous differences between force fields, one of the most important for cellulose is hydrogen bonding, because it is critical for simulating structural and mechanical properties, as well as the interaction of cellulose nanomaterials with the environment. Both implicit and explicit hydrogen bond models have been developed using a variety of combinations of parameters to describe force fields. However, the difficulty in utilizing

this information is the lack of a consistent definition of a hydrogen bond, which has typically been identified by the distance between hydrogen and acceptor and sometimes by the angle between donor, hydrogen, and acceptor. Although the hydrogen bond model selected can have a significant influence on simulation predictions, the advantages of using one approach over another have not yet been conclusively determined.

5.3.3.2 Predicting cellulose nanomaterial properties

Molecular modeling has been used to predict a variety of cellulose material properties, including elastic modulus, thermal expansion, and Poisson's ratio. The most frequently predicted properties are elastic properties because they can be calculated using molecular models relatively easily and are experimentally measurable. Molecular simulation has shown that numerical removal of hydrogen bonds can cause predicted elastic properties to decrease on the order of 50 to 60 percent. In addition, molecular modeling has shown that cooperative hydrogen bonding plays such a critical role in the behavior of cellulose that omitting interchain hydrogen bonding, as is necessarily the case for a single cellulose chain, will affect intrachain hydrogen bonding.

5.3.3.3 Predicting interface properties of cellulose

Models have been used to investigate the interaction of cellulose and other glycans with other materials, primarily liquid solvents and other polymeric materials. Most solvent studies are focused on water, although some have modeled the behavior of cellulose in benzene and cyclohexane. Some studies have used the radial distribution function to characterize water structure relative to specific surfaces, referred to as solvent-accessible surfaces. Instead of surrounding cellulose with water, some studies have introduced a water droplet into the model to investigate solvent-accessible surfaces in terms of the contact angle. Modeling studies have even shown that water induces changes in the crystal structure itself in ways that include affecting the twist of the cellulose strand. New modeling techniques and structure-property linkages will provide additional insight into how to dissemble cellulose materials and on the interaction of cellulose nanoparticles and their environment, which will be useful for suspension-based composite-processing routes.

To date, the primary focus of modeling studies of the interaction of cellulose with other polymeric materials has been on characterizing molecular interactions of the various surfaces of crystalline cellulose with the desired polymer. The noncovalent interactions between cellulose and an adjacent molecule are characterized in terms of interaction energies,

density profiles, and orientation changes. The interaction between a cellulose crystal and a single cellulose chain has also been characterized by numerically pulling the chain away from the surface and calculating the pull-off force as a function of the initial chain orientation. In addition, the potential for chemical grafting as a means of strengthening the interaction between cellulose and a polymer matrix has been investigated. Further advances in this area of modeling, combined with the development of new structure-property linkages, will provide additional insight that will aid in the development of cellulose nanomaterial composites, particularly in how to tailor the interface chemistries to provide the desired properties in composite structures.

5.3.4 Key Messages on Computational Analysis of Glycans

Computational modeling of glycans and the interactions of glycans with each other and with other molecules is often a complementary tool to other analytical techniques for understanding glycan structures and properties. Although significant advances have been made in the development of computer models and force fields for glycans, accurate predictions remain challenging. This is due to such factors as the flexibility of glycan molecules, the fact that many glycan interactions involve multivalent and weak molecular interactions, the complex role of water in determining glycan three-dimensional structures, and the need for models to accommodate electrical charges on certain subsets of polysaccharides.

5.4 GLYCOENZYMES

The glycome of an organism is defined by glycoenzymes, encoded by the genome, that synthesize and degrade glycans. The high specificity and efficiency of these enzymes can be exploited as tools to produce or degrade glycoconjugates of interest and to manipulate glycans in complex biological systems to study their functions (Kiessling and Splain 2010). These enzymes can also serve as targets for inhibition to alter glycan structures in situ, with potential for applications that benefit human health.

5.4.1 Classes of Glycoenzymes

Numerous enzymes are needed to synthesize and degrade the glycans of animals, plants, and microorganisms. Coordinated expression of these enzymes is required for normal production and degradation of glycans in any cell, and deficiencies or mutations in any can result in abnormal biology and disease. For the biosynthesis of glycans, glycosyltransferases

are responsible for building glycan chains of defined structure needed by a cell. The resulting glycans can be modified by other classes of enzymes, such as sulfotransferases, *O*-acetyltransferases, and epimerases, resulting in additional structural diversity that influences the functions of the glycans. Degradation of glycans is carried out by glycosidases, which all organisms rely on for digestion of glycans as nutrient sources. Glycosidases are also required for the normal turnover of glycans produced by cells, and many human deficiencies in these enzymes result in the buildup of products. Genetic deficiencies in many of these enzymes are the basis of many severe human disorders recognized as childhood disorders of glycosylation.

Despite the importance of glycoenzymes, little is known about how they carry out the biosynthesis and degradation of glycans of animals, plants, and microorganisms. Of high current interest is understanding how glycosyltransferases are able to produce glycans that contain information needed to mediate diverse biological processes.

5.4.1.1 Glycosyltransferases

Glycosyltransferases carry out the nontemplate-driven synthesis of glycans in all organisms and in the process transfer information encoded by the genome to the glycan structures that comprise the glycome of that organism. Most glycosyltransferases act by transferring a single sugar to a specific hydroxyl group of another saccharide in a growing glycan chain. In this case the fidelity of glycan structures is determined by the high specificity of glycosyltransferases for their substrates, with the product of one enzyme being recognized as the acceptor substrate of the next enzyme, allowing the assembly of glycans of defined structure. It is estimated that the human genome encodes approximately 250 glycosyltransferases (Narimatsu 2006; Henrissat et al. 2009). Some glycosyltransferases, such as those that synthesize core structures, are expressed in nearly every cell, whereas the subset of glycosyltranferases that elaborate terminal sequences on glycoprotein and glycolipids glycans are differentially expressed (Lowe and Marth 2003; Comelli et al. 2006).

Although most glycosyltransferases catalyze the formation of one type of linkage, some promote the assembly of polysaccharides. Mammalian polysaccharides include medically important substances such as heparin or hyaluronan, which are composed of multiple linkages. Plant polysaccharides, some of which are structurally very complex, are the products of plant glycosyl transferases, and these polysaccharides form the majority of plant biomass. Bacterial polysaccharides are often essential to bacterial viability, and they also serve as vaccine candidates. For polymerizing glycosyltransferases the factors that determine polysac-

charide length, sequence, and fidelity are much less well known. What is apparent is that each cell type produces a characteristic set of glycan structures encoding information that contributes to the biology of that cell (Lowe and Marth 2003; Comelli et al. 2006). While the precise organization of the glycosylation machinery differs somewhat for plants, bacteria, and other microorganisms, the central role of glycosyltransferases in mediating nontemplate-driven biosynthesis of defined glycan structures is the same.

5.4.2 Applications of Glycosyltransferases and Other Glycoenzymes

As a result of their exquisite specificity, glycosyltransferases and glycosidases are widely recognized as important tools for chemoenzymatic synthesis of glycans and as enzymatic probes of glycan structure. In this context they can be likened to other enzyme classes that have enabled the analysis and synthesis of other biopolymers, such as polymerases and endorestriction nucleases for DNA, and proteases for proteins. However, major barriers to the routine use of these enzymes as tools include lack of widespread availability, lack of well-characterized enzymes of desired specificity, lack of access to the appropriate nucleotide sugars, and difficulty in producing stable enzymes in sufficient quantity. As a result, only a few laboratories have the capacity to exploit the power of glycosyltransferases as synthetic tools despite widespread interest in using them. Having a diverse enzymatic toolbox that is widely accessible to the glycoscience community would dramatically accelerate their use and solidify their roles in chemoenzymatic synthesis of glycans and as probes for analysis of glycan structure.

To exploit and understand glycosyltransferases, insights into the generation of sugar-nucleotide donors also are needed. Advances on this front are enhancing access to a wide range of glycosyl donors. Specifically, the engineering of glycosyltransferases such that they run in reverse to catalyze nucleotide-sugar formation is enabling the production of natural and nonnatural building blocks (Gantt et al. 2011a). Similar methods to rapidly generate lipid-linked sugar donors, which are used by many microbial enzymes, could lead to advances in our understanding of host-pathogen interactions and fuel efforts to devise novel antimicrobial agents. Moreover, inhibitors that block the production of unique nucleotide sugars in pathogens could lead to new classes of antimicrobial agents. For example, compounds of this type could lead to new strategies to combat infectious diseases, including tuberculosis (Dykhuizen et al. 2008).

In the same way, inhibitors of glycosyltransferases are needed to investigate their roles and the glycans they produce in biology. Currently, there are only a few glycosyltransferases for which enzyme inhibitors

have been identified, and there are inhibitors of only one or two that have in vivo activity. As a result, the field has adopted the much more difficult strategy of using genetic knockouts in mice or Arabidopsis to gain information on the roles of individual enzymes and the structures they produce in vivo. The phenotypes of glycosyltransferase knockout mice strongly suggest that inhibitors of selected enzymes would have therapeutic potential, while Arabidopsis knockouts are helping to identify targets for overcoming recalcitrance in plant biomass. The availability of inhibitors for key glycosyltransferases would be of enormous benefit in elucidating the functions of their glycan products and would validate these enzymes as targets for therapeutic intervention in human disease. It is envisioned that inhibitors of classes of enzymes as well as inhibitors of single specific enzymes will be useful tools to understand the biology of glycans and to provide altered glycans for discovery or therapeutic use.

To address these critical needs, a systematic effort is needed to identify enzymes for glycan assembly and degradation that exhibit the range of specificities and physical properties, such as stability and turnover rate, needed to build an enzymatic toolbox. A parallel effort is needed to identify inhibitors of glycan assembly. The sections below expand on these and other needs for glycosyltransferases and glycan-processing enzymes.

5.4.2.1 Chemo-enzymatic synthesis of glycans

There are two general strategies for chemoenzymatic synthesis of glycans: one is to use engineered glycosidases, and the other is to use glycosyltransferases (see also Section 5.1.2.1, which discusses enzymatic synthesis of glycans). The latter is the most widely used method today. The specificity of glycosyltransferases makes them unique synthetic catalysts for production of glycans of defined structure that will serve as reagents to investigate the biological roles of glycans and glycan-binding proteins (Vasiliu et al. 2006; Boltje et al. 2009; Palcic 2011; Xu et al. 2011). As synthetic tools, glycosyltransferases complement chemical synthesis technologies, and the combination of chemical and enzymatic synthesis can produce most natural glycan structures. In general, because of the specificity and regioselectivity of glycosyltransferases, addition of a single monosaccharide by a glycosyltransferase replaces 10 to 12 chemical steps to achieve the same end, which vastly simplifies the synthesis of complex glycans.

Although they exhibit high specificity for natural glycans, many glycosyltransferases exhibit high promiscuity for donor and acceptor substrates with unnatural substituents at positions that do not interfere with the enzymatic activity (Yu et al. 2006; Blixt et al. 2008; Palcic 2011). This allows them to be used for synthesis of glycans with substituents that

improve biological activity or contain functional groups for subsequent chemical tagging. This biosynthetic promiscuity can be exploited to introduce functionality into the glycans of living cells by feeding them with chemically modified monosaccharides that are used by the cell's biosynthetic machinery in place of the natural sugar.

The major limitation to widespread use of enzymatic synthesis is the paucity of glycosyltransferases available to the research and industrial communities. In part the shortage of suitable enzymes results from the difficulty of producing mammalian glycosyltransferases in quantities sufficient to meet demand. The most useful and robust synthetic enzymes are bacterial glycosyltransferases that synthesize mammalian-type structures (Gilbert et al. 1998; Yu et al. 2006; Sauerzapfe et al. 2009; Palcic 2011). Identification of additional bacterial enzymes that fill the gaps in the enzymatic armamentarium would be extraordinarily useful. Genomic and metagenomic sequencing of all organisms, to date, has revealed more than 50,000 putative glycosyltransferases, of which only several hundred have been tested and confirmed to be glycosyltransferases (Narimatsu 2006; Henrissat et al. 2009; Palcic 2011). This genetic resource might be tapped to identify enzymes that can be produced to meet the need for these tools.

An alternative to the use of glycosyltransferases for synthesis of glycoconjugates is the application of engineered glycosidases, termed "glycosynthases" (Hancock et al. 2006; Wang and Lomino 2012). Recently, enzymes of this class have been shown to be especially effective for the en block transfer of oligosaccharides onto proteins bearing a single glycan moiety. These methods, as well as other advances in protein engineering such as expressed protein ligation, are being used to generate new types of defined glycoproteins.

Although most enzymatic synthesis has been done on a research scale, manufacturing scale is limited only by the lack of current technologies to produce enzymes and substrates cheaply. Engineering bacteria to produce small mammalian oligosaccharides has shown some promise as one approach to synthesis at large scale (Antoine et al. 2003, 2005; Drouillard et al. 2006), but the availability of glycosyltransferases produced by bacteria, yeast, or fungi also would address the scale problem.

5.4.2.2 Glycosidases and glycosyltransferases in glycan structure determination

Although methods for rapid profiling of glycans have advanced the glycoscience field, robust methods for the complete description of glycan structures are lacking. Methods in wide use include mass spectrometry approaches, such as nano-LC/MS and MS/MS, NMR, and x-ray crystallography, and conventional biochemical methods involving radiolabeling,

glycosidase digestion, and chromatographic fractionation (Marino et al. 2010). None of these methods provides complete structure information. The lack of a single high-throughput method is made more challenging by the small amounts of glycans that can be obtained from analysis from biological sources. Indeed, as Marino et al. have noted, "No universal method for the rapid and reliable identification of glycan structure is currently available; hence, research goals must dictate the best method or combination of methods" (2010, p. 713).

Glycosidases have long been used to aid in the sequencing of glycans, providing key information for complete structure determination. However, their use is not routine and has not been exploited for high-throughput structure determination. In part this is due to routine availability and gaps in the enzymatic toolbox to assist in structure determination of diverse glycan structures. Because glycosidases are found in every organism and are critically important for degrading glycans for nutrients, there is an enormous genetic resource for enzymes that will fill the toolbox. This is exemplified by the existing CAZY database, which has identified glycosidase-related genes in more than 2,000 eukaryotic and prokaryotic species (www.cazy.org).

In principle, the high-substrate specificity of glycosyltransferases also offers the potential for a systematic approach to sequencing glycans. By analogy with the Sanger methods of DNA sequencing, where template-directed, chain elongation/termination reveals DNA sequence, a repertoire of glycosyltransferases could be applied to a glycan moiety of unknown structure, together with nucleotide sugars (radiolabeled or labeled with other report groups) in vitro. As applied to glycan sequencing, the "template" is the structure of the acceptor glycan whose structure is sought, but this template is three-dimensional instead of the two dimensions characteristic of a DNA sequence. Defining the structure of the more information-rich three-dimensional glycan template could be enabled with a repertoire of glycosyltransferases capable of "reading" the template. Some glycosyltransferases have strict acceptor specificity for disaccharide or trisaccharide sequences enabling them to be used to advantage over glycosidases to gain linkage information, a concept that is also being exploited to detect glycan epitopes of defined sequences on cell surfaces. Mutant glycosyltransferases engineered for informative acceptor and/or nucleotide sugar substrate specificity could make this approach more penetrating and more widely useful (Palcic 2011).

Such an approach could also be useful in expanding the utility of NMR-based glycan structural determinations, by analogy to approaches involving segmental isotope labeling of proteins that are expanding the size limit of NMR spectroscopy to larger proteins (Skrisovska et al. 2010). This technique could be applied to large glycan structures either as iso-

lated moieties or as components of a glycoprotein or glycolipid, using one or more glycosyltransferases with defined substrate specificity and sugar nucleotide substrates whose sugar moieties are labeled with stable isotopes (Macnaughtan et al. 2008; Skrisovska et al. 2010). Glycosyltransferases could be used to assess the presence of a determinant as defined by its substrate specificity, although it is important to note that they cannot be used to solve entire structures or novel structures on their own. However, this is an extremely promising technology that should be further developed to help address the current shortcomings in NMR technologies for determining tertiary structure of glycans, especially those attached to their native proteins.

5.4.2.3 Glycosyltransferases in glycan engineering of cells

Because they are major components of the cell surface, glycans represent attractive targets for imaging physiology and pathophysiology, both in vitro and in vivo. As mentioned above, many glycosyltransferases are promiscuous and can accept unnatural substituents. At least one group has used this principle to introduce "bio-orthogonal chemical reporters" into monosaccharides that are taken up by cells and incorporated into cell surface glycans. These reporters enable the detection and imaging of glycan structures of living cells in model organisms using bio-orthogonal chemistry to attach fluorescent label or other biological tag that make the glycans "visible" (Laughlin and Bertozzi 2009; Sletten and Bertozzi 2011). This approach has been used to label cells for in vivo imaging in mice (Chang et al. 2010) and recently to image the sialome in zebrafish (Dehnert et al. 2012). The same principle can be used to introduce modified sugars directly on to the surface of cells using glycosyltransferases (Ramya et al. 2010; Zheng et al. 2011). This approach also provides information on the underlying glycans on the cell, because the enzyme has a strict specificity for the acceptor sequence it uses to form the product (Khidekel et al. 2004; Boeggeman et al. 2007; Ramya et al. 2010; Zheng et al. 2011). With a large toolbox of glycosyltransferases with well-characterized specificities, this approach can be used to gain much structural information about the glycans on the cell surface.

Glycoengineering approaches are also being used to influence cellular trafficking. For example, using a platform called "glycosyltransferase-programmed stereosubstitution," scientists have modified existing cellular glycans to create the selectin ligand HCELL (hematopoietic cell E-/L-selectin ligand), which is involved in the attachment of circulating stem cells and white blood cells to endothelial cells. This technique has potential applications to the development of cell-based therapies (Sackstein 2009).

5.4.2.4 Glycosyltransferase inhibitors

Because of the central importance of glycosyltransferases to the synthesis of glycan structures, inhibitors of key enzymes would be of enormous benefit to elucidate the functions of glycans in cell communication and the roles of specific enzymes in the biosynthesis of glycans. Genetic ablation of specific glycosyltransferases in mice has already revealed important biological roles for glycans synthesized by the missing enzyme (Lowe and Marth 2003; Satoh et al. 2005; Ohtsubo et al. 2011). Many phenotypes from these mice have validated individual glycosyltransferases as targets for the development of inhibitors that would provide a therapeutic benefit. Small-molecule inhibitors to such enzymes would be invaluable to the research community as probes to uncover the biological roles of glycans and to assess their therapeutic utility. Specific inhibitors could also be used in place of or in combination with glycosyltransferase knockout mice to reveal additional novel phenotypes that provide information about the functions of glycan ligands and glycan-binding proteins. Despite the obvious need, few glycosyltransferase inhibitors capable of blocking glycosylation in vivo have been identified to date (Lachmann 2003; Brown et al. 2009). Several recent reports describe approaches for high-throughput screening of glycosyltransferase inhibitors that demonstrate the feasibility of screening for inhibitors of these enzymes (Helm et al. 2003; Gross et al. 2005; Rillahan et al. 2011). A systematic effort to screen for inhibitors of a panel of key glycosyltransferases is sure to open a path to the development of inhibitors that will benefit the research community and assess the potential of glycosyltransferase inhibitors as drug development targets.

5.4.3 Key Messages on Glycoenzymes

Enzymes have a range of uses as tools to study glycoscience, including in enzymatic synthesis of glycans, as biochemical probes, and in structural determination. Similarly, inhibitors of enzymes such as glycosyltransferases can be used as important tools in trying to better understand glycan biology and function. Despite their utility as part of the glycoscience toolkit, only limited numbers of glycan-active enzymes from both bacteria and mammalian species are available, and few three-dimensional enzyme structures, particularly from mammals, are known.

5.5 SYSTEMS GLYCOBIOLOGY

As a recent National Research Council report described:

> The field of systems biology seeks to integrate . . . multiple levels of biological knowledge into descriptive, and ultimately predictive, mathe-

> matical models, combining experimental knowledge with computational tools in order to study the interactions between the components that make up a particular biological system. As a result, a primary goal of systems biology is to understand how the system being studied functions, what its properties are that arise from the interactions of its individual components (also referred to as emergent properties), and the design principles on which it operates (NRC, 2011, p. 27).

Similarly, systems glycobiology is an approach that integrates biological and chemical information about glycans with mathematical modeling and bioinformatics-enabled data analysis in an effort to understand the networks that control glycan structure and function. Informatics tools are key enablers for processing the data that arise from multiple sources—biochemical pathways (Hossler et al. 2007) and multiple types of analytical structure determination techniques, as well as mathematical and computational modeling. By analyzing and extracting information from this sea of data, glycoscience can be studied in this systems context and ultimately understood and manipulated in controlled ways.

Such research as the above illustrates the possibility of whole cell simulation, but in order to perform simulations of higher organisms' cells, glycosylation and other posttranslational modifications and their kinetic reaction data will need to be incorporated. Greater advances in the Analytical Tools will aid in this. Moreover, predictions or assumptions can also be incorporated to perform simulations as necessary. With the availability and success of such simulations, perturbations to the model will enable predictions of phenotypic effects (Karr 2012).

5.6 INFORMATICS AND DATABASES

It is becoming increasingly evident that complex relationships between genomic DNA, transcripts, proteins, and their posttranslational modifications, such as phosphorylation and glycosylation, critically govern phenotypes of whole organisms. The development of informatics to capture, analyze, mine, and disseminate sequence information and datasets associated with genes and proteins has been instrumental in advancing genomics and proteomics. One major area of glycomics deals with understanding complex glycans that are attached to proteins during posttranslational modification and the biological functions mediated by these glycan modifications. Informatics applied to glycomics has been faced with unique challenges. The biosynthesis of glycans is complex, nontemplate driven, and involves tissue-specific isoforms of several glycan biosynthetic enzymes. As a result, it becomes challenging to decipher

the entire glycome of a whole organism in the same way that it has been possible for the genome and proteome.

The chemical heterogeneity of glycans also makes it challenging for any single analytical approach to provide a complete description of each glycan structure isolated from a glycoprotein or a cell type. Furthermore, glycan-protein interactions, leading to either the activation or inhibition of a biological response, are often not binary but rather involve more subtle mediation of a signaling pathway. In addition, glycan-protein interactions typically involve multivalency with regard to both the protein and the glycan. Because of these challenges, there are layers of ambiguity in determining primary sequence or chemical structure of a glycan that also impinge on understanding the specificity of glycan-protein interactions that modulate key biological functions.

An important factor in broadening appreciation of glycomics to the larger scientific community is the urgent need to develop databases, computational, and informatics tools to acquire, integrate, annotate, mine, and disseminate glycomics datasets such as analytical data, glycan array data, and glycogene expression data (Packer et al. 2008). Many earlier efforts in glycomics focused on structural characterization of glycans and on the development of glycan structure databases and computational tools to assist assignment of glycan structures from high-throughput analytical datasets. The development of these tools has advanced to a point where it is possible to obtain robust and detailed profiling of a majority of glycans isolated from cells, tissues, and individual glycoproteins.

To accelerate the development of additional databases and informatics tools, glycomics can to some extent borrow many of the tools that were developed for proteomics and genomics, but there are specific characteristics of glycans that require the development of different, and unique, tools. The most obvious difference is that glycans, unlike proteins or nucleic acids, are branched, isomeric, and constructed using several types of linkages. A common theme that unites these challenges is that there is no template from which glycan structure originates, and thus an "ensemble" of structures is created. Representing the complexity of glycan structures and the diversity of context—the fact that expression levels for each glycan, as well as glycosylation patterns, differ across cells and tissues—presents a significant challenge for bioinformatics approaches.

5.6.1 Limited Successes in Developing Broadly Available Informatics Tools

Many notable advances in glycomics informatics and database development have focused on interpretation of analytical data, including

assignment of NMR and mass spectrometry peaks. These advances have, to a certain extent, made glycan analysis more accessible to the broader research community. For example, to assist researchers in the assignment of glycan structures and features based on NMR data, the characteristic NMR chemical shifts and coupling constants of glycans reported in the literature have been compiled in accessible databases such as at the glycosciences.de portal (http://www.glycosciences.de/sweetdb/; Lütteke et al. 2006) and CASPER (http://www.casper.organ.su.se/casper/; Loss et al. 2006), thus improving the accessibility of NMR as a tool for glycoscientists. Several tools, including Glyco-Search-MS (http://www.glycosciences.de/sweetdb/start.php?action=form_ms_search; Loss et al. 2002) and GlycoWorkbench (http://www.glycoworkbench.org/; Ceroni et al. 2008), have focused on interpretation of mass spectrometry fragmentation patterns through comparison to reference datasets, thereby deducing the most likely glycan structure. Additional development in this area, including pairing with proteomics to enable analysis of glycopeptides and proteins, will further increase the usefulness of these tools.

More recent efforts have focused on developing computational tools to mine multiple high-throughput datasets associated with gene expression studies, glycan profiling, and glycan array screening. One area of application of these tools has been in correlating and predicting profiles of glycan structures in a cell based on expression of glycan biosynthesis enzymes (Kawano et al. 2005). Another area of active development has been in mining glycan array datasets to identify glycan sequence motifs recognized by various proteins, such as plant and animal lectins, pathogen proteins, and antibodies (Hizukuri et al. 2005; Aoki-Kinoshita et al. 2006; Kuboyama et al. 2006; Hashimoto et al. 2008a,b; Porter et al. 2010; Jiang et al. 2011b). These glycan sequence motifs represent a combination of substructures that favor binding and those that are detrimental to binding. Furthermore, identification of such binding motifs facilitates using protein-glycan co-crystal structures to translate biochemical and biophysical aspects of glycan-protein interactions to the biology mediated by these interactions.

Data mining methods and tools have been developed to solve a variety of problems in glycobiology, including extraction of potential glycan biomarkers; prediction of glycan-binding patterns (Hashimoto et al. 2008a,b; Aoki-Kinoshita et al. 2006); and the analysis of glycan biosynthesis pathways (Krambeck et al. 2009), as provided by the RINGS Web resource (http://www.rings.t.soka.ac.jp). Many computer theoretical methods have been applied to glycan analysis, including pairwise (Aoki et al. 2003) and multiple alignment of glycans and the development of "score matrices" for analysis of glycosidic linkages (Aoki et al. 2005). Such applications of existing bioinformatics methods to glycobiology

can be made to further elucidate glycan function (Aoki-Kinoshita, 2010). However, currently there is a severe lack in interest by the bioinformatics community in glycoscience as a result of the lack of a consistent database with relevant links to major databases and an understandable glycan representation format. Without easily available data of biological interest, bioinformatics research will not progress very far in the glycosciences, creating an ever-increasing gap between the genomics and proteomics world and glycomics. Moreover, without a consistent format for representing glycan structures, not only is there confusion regarding a "correct" representation of glycans, but also the integration of various computational tools becomes difficult.

5.6.2 Critical Need for Development of a Single Integrated Database

Clearly, developing a structural assignment database is key to a larger integrative effort to make glycomics accessible and relevant. Indeed, in the absence of a centralized database at a location such as the National Center for Biotechnology Information (NCBI), glycomics will not gain the attention and respect of the scientific community. Long-term funding and long-term stability of such an internationally supported database is absolutely critical for the future of glycosciences.

Larger and more complete informatics efforts can then focus on development of computational tools to correlate glycan structure with expression of biosynthetic enzymes to link biosynthesis and end product. Also, the development of new technologies, such as glycan array platforms to characterize glycan-protein interactions, have necessitated development of novel tools and database strategies for these high-throughput sources of data. In addition, there is a wealth of data on phenotypic analysis of knockout mice that lack specific glycan biosynthesis enzymes that could benefit from a database. Integration of gene expression, structural characterization, glycan motif recognition by various proteins, and whole-organism phenotyping data will enable a critical understanding of glycan diversity in a normal versus a perturbed cell and how these differences correlate with the physiological state of the cell. To truly "reduce this to practice," the field will need relational databases to make sense of the huge amount of data that will come from such studies and to develop trait correlations that will ultimately lead back to candidate genes.

Unfortunately, current glycobiology databases are largely incomplete, disconnected, and inaccessible to the broader community and have a high percentage of incorrect entries that require correction. Analytical databases today provide only "sound bytes" and are missing a great deal of the complexity. In addition, other structural databases need to be made "glycan aware."

To circumvent these challenges, it is critical that a centralized glycan database is created wherein all glycan structures that have been sequenced and published are registered. This database, then, could be expanded to include information on gene expression and organism phenotyping data. Also needed are reporting standards that specify the minimum information that should be reported about a dataset or an experimental process that allows a user to interpret and use the data entered. This may require manual independent curation of data, although it should be possible to develop a curation system to assist in annotations to a certain extent. In addition, there needs to be a glycan equivalent of the Phred Score for nucleic acid bases that can provide the user with a measure of the level of certainty of information on a given linkage in a given structure in a database (Ewing and Green 1998; Ewing et al. 1998). This will allow incomplete yet useful structural data to be included in databases.

Currently, the GlycomeDB database has incorporated many major databases in a way that consolidates unique structures and provides links so that the original database entries can be retrieved (Ranzinger et al. 2011). In contrast, the GlycoSuiteDB database is a manually curated database of structures from the literature, and thus the number of entries is small, less than 4,000 versus more than 36,000 in GlycomeDB (Cooper et al. 2003). The total GlycomeDB entries represents the sum of the several incorporated databases, and only about 1,000 of the structures are fully characterized, used in biologically known pathways, and nonredundant. Similarly, GlycoSuiteDB contains approximately 1,500 eukaryotic structures fulfilling those requirements. In general, there are 10,000 structures on average in the major glycan structure databases—EurocarbDB, KEGG Glycan, Bacterial Carbohydrate Structure Database (BCSDB), and Consortium for Functional Glycomics (CFG)—although it should be noted that, with the exception of BCSDB, most of these contain mainly eukaryotic glycans. These databases also mainly contain *N*- and *O*-linked glycan structures, whereas glycolipid and proteoglycan structures are few. Moreover, fully characterized glycan structures (including all linkage information) are limited to about 2,000 structures. Therefore, several issues must be addressed in developing a comprehensive glycan (or glycoconjugate) structure database.

5.6.2.1 Standardized representations of glycan (or glycoconjugate) structures

Because glycan structures are not linear, a simple single-letter code for monosaccharides is insufficient to represent glycan structures accurately. This is further complicated by the various naming schemes of monosaccharides. While a database of monosaccharides is currently available (MonosaccharideDB; http://www.monosaccharidedb.org),

different researchers prefer to use different methods for representing glycan structures, including IUPAC, LINUCS (Bohne-Lang et al. 2001), Linear Code (Banin et al. 2002), GlycoCT (Herget et al. 2008), and KCF (Aoki-Kinoshita 2010). Although Glyde-II (Sahoo et al. 2005) has been established as the standard format for exchanging glycan structures, it is not human readable. There are also a number of ways to graphically represent glycan structures as cartoons, including the system originating from Stuart Kornfeld, expanded and optimized by Varki et al. (2009) and adopted by the CFG, and the Oxford system (Harvey 2011). To resolve this issue, informatics methods will need to accurately convert across different formats, which involves creating a knowledge base on the chemical structures behind the nomenclature for each naming scheme so that the residues are mapped accurately. MonosaccharideDB stores monosaccharide data as chemical information and provides mappings to various database formats. Database developers will need to keep in mind the various formats that are available and allow queries using different formats; this may be possible by linked glycan structure components with MonosaccharideDB.

5.6.2.2 Comprehensive representation of glycan and glycoconjugate structures

To characterize accurately the cellular glycome, development of more sensitive analytical techniques, including likely NMR and mass spectrometry, will be vital. In turn, the development of informatics methods to aid in structure determination will also be important, including ones that can integrate data from multiple techniques. It is likely that the development of such informatics methods will require collaboration among computer scientists, analytical chemists, and others.

5.6.2.3 Standard ontology for glycan function and localization

An ontology for representing glycan structures has been proposed, called "GlycO." However, beyond structures, a formal representation of glycans and how they were determined, their functions, and their relationship to other molecules still needs to be established. MIRAGE—Minimum Information Required for a Glycomics Experiment standard—is currently being developed as a reporting standard for glycomics experiments, based on MIAME and MIAPE. MIRAGE aims to specify "the minimum information that should be reported about a data set or an experimental process, to allow a reader to interpret and critically evaluate the conclusions reached, and to support their experimental corroboration." Such a standard will serve as the first step toward establishing a well-documented glycan structure database that can be linked back to the

original experimental data. Further ontologies for annotating glycan function may be similarly based on existing ontologies for genes and proteins.

5.6.2.4 Links to protein, lipid, and other related databases

To integrate knowledge about glycans with the broader community, glycan structures registered in any database should be linked to the proteins, lipids, cells, and other entities to which the glycan structures were bound or in which they were found. Furthermore, links to the proteins, viruses, and other binders with glycans must be documented and linked wherever possible. Currently, to the committee's knowledge, the UniProt database is the only major protein database that contains information regarding potential glycosylation sites in amino acid sequences. To get to this information, however, the user must know to look for it in UniProt, because it is not directly accessible from GenBank or InterPro. Such links to the major protein and lipid databases will facilitate more communication with other related fields, and some progress is occurring in this area. Glycan information in GlycoSuiteDB is currently linked to UniProt. There are plans to link it to the UniCarbKB database as a combination of GlycoSuiteDB and EuroCarbDB. Bioinformatics methods can also be applied more easily when linked with larger resources of data. Additionally, this effort should link with other structural biology efforts aimed at defining conformation of glycan structures and their interaction with binding partners, because conformation has proven to be one of the driving parameters for specificity and affinity.

5.6.3 Key Messages on Glycan Bioinformatics and Databases

The current challenge for the bioinformatics field is to develop a unified, curated, stable database, with long-term funding, that encompasses glycobiology in a broader context. Although significant efforts have been made, a range of issues remain to be addressed and information about glycans is not accessible in a manner similar to other types of biological information. A particular challenge is standardization and annotation of glycan information for databases, including representation, level of structural certainty, and minimal information.

The development of a unified and integrated database resource not only would aid the field directly but would also help scientists from other disciplines, including clinicians, better appreciate, understand, and become involved in glycoscience. There is a need to develop bioinformatics tools that can make connections between disease and glycan structure and represent those connections in a straightforward manner. The initial

efforts to create a database worthy of long-term support will require focus regarding its content and function, as defined by consensus of the community that will use it. Such a database will need to do a few things very well in a sustainable and unambiguous way that is independent of new methodologies. One first step could be the creation of a centralized structural database that can be extended by connecting it to other resources. Such a database must be based at a centralized location to assure long-term stability and continuity and cannot be dependent on any individual scientist or institution. Other supplemental databases with incomplete information may add value if made available in parallel to a fully curated and centralized database. A revolution in the development of such databases would bring other scientists into the field, demystify it, and provide a tool to educate individuals about glycoscience.

5.7 SUMMARY AND FINDINGS

As this chapter makes clear, a diverse suite of tools are available to synthesize glycans; understand glycan structures, functions, and interactions; and share and communicate glycan information across the research community. Important limitations in the toolkit currently restrict glycoscience to a field that is actively practiced by only a relatively small group of specialists. Existing tools are useful and provide a base from which to answer glycoscience questions; however, they are not adequate to advance the field to the point where it can realize its potential widely across biology, chemistry, and materials science. New energy and creative solutions, stemming not only from glycoscience specialists but from many others in the broader scientific community too, will be needed to address some of these technical challenges.

As a result, the committee finds that:

- Scientists and engineers need access to a broad array of chemically well-defined glycans.
- Over the past 30 years, tremendous advances have been made in chemical and enzymatic synthesis of glycans, but these methods remain relegated to specialized laboratories capable of producing only small quantities of a given glycan. For glycoscience to advance, significant further progress in glycan synthesis is needed to create widely applicable methodologies that generate both large and small quantities of any glycan on demand.
- A suite of widely applicable tools, analogous to those available for studying nucleic acids and proteins, is needed to detect, describe, and fully purify glycans from natural sources and then to characterize their chemical composition and structure.

- Continued advances in molecular modeling, verified by advanced chemical analysis and solution characterization tools, can generate insights for understanding glycan structures and properties.
- An expanded toolbox of enzymes and enzyme inhibitors for manipulating glycans would drive progress in many areas of glycoscience.
- A centralized accessible database linked to other molecular databases is needed to fully realize advancements in knowledge generated by an expanded effort in glycoscience. Glycan information is not currently accessible to the research community in an integrated and centralized manner similar to other biological information.

6

Deciphering the Glycome for Human Health and Sustainability: Findings, Recommendations, and Roadmap

As this report highlights, glycans are the fourth great class of macromolecules on which all life depends for existence, yet our knowledge of these substances continues to lag that of nucleic acids and proteins. Glycans are universal in living systems and play a central role in the etiology of all major human diseases. Thus, advances in glycoscience will be needed to realize the full potential of our investments in such areas as genomics and proteomics and in efforts to address and possibly prevent the root causes of human illness. Glycans represent Earth's largest and potentially most versatile natural resource, and advances in glycoscience are needed to turn that resource into a varied and sustainable source of food, fuel, chemicals, and materials.

Over the course of its data-gathering efforts and deliberations, the committee concluded there is good reason that our understanding of glycans pales in comparison to what is known about nucleic acids, proteins, and lipids: Glycoscience lacks the necessary technologies needed to fully decipher the glycome and make sense of the last step of information flow that transforms genetic information into phenotype. However, the committee also determined that enormous advances in technologies, spurred by the Human Genome Project, the modern molecular biology revolution, nanotechnology, microfluidics, information technologies, and other fields, have created a new opportunity to create the tools and methods needed to bring glycoscience up to par with genomics and proteomics.

The committee believes that a concerted effort in glycoscience is now necessary and will be sufficient to create the needed tools and methods.

More importantly, such an effort will attract the attention of researchers in a wide range of disciplines and democratize the field. Glycoscience, like genomics and nanotechnology, will then become a core discipline that is integrated across the entire scientific enterprise, spawning both advances in knowledge and economic activity. The return on an investment in glycoscience, as with genomics and nanotechnology, is likely to be substantial and to contribute significantly to the recently announced effort to develop a national bioeconomy.

The committee's recommendations seek to enable the development of better and more readily accessible tools for studying glycoscience and for applying glycoscience knowledge to questions across multiple fields. The committee has sought to prioritize areas where advances will be broadly applicable and where gaps in current capabilities cut across and currently limit research. Such areas include the chemical synthesis of glycans and the determination of glycan structures. Having accessible databases and bioinformatics tools is similarly of fundamental utility to the field. Longer-term, education of the scientific community and students about the functions of glycans will be important to achieving a roadmap for the future of the field. How to most effectively deploy resources to achieve these priorities and to enable glycoscience to contribute to advances in health, energy, and materials science will require additional discussion among multiple federal agencies as well as members of the broader scientific community, a discussion that extends beyond the committee's mandate in this report.

The committee's findings are detailed in preceding chapters; the findings fall into four general categories that can be summarized here. In the area of human health the committee finds that:

- Glycans are directly involved in the pathophysiology of every major disease.
- Additional knowledge from glycoscience will be needed to realize the goals of personalized medicine and to take advantage of the substantial investments in human genome and proteome research and its impact on human health.
- Glycans are increasingly important in pharmaceutical development.

In the area of energy the committee finds that:

- Plant cell walls, made mostly of glycans, represent the planet's dominant source of biological carbon sequestration, or biomass, and are a potentially sustainable and economical source of non-petroleum-based energy.

- Understanding cell wall structure and biosynthesis and overcoming the recalcitrance of plant cell walls to conversion into feedstocks that can be transformed into liquid fuels and other energy sources will be important to achieving a sustainable energy revolution. Glycoscience research will be necessary to advance this area.
- Glycoscience can contribute significantly to bioenergy development by advancing the understanding of how to increase biomass production per hectare and how to increase the yield of fermentable sugar per ton of biomass.

In the area of materials the committee finds that:

- By fostering a greater understanding of the properties of glycans and of plant cell wall construction and deconstruction, glycoscience can play an important role in the development of nonpetroleum-based sustainable new materials.
- Glycan-based materials have wide-ranging uses in such areas as fine chemicals and feedstocks, polymeric materials, and nanomaterials.
- There are many pathways to create a variety of functionalities on a glycan, creating a wide range of options for tailoring material properties.

Based on the above, the committee makes the following findings on the toolkit needed to advance glycoscience:

- Scientists and engineers need access to a broad array of chemically well-defined glycans.
- Over the past 30 years, tremendous advances have been made in chemical and enzymatic synthesis of glycans, but these methods remain relegated to specialized laboratories capable of producing only small quantities of a given glycan. For glycoscience to advance, significant further progress in glycan synthesis is needed to create widely applicable methodologies that generate both large and small quantities of any glycan on demand.
- A suite of widely applicable tools, analogous to those available for studying nucleic acids and proteins, is needed to detect, describe, and fully purify glycans from natural sources and then to characterize their chemical composition and structure.
- Continued advances in molecular modeling, verified by advanced chemical analysis and solution characterization tools, can generate insights for understanding glycan structures and properties.

- An expanded toolbox of enzymes and enzyme inhibitors for manipulating glycans would drive progress in many areas of glycoscience.
- A centralized accessible database linked to other molecular databases is needed to fully realize advancements in knowledge generated by an expanded effort in glycoscience. Glycan information is not currently accessible to the research community in an integrated and centralized manner similar to other biological information.

Based on these findings, the committee makes the following recommendations in order to achieve a more complete understanding of the importance of glycoscience and its impacts on health, energy, and material sciences. Each recommendation is followed by a series of roadmap goals. The capabilities created by the achievement of these recommendations will ensure that all interested researchers can efficiently and effectively incorporate glycoscience into their work.

1. The committee recommends that the development of transformative methods for the facile synthesis of carbohydrates and glycoconjugates be a high priority for NIH, NSF, DOE, and other relevant stakeholders.

Roadmap Goals

Within 7 years, have synthetic tools to be able to synthesize all known carbohydrates of up to octasaccharides, including substituents (e.g., acetyl, sulfate groups). This goal encompasses human glycoprotein and glycolipid glycans and proteoglycans, which are currently estimated to be 10,000-20,000 structures, along with plant and microbial glycans and polymers.

Within 10 years, have synthetic tools to be able to synthesize uniform batches, in milligram quantities, of all linear and branched glycans that will enable glycan arrays for identifying protein binding epitopes, provide standards for analytical methods development, and enable improved polysaccharide materials engineering and systematic studies for all fields to be conducted. This includes methods for synthesis of structures with isotopic enrichment of specific desired atoms that may be needed for a wide variety of studies.

Within 15 years, be able to synthesize any glycoconjugate or carbohydrate in milligram to gram quantities using routine procedures. Community access should be available through a web ordering system with rapid delivery.

2. The committee recommends that the development of transformative tools for detection, imaging, separation, and high-resolution structure determination of carbohydrate structures and complex mixtures be a high priority for NIH, NSF, DOE, FDA, and other relevant stakeholders.

Roadmap Goals

Over the next 5-10 years, develop the technology to purify, identify, and determine the structures of all the important glycoproteins, glycolipids, and polysaccharides in any biological sample. For glycoproteins, determine the significant glycans present at each glycosylation site. Develop agreed upon criteria for what constitutes the acceptable level of structural detail and purity.

Within 10 years, have the ability to undertake high-throughput sequencing of all *N*- and *O*-linked glycans from a single type of cell in a single week.

Within 10 years, have the ability to routinely determine the complete carbohydrate structure of any glycan or polymer repeat sequence including branching, anomeric linkages between glycans, and substituents.

Within 15 years, have the ability to determine glycoforms (a complete description of molecular species within a population that have the same polypeptide sequence) of any glycoprotein in a biological sample.

For example, one specific achievable step could be to apply the tools developed in the roadmap to characterize the set of glycomes in blood, including those of blood cells and plasma.

3. The committee recommends that the development of transformative capabilities for perturbing carbohydrate and glycoconjugate structure, recognition, metabolism, and biosynthesis be a high priority for NIH, NSF, DOE and other relevant stakeholders.

Roadmap Goals

Within 5 years, identify the genes involved in glycan and glycoconjugate metabolism in any organism whose genome has been sequenced, and identify the activities of at least 1,000 enzymes that may have utility as synthetic and research tools.

Within 10 years, be able to use all glyco-metabolic enzymes (e.g., glycosyltransferases, glycosidases) as well as other state of the art tools for perturbing and modifying glyco-metabolic pathways (knockouts, siRNAs, etc.) of utility to the biomedical and plant research communities.

Within 10 years, develop methods for creating specific inhibitors to any human, plant, or microbial glycosyltransferase suitable for in vitro and in vivo studies in order to perturb the biology mediated by these enzymes.

Within 15 years, develop imaging methods for studying glycan structure, localization, and metabolism in both living and non-living systems.

4. The committee recommends that robust, validated informatics tools be developed in order to enable accurate carbohydrate and glycoconjugate structural prediction, computational modeling, and data mining. This capability will broaden access of glycoscience data to the entire scientific community.

Roadmap Goals

Within 5 years, develop an open-source software package that can automatically annotate an entire glycan profile (such as from a mass spectrometry experiment) with minimal user interaction.

Within 5 years, develop the technology to perform computer simulations of carbohydrate interactions with other entities such as proteins and nucleic acids.

Within 10 years, develop the software to simulate a cellular system to predict the effects of perturbations in glycosylation of particular glycoconjugates and polysaccharides.

5. The committee recommends that a long-term-funded, stable, integrated, centralized database, including mammalian, plant and microbial carbohydrates and glycoconjugates, be established as a collaborative effort by all stakeholders. The carbohydrate structural database needs to be fully cross-referenced with databases that provide complementary biological information (e.g., PDB and GenBank). Furthermore, there should be a requirement for deposition of new structures into the database using a reporting standard for minimal information.

Roadmap Goals

Within 5 years, develop a long-term-funded, centralized glycan structure database with each entry highly annotated using standards adopted by the community and all the world's repositories of glycan structures. The database should be cross-referenced and open source to allow the community to develop database resources that draw on this resource and

improve its utility to investigators that wish to incorporate glycoscience in their work

Within 5 years, employ an active curation system to automatically validate glycan structures deposited into a database so that journals can provide authors with an easily accessible interface for submitting new glycan structures to the database.

To achieve the roadmap goals articulated in its recommendations, the committee notes that it will be of critical importance for the field to reach agreement on the standards of evidence and the nature of the assumptions that will be used to annotate and validate glycan structures within the next 2 to 3 years. For example, a level of certainty should be assigned to each linkage in the database, using a defined convention. Agreement on these standards is needed to avoid depositing large numbers of structures into databases that will ultimately prove more confusing than useful.

Finally, the committee noted that there is widespread lack of understanding and appreciation of glycoscience within the scientific and medical communities and among the general public. Glycans are integral components of living organisms, whether human, animal, plant, or microbe, and glycan products have applications in health, energy, and materials science.

The committee concludes that integrating glycoscience into relevant disciplines in high school, undergraduate, and graduate education, and developing curricula and standardized testing for science competency would increase public as well as professional awareness.

Roadmap Goals

Within 5 years, integration of glycoscience as a significant part of the science curriculum would include glycoscience as both lecture materials and hands-on experiments or activities.

Within 10 years, glycoscience will be integrated and taught at every level wherever it is relevant to understand the scientific content. Competency in glycoscience could also be included in all standardized testing wherever relevant (for example, as part of the SAT and GRE Subject Tests, the MCAT, and Medical Board Exams).

To achieve these goals, the committee notes that mechanisms would need to be implemented to define appropriate glycoscience competency and to incorporate glycoscience topics into educational frameworks at multiple levels. The process of setting education policies and developing and implementing curricula varies from state to state, university to university, and country to country. Although the committee cannot define

the specific steps needed to achieve its recommended education goals, the committee encourages the engagement of glycoscience experts in these processes.

Glycoscience is a vibrant field filled with challenging problems. Research can make significant contributions toward understanding and improving human health, creating next-generation fuels and materials, and contributing to economic innovation and development. Now is the time for glycoscience to be embraced broadly by the research community. Drawing in members from the full spectrum of chemistry, biology, materials science, engineering, medicine, and other disciplines will be needed to address the technical challenges described here. Although these challenges are substantial and complex, the results of achieving these goals have the potential to impact science in exciting ways.

References

Achyuthan, K. E., A. M. Achyuthan, P. D. Adams, S. M. Dirk, J. C. Harper, B. A. Simmons, and A. K. Singh. 2010. Supramolecular self-assembled chaos: polyphenolic lignin's barrier to cost-effective lignocellulosic biofuels. *Molecules* 15(12):8641-8688.

Adamczyk, B., T. Tharmalingam, and P. M. Rudd. 2011. Glycans as cancer biomarkers. *Biochimica et Biophysica Acta* 1820(9):1347-1353.

Albersheim, P., A. Darvill, K. Roberts, R. Sederoff, and A. Staehelin. 2010. *Plant Cell Walls*. New York: Garland Science, Taylor & Francis Group, LLC.

Antoine, T., B. Priem, A. Heyraud, L. Greffe, M. Gilbert, B. W. Wakarchuk, J. S. Lam, and E. Samain. 2003. Large-scale in vivo synthesis of the carbohydrate moieties of gangliosides GM1 and GM2 by metabolically engineered *Escherichia coli*. *ChemBioChem* 4(5):406-412.

Antoine, T., A. Heyraud, C. Bosso, and E. Samain. 2005. Highly efficient biosynthesis of the oligosaccharide moiety of the GD3 ganglioside by using metabolically engineered *Escherichia coli*. *Angewandte Chemie—International Edition* 44(9):1350-1352.

Aoki, K. F., A. Yamaguchi, Y. Okuno, T. Akutsu, N. Ueda, M. Kanehisa, and H. Mamitsuka. 2003. Efficient tree-matching methods for accurate carbohydrate database queries. *Genome Inform* 14:134-143.

Aoki, K. F., H. Mamitsuka, T. Akutsu, and M. Kanehisa. 2005. A score matrix to reveal the hidden links in glycans. *Bioinformatics* 21(8):1457-1463.

Aoki-Kinoshita, K., N. Ueda, H. Mamitsuka, and M. Kanehisa. 2006. ProfilePSTMM: Capturing tree-structure motifs in carbohydrate sugar chains. *Bioinformatics* 22(14):25-34.

Aoki-Kinoshita, K. F. 2010. *Glycome Informatics: Methods and Applications*. Boca Raton, FL: CRC Press.

Arnold, J. N., M. R. Wormald, R. B. Sim, P. M. Rudd, and R. A. Dwek. 2007. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annual Review of Immunology* 25:21-50.

Astronomo, R. D., and D. R. Burton. 2010. Carbohydrate vaccines: Developing sweet solutions to sticky situations? *Nature Reviews Drug Discovery* 9(4):308-324.

Avci, U., S. Pattathil, J. S. Miller, and M. G. Hahn. 2011. Immunological approaches to cell wall and biomass characterization. In: *Biomass Conversion: Methods and Protocols.* M. Himmel, ed. Athens, GA: Humana Press.

Banin, E., Y. Neuberger, Y. Altshuler, A. Halevi, O. Inbar, D. Nir, and A. Dukler. 2002. A novel linear code(r) nomenclature for complex carbohydrates. *Trends in Glycoscience and Glycotechnology* 14:127-137.

Bar-Peled, M., and M. A. O'Neill. 2011. Plant nucleotide sugar formation, interconversion, and salvage by sugar recycling. *Annual Review of Plant Biology* 62:127-155.

Bardoel, B. W., and J. A. G. van Strijp. 2011. Molecular battle between host and bacterium: Recognition in innate immunity. *Journal of Molecular Recognition* 24(6):1077-1086.

Barreiro, O., and F. Sanchez-Madrid. 2009. Molecular basis of leukocyte-endothelium interactions during the inflammatory response. *Revista Espanola De Cardiologia* 62(5):552-562.

Bendiak, B., T. T. Fang, and D. N. M. Jones. 2002. An effective strategy for structural elucidation of oligosaccharides through NMR spectroscopy combined with peracetylation using doubly C-13-labeled acetyl groups. *Canadian Journal of Chemistry—Revue Canadienne De Chimie* 80(8):1032-1050.

Bishop, J. R., and P. Gagneux. 2007. Evolution of carbohydrate antigens—microbial forces shaping host glycomes? *Glycobiology* 17(5):23R-34R.

Bishop, J. R., M. Schuksz, and J. D. Esko. 2007. Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* 446(7139):1030-1037.

Blixt, O., K. Allin, O. Bohorov, X. Liu, H. Andersson-Sand, J. Hoffmann, and N. Razi. 2008. Glycan microarrays for screening sialyltransferase specificities. *Glycoconjugate Journal* 25(1):59-68.

Boeggeman, E., B. Ramakrishnan, C. Kilgore, N. Khidekel, L. C. Hsieh-Wilson, J. T. Simpson, and P. K. Qasba. 2007. Direct identification of nonreducing GlcNAc residues on *N*-glycans of glycoproteins using a novel chemoenzymatic method. *Bioconjugate Chemistry* 18(3):806-814.

Bohne-Lang, A., E. Lang, T. Förster, and C. W. von der Lieth. 2001. LINUCS: Linear notation for unique description of carbohydrate sequences. *Carbohydrate Research* 336(1):1-11.

Boltje, T. J., T. Buskas, and G. J. Boons. 2009. Opportunities and challenges in synthetic oligosaccharide and glycoconjugate research. *Nature Chemistry* 1(8):611-622.

Bose-Basu, B., J. Zajicek, G. Bondo, S. Zhao, M. Kubsch, I. Carmichael, and A. S. Serianni. 2000. Deuterium nuclear spin-lattice relaxation times and quadrupolar coupling constants in isotopically labeled saccharides. *Journal of Magnetic Resonance* 144(10828189):207-216.

Bozell, J. J., and G. R. Petersen. 2010. Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. *Green Chemistry* 12(4):539-554.

Broekhuizen, L. N., H. L. Mooij, J. J. P. Kastelein, E. S. G. Stroes, H. Vink, and M. Nieuwdorp. 2009. Endothelial glycocalyx as potential diagnostic and therapeutic target in cardiovascular disease. *Current Opinion in Lipidology* 20(1):57-62.

Brown, J. R., F. Yang, A. Sinha, B. Ramakrishnan, Y. Tor, P. K. Qasba, and J. D. Esko. 2009. Deoxygenated disaccharide analogs as specific inhibitors of beta 1-4-galactosyltransferase 1 and selectin-mediated tumor metastasis. *Journal of Biological Chemistry* 284(8):4952-4959.

Buskas, T., P. Thompson, and G. J. Boons. 2009. Immunotherapy for cancer: Synthetic carbohydrate-based vaccines. *Chemical Communications* 36:5335-5349.

Butters, T. D. 2007. Gaucher disease. *Current Opinion in Chemical Biology* 11(4):412-418.

Carroll, A., and C. Somerville. 2009. Cellulosic biofuels. *Annual Review of Plant Biology* 60:165-182.

Celie, J., R. H. J. Beelen, and J. van den Born. 2009. Heparan sulfate proteoglycans in extravasation: Assisting leukocyte guidance. *Frontiers in Bioscience* 14:4932-4949.

Ceroni, A., K. Maass, H. Geyer, R. Geyer, A. Dell, and S. M. Haslam. 2008. GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of glycans. *Journal of Proteome Research* 7(4):1650-1659.

Chabre, Y. M., D. Giguere, B. Blanchard, J. Rodrigue, S. Rocheleau, M. Neault, S. Rauthu, A. Papadopoulos, A. A. Arnold, A. Imberty, and R. Roy. 2011. Combining glycomimetic and multivalent strategies toward designing potent bacterial lectin inhibitors. *Chemistry—A European Journal* 17(23):6545-6562.

Chang, P. V., J. A. Prescher, E. M. Sletten, J. M. Baskin, I. A. Miller, N. J. Agard, A. Lo, and C. R. Bertozzi. 2010. Copper-free click chemistry in living animals. *Proceedings of the National Academy of Sciences of the United States of America* 107(5):1821-1826.

Chiba, Y., and Y. Jigami. 2007. Production of humanized glycoproteins in bacteria and yeasts. *Current Opinion in Chemical Biology* 11(6):670-676.

Chichlowski, M., J. B. German, C. B. Lebrilla, and D. A. Mills. 2011. The influence of milk oligosaccharides on microbiota of infants: Opportunities for formulas. *Annual Review of Food Science and Technology* 2:331-351.

Chuck, G. S., C. Tobias, L. Sun, F. Kraemer, C. L. Li, D. Dibble, R. Arora, J. N. Bragg, J. P. Vogel, S. Singh, B. A. Simmons, M. Pauly, and S. Hake. 2011. Overexpression of the maize Corngrass1 micro-RNA prevents flowering, improves digestibility, and increases starch content of switchgrass. *Proceedings of the National Academy of Sciences of the United States of America* 108(42):17550-17555.

Clark, R. J., P. M. McDonough, E. Swanson, S. U. Trost, M. Suzuki, M. Fukuda, and W. H. Dillmann. 2003. Diabetes and the accompanying hyperglycemia impairs cardiomyocyte calcium cycling through increased nuclear *O*-GlcNAcylation. *Journal of Biological Chemistry* 278(45):44230-44237.

Cohen, M., and A. Varki. 2010. The Sialome—Far more than the sum of its parts. *OMICS: A Journal of Integrative Biology* 14(4):455-464.

Comelli, E. M., S. R. Head, T. Gilmartin, T. Whisenant, S. M. Haslam, S. J. North, N. K. Wong, T. Kudo, H. Narimatsu, J. D. Esko, K. Drickamer, A. Dell, and J. C. Paulson. 2006. A focused microarray approach to functional glycomics: Transcriptional regulation of the glycome. *Glycobiology* 16(2):117-131.

Cooper, C. A., H. J. Joshi, M. J. Harrison, M. R. Wilkins, and N. H. Packer. 2003. GlycoSuiteDB: A curated relational database of glycoprotein glycan structures and their biological sources; 2003 update. *Nucleic Acids Research* 31(1):511-513.

Copeland, R. J., J. W. Bullen, and G. W. Hart. 2008. Cross-talk between GlcNAcylation and phosphorylation: Roles in insulin resistance and glucose toxicity. *American Journal of Physiology—Endocrinology and Metabolism* 295(1):E17-28.

Costantino, P., R. Rappuoli, and F. Berti. 2011. The design of semi-synthetic and synthetic glycoconjugate vaccines. *Expert Opinion on Drug Discovery* 6(10):1045-1066.

Coutinho, M. F., L. Lacerda, and S. Alves. 2012. Glycosaminoglycan storage disorders: A review. *Biochemistry Research International* 2012:471325.

Cox, T. M., and M. B. Cachon-Gonzalez. 2012. The cellular pathology of lysosomal diseases. *Journal of Pathology* 226(2):241-254.

Coxon, B. 2009. Developments in the Karplus equation as they relate to the NMR coupling constants of carbohydrates. *Advances in Carbohydrate Chemistry and Biochemistry, Vol. 62*. D. Horton, ed. San Diego: Elsevier Academic Press.

Crocker, P. R., J. C. Paulson, and A. Varki. 2007. Siglecs and their roles in the immune system. *Nature Reviews Immunology* 7(4):255-266.

Crocker, P. R., and P. Redelinghuys. 2008. Siglecs as positive and negative regulators of the immune system. *Biochemical Society Transactions* 36:1467-1471.

Culyba, E. K., J. L. Price, S. R. Hanson, A. Dhar, C.-H. Wong, M. Gruebele, E. T. Powers, and J. W. Kelly. 2011. Protein native-state stabilization by placing aromatic side chains in *N*-glycosylated reverse turns. *Science* 331(6017):571-575.

Cummings, R. D. 2009. The repertoire of glycan determinants in the human glycome. *Molecular BioSystems* 5(10):1087-1104.

Dai, H., and X. M. Gao. 2011. Elevated levels of serum antibodies against alpha-1, 6-glucan in patients with systemic lupus erythematosus or rheumatoid arthritis. *Protein and Cell* 2(9):739-744.

Dehnert, K. W., J. M. Baskin, S. T. Laughlin, B. J. Beahm, N. N. Naidu, S. L. Amacher, and C. R. Bertozzi. 2012. Imaging the sialome during zebrafish development with copper-free click chemistry. *ChemBioChem* 13(3):353-357.

Dell, A., and H. R. Morris. 2001. Glycoprotein structure determination by mass spectrometry. *Science* 291(5512):2351-2356.

Dell, A., A. J. Reason, K. H. Khoo, M. Panico, R. A. McDowell, and H. R. Morris. 1994. Mass spectrometry of carbohydrate-containing biopolymers. *Methods in Enzymology* 230:108-132.

Demotte, N., V. Stroobant, P. J. Courtoy, P. Van Der Smissen, D. Colau, I. F. Luescher, C. Hivroz, J. Nicaise, J. L. Squifflet, M. Mourad, D. Godelaine, T. Boon, and P. van der Bruggen. 2008. Restoring the association of the T-cell receptor with CD8 reverses anergy in human tumor-infiltrating lymphocytes. *Immunity* 28(3):414-424.

Dennis, J. W., M. Granovsky, and C. E. Warren. 1999. Glycoprotein glycosylation and cancer progression. *Biochimica et Biophysica Acta—General Subjects* 1473(1):21-34.

Dias, W. B., and G. W. Hart. 2007. *O*-GlcNAc modification in diabetes and Alzheimer's disease. *Molecular BioSystems* 3(11):766-772.

Ding, S.-Y., Q. Xu, M. Crowley, Y. Zeng, M. Nimlos, R. Lamed, E. A. Bayer, and M. E. Himmel. 2008. A biophysical perspective on the cellulosome: New opportunities for biomass conversion. *Current Opinion in Biotechnology* 19(3):218-227.

DOE (U.S. Department of Energy). 2006. *Breaking the Biological Barriers to Cellulose Ethanol—A Joint Research Agenda*. Report from the December 2005 Workshop, DOE/SC-0095. U.S. Department of Energy Office of Science, Washington, D.C.

DOE. 2011. *U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry*. R. D. Perlack and B. J. Stokes (Leads), ORNL/TM-2011/224. Oak Ridge National Laboratory, Oak Ridge, TN. 227 pp.

Drake, P. M., W. Cho, B. S. Li, A. Prakobphol, E. Johansen, N. L. Anderson, F. E. Regnier, B. W. Gibson, and S. J. Fisher. 2010. Sweetening the pot: Adding glycosylation to the biomarker discovery equation. *Clinical Chemistry* 56(2):223-236.

Drouillard, S., H. Driguez, and E. Samain. 2006. Large-scale synthesis of H-antigen oligosaccharides by expressing Helicobacter pylori alpha 1,2-fucosyltransferase in metabolically engineered *Escherichia coli* cells. *Angewandte Chemie, International Edition* 45(11):1778-1780.

Drozdova, A., P. Bojarova, K. Krenek, L. Weignerova, B. Henssen, L. Elling, H. Christensen, H. H. Jensen, H. Pelantova, M. Kuzma, K. Bezouska, M. Krupova, D. Adamek, K. Slamova, and V. Kren. 2011. Enzymatic synthesis of dimeric glycomimetic ligands of NK cell activation receptors. *Carbohydrate Research* 346(12):1599-1609.

Dube, D. H., and C. R. Bertozzi. 2005. Glycans in cancer and inflammation: Potential for therapeutics and diagnostics. *Nature Reviews Drug Discovery* 4(6):477-488.

Duus, J., C. H. Gotfredsen, and K. Bock. 2000. Carbohydrate structural determination by NMR spectroscopy: Modern methods and limitations. *Chemical Reviews* 100(12):4589-4614.

Dwek, R. A. 1996. Glycobiology: Toward understanding the function of sugars. *Chemical Reviews* 96(2):683-720.

Dykhuizen, E. C., J. F. May, A. Tongpenyai, and L. L. Kiessling. 2008. Inhibitors of UDP-galactopyranose mutase thwart mycobacterial growth. *Journal of the American Chemical Society* 130(21):6706-6707.

Erbacher, A., F. Gieseke, R. Handgretinger and I. Muller. 2009. Dendritic cells: Functional aspects of glycosylation and lectins. *Human Immunology* 70(5):308-312.

Ernst, B., and J. L. Magnani. 2009. From carbohydrate leads to glycomimetic drugs. *Nature Reviews Drug Discovery* 8(8):661-677.

Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* 8(3):186-194.

Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* 8(3):175-185.

Fadda, E., and R. J. Woods. 2010. Molecular simulations of carbohydrates and protein-carbohydrate interactions: Motivation, issues and prospects. *Drug Discovery Today* 15(15-16):596-609.

Fang, T. T., and B. Bendiak. 2007. The stereochemical dependence of unimolecular dissociation of monosaccharide-glycolaldehyde anions in the gas phase: A basis for assignment of the stereochemistry and anomeric configuration of monosaccharides in oligosaccharides by mass spectrometry via a key discriminatory product ion of disaccharide fragmentation, m/z 221. *Journal of the American Chemical Society* 129(31):9721-9736.

Fattal I., N. Shental, D. Mevorach, J. M. Anaya, A. Livneh, P. Langevitz, G. Zandman-Goddard, R. Pauzner, M. Lerner, M. Blank, M. E. Hincapie, U. Gafter, Y. Naparstek, Y. Shoenfeld, E. Domany, and I. R. Cohen. 2010. An antibody profile of systemic lupus erythematosus detected by antigen microarray. *Immunology* 130(3):337-343.

Frank, M., and S. Schloissnig. 2010. Bioinformatics and molecular modeling in glycobiology. *Cellular and Molecular Life Sciences* 67(16):2749-2772.

Freeze, H. H., and B. G. Ng. 2011. Golgi glycosylation and human inherited diseases. *Cold Spring Harbor Perspectives in Biology* 3(9):a005371.

Freeze, H. H., and V. Sharma. 2010. Metabolic manipulation of glycosylation disorders in humans and animal models. *Seminars in Cell & Developmental Biology* 21(6):655-662.

Fulop, N., M. M. Mason, K. Dutta, P. P. Wang, A. J. Davidoff, R. B. Marchase, and J. C. Chatham. 2007. Impact of Type 2 diabetes and aging on cardiomyocyte function and *O*-linked *N*-acetylglucosamine levels in the heart. *American Journal of Physiology—Cell Physiology* 292(4):C1370-C1378.

Galan, M. C., D. Benito-Alifonso, and G. M. Watt. 2011. Carbohydrate chemistry in drug discovery. *Organic & Biomolecular Chemistry* 9(10):3598-3610.

Gantt, R. W., P. Peltier-Pain, W. J. Cournoyer, and J. S. Thorson. 2011a. Using simple donors to drive the equilibria of glycosyltransferase-catalyzed reactions. *Nature Chemical Biology* 7(10):685-691.

Gantt, R. W., P. Peltier-Pain, and J. S. Thorson. 2011b. Enzymatic methods for glyco(diversification/randomization) of drugs and small molecules. *Natural Product Reports* 28(11):1811-1853.

Garber, K. C. A., K. Wangkanont, E. E. Carlson, and L. L. Kiessling. 2010. A general glycomimetic strategy yields non-carbohydrate inhibitors of DC-SIGN. *Chemical Communications* 46(36):6747-6749.

Garcia, I., M. Marradi, and S. Penades. 2010. Glyconanoparticles: Multifunctional nanomaterials for biomedical applications. *Nanomedicine* 5(5):777-792.

Ghaderi, D., R. E. Taylor, V. Padler-Karavani, S. Diaz, and A. Varki. 2010. Implications of the presence of *N*-glycolylneuraminic acid in recombinant therapeutic glycoproteins. *Nature Biotechnology* 28(8):863-867.

Gilbert, M., R. Bayer, A. M. Cunningham, S. Defrees, Y. H. Gao, D. C. Watson, N. M. Young, and W. W. Wakarchuk. 1998. The synthesis of sialylated oligosaccharides using a CMP-Neu5Ac synthetase/sialyltransferase fusion. *Nature Biotechnology* 16(8):769-772.

Godfrey, C., A. R. Foley, E. Clement, and F. Muntoni. 2011. Dystroglycanopathies: Coming into focus. *Current Opinion in Genetics & Development* 21(3):278-285.

Golks, A., T. T. T. Tran, J. F. Goetschy, and D. Guerini. 2007. Requirement for *O*-linked *N*-acetylglucosaminyltransferase in lymphocytes activation. *EMBO Journal* 26(20):4368-4379.

Gomord, V., P. Chamberlain, R. Jefferis, and L. Faye. 2005. Biopharmaceutical production in plants: Problems, solutions and opportunities. *Trends in Biotechnology* 23(11):559-565.

Gray, G. R. 1990. Linkage analysis using reductive cleavage method. *Methods in Enzymology* 193:573-587.

Greene, R. V. 2011. Basic energy sciences (BES) perspective. Presentation to the Committee on Assessing the Importance and Impact of Glycomics and Glycosciences, October 11, Washington, DC.

Gross, B. J., B. C. Kraybill, and S. Walker. 2005. Discovery of *O*-GlcNAc transferase inhibitors. *Journal of the American Chemical Society* 127(42):14588-14589.

Guerrini, M., D. Beccati, Z. Shriver, A. Naggi, K. Viswanathan, A. Bisio, I. Capila, J. C. Lansing, S. Guglieri, B. Fraser, A. Al-Hakim, N. S. Gunay, Z. Zhang, L. Robinson, L. Buhse, M. Nasr, J. Woodcock, R. Langer, G. Venkataraman, R. J. Linhardt, B. Casu, G. Torri, and R. Sasisekharan. 2008. Oversulfated chondroitin sulfate is a contaminant in heparin associated with adverse clinical events. *Nature Biotechnology* 26(6):669-675.

Habibi, Y., L. A. Lucia, and O. J. Rojas. 2010. Cellulose nanocrystals: Chemistry, self-assembly, and applications. *Chemical Reviews* 110(6):3479-3500.

Hakomori, S. 1964. A rapid permethylation of glycolipid, and polysaccharide catalyzed by methylsulfinyl carbanion in dimethyl sulfoxide. *Journal of Biological Chemistry* 55: 205-208.

Hakomori, S. 2002. Glycosylation defining cancer malignancy: New wine in an old bottle. *Proceedings of the National Academy of Sciences of the United States of America* 99(16): 10231-10233.

Haltiwanger, R. S., and J. B. Lowe. 2004. Role of glycosylation in development. *Annual Review of Biochemistry* 73:491-537.

Hancock, S. M., M. D. Vaughan, and S. G. Withers. 2006. Engineering of glycosidases and glycosyltransferases. *Current Opinion in Chemical Biology* 10(5):509-519.

Hanson, S. R., E. K. Culyba, T.-L. Hsu, C.-H. Wong, J. W. Kelly, and E. T. Powers. 2009. The core trisaccharide of an *N*-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proceedings of the National Academy of Sciences of the United States of America* 106(9):3131-3136.

Hart, G. W., and R. J. Copeland. 2010. Glycomics hits the big time. *Cell* 143(5):672-676.

Hart, G. W., M. P. Housley, and C. Slawson. 2007. Cycling of *O*-linked beta-*N*-acetylglucosamine on nucleocytoplasmic proteins. *Nature* 446(7139):1017-1022.

Hart, G. W., C. Slawson, G. Ramirez-Correa, and O. Lagerlof. 2011. Cross talk between *O*-glcNAcylation and phosphorylation: Roles in signaling, transcription, and chronic disease. *Annual Review of Biochemistry* 80:825-858.

Hashimoto, K., K. F. Aoki-Kinoshita, N. Ueda, M. Kanehisa, and H. Mamitsuka. 2008a. A new efficient probabilistic model for mining labeled ordered trees applied to glycobiology. *ACM Transactions on Knowledge Discovery from Data* 2(1):1-30.

Hashimoto, K., I. Takigawa, M. Shiga, M. Kanehisa, and H. Mamitsuka. 2008b. Mining significant tree patterns in carbohydrate sugar chains. *Bioinformatics* 24(16):167-173.

Heimburg-Molinaro, J., M. Lum, G. Vijay, M. Jain, A. Almogren, and K. Rittenhouse-Olson. 2011. Cancer vaccines and carbohydrate epitopes. *Vaccine* 29(48):8802-8826.

Helm, J. S., Y. N. Hu, L. Chen, B. Gross, and S. Walker. 2003. Identification of active-site inhibitors of MurG using a generalizable, high-throughput glycosyltransferase screen. *Journal of the American Chemical Society* 125(37):11168-11169.

Hennet, T. 2012. Diseases of glycosylation beyond classical congenital disorders of glycosylation. *Biochimica et Biophysica Acta* 1820(9):1306-1317.

Hennet, T., and L. G. Ellies. 1999. The remodeling of glycoconjugates in mice. *Biochimica et Biophysica Acta—General Subjects* 1473(1):123-136.

Henrissat, B., A. Surolia, and P. Stanley. 2009. A genomic view of glycobiology. In *Essentials of Glycobiology*. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, M. E. Etzler, eds. New York: Cold Spring Harbor.

Herget, S., R. Ranzinger, K. Maass, and C. W. Lieth. 2008. GlycoCT—a unifying sequence format for carbohydrates. *Carbohydrate Research* 343(12):2162-2171.

Hess, M. A., R. Sczyrba, T.-W. Egan, H. Kim, H. Chokhawala, G. Schroth, S. Luo, D. S. Clark, F. Chen, T. Zhang, R. I. Machie, L. A. Pennacchio, S. G. Tringe, A. Visel, T. Woyke, Z. Wang, and E. M. Rubin. 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331:463-467.

Himmel, M. E., ed. 2008. Front matter. *Biomass Recalcitrance: Deconstructing the Cell Wall for Bioenergy*. United Kingdom: Blackwell Publishing, John Wiley & Sons.

Himmel, M. E., S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady, and T. D. Foust. 2007. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* 315(5813):804-807.

Hizukuri, Y., Y. Yamanishi, O. Nakamura, F. Yagi, S. Goto, and M. Kanehisa. 2005. Extraction of leukemia-specific glycan motifs in humans by computational glycomics. *Carbohydrate Research* 340(14):2270-2278.

Horton, D., and J. D. Wander. 1980. Naturally occurring amino sugars. Pp. 726-760 in *The Carbohydrates, Chemistry and Biochemistry*. W. Pigman, D. Horton, and J. D. Wander, eds. New York: Academic Press.

Hossler, P., B. C. Mulukutla, and W.-S. Hu. 2007. Systems analysis of *N*-glycan processing in mammalian cells. *PLoS ONE* 2(8):e713.

Hsu, C. H., S. C. Hung, C. Y. Wu, and C. H. Wong. 2011. Toward automated oligosaccharide synthesis. *Angewandte Chemie, International Edition* 50(50):11872-11923.

Hu, Y., J. Suarez, E. Fricovsky, H. Wang, B. T. Scott, S. A. Trauger, W. L. Han, M. O. Oyeleye, and W. H. Dillmann. 2009. Increased enzymatic *O*-GlcNAcylation of mitochondrial proteins impairs mitochondrial function in cardiac myocytes exposed to high glucose. *Journal of Biological Chemistry* 284(1):547-555.

Huang, Y. L., and C. Y. Wun. 2010. Carbohydrate-based vaccines: Challenges and opportunities. *Expert Review of Vaccines* 9(11):1257-1274.

Imberty, A., Y. M. Chabre, and R. Roy. 2008. Glycomimetics and glycodendrimers as high-affinity microbial anti-adhesins. *Chemistry—A European Journal* 14(25):7490-7499.

Jaeken, J. 2011. Congenital disorders of glycosylation (CDG): It's (nearly) all in it! *Journal of Inherited Metabolic Disease* 34(4):853-858.

Jaipuri, F. A., and N. L. Pohl. 2008. Toward solution-phase automated iterative synthesis: Fluorous-tag assisted solution-phase synthesis of linear and branched mannose oligomers. *Organic & Biomolecular Chemistry* 6(15):2686-2691.

Jandus, C., H. U. Simon, and S. von Gunten. 2011. Targeting siglecs—a novel pharmacological strategy for immuno- and glycotherapy. *Biochemical Pharmacology* 82(4):323-332.

Jang-Lee J., S. J. North, M. Sutton-Smith, D. Goldberg, M. Panico, H. Morris, S. Haslam, and A. Dell. 2006. Glycomic profiling of cells and tissues by mass spectrometry: Fingerprinting and sequencing methodologies. *Methods in Enzymology* 415:59-86.

Jefferis, R. 2009. Glycosylation as a strategy to improve antibody-based therapeutics. *Nature Reviews Drug Discovery* 8(3):226-234.

Jeffrey, G. A., and M. Sundaralingam. 1985. Bibliography of crystal-structures of carbohydrates, nucleosides, and nucleotides for 1979 and 1980—addenda and errata for 1970-1978—and index for 1935-1980. *Advances in Carbohydrate Chemistry and Biochemistry* 43:203-421.

Jiang, D. H., J. R. Liang, and P. W. Noble. 2011a. Hyaluronan as an immune regulator in human diseases. *Physiological Reviews* 91(1):221-264.

Jiang, H., K. F. Aoki-Kinoshita, and W. K. Ching. 2011b. Extracting glycan motifs using a biochemically weighted kernel. *Bioinformation* 7(8):405-412.

Johnson, R. A. 2012. Emerging trends and drivers in the global biotechnology industry. Available at www.unog.ch/bwc/meeting [accessed August 1, 2012].

Jokela, T. A., K. M. Makkonen, S. Oikari, R. Karna, E. Koli, G. W. Hart, R. H. Tammi, C. Carlberg, and M. I. Tammi. 2011. Cellular content of UDP-*N*-acetylhexosamines controls hyaluronan synthase 2 expression and correlates with *O*-linked *N*-acetylglucosamine modification of transcription factors YY1 and SP1. *Journal of Biological Chemistry* 286(38):33632-33640.

Jones, S. P. 2005. A bittersweet modification *O*-GlcNAc and cardiac dysfunction. *Circulation Research* 96(9):925-926.

Kaddurah-Daouk, R., B. S. Kristal, and R. M. Weinshilboum. 2008. Metabolomics: A global biochemical approach to drug response and disease. *Annual Review of Pharmacology and Toxicology* 48:653-683.

Kaeothip, S., and A. V. Demchenko. 2011. Expeditious oligosaccharide synthesis via selective, semi-orthogonal, and orthogonal activation. *Carbohydrate Research* 346(12):1371-1388.

Kaida, K., T. Ariga, and R. K. Yu. 2009. Antiganglioside antibodies and their pathophysiological effects on Guillain-Barré syndrome and related disorders—a review. *Glycobiology* 19(7):676-692.

Karlsson, K. A. 1998. Meaning and therapeutic potential of microbial recognition of host glycoconjugates. *Molecular Microbiology* 29(1):1-11.

Karr, J. R., J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass, and M. W. Covert. 2012. A whole-cell computational model predicts phenotype from genotype. *Cell* 150(2):389-401.

Kawano, S., K. Hashimoto, T. Miyama, S. Goto, and M. Kanehisa. 2005. Prediction of glycan structures from gene expression data based on glycosyltransferase reactions. *Bioinformatics* 21(21):3976-3982.

Khidekel, N., S. B. Ficarro, E. C. Peters, and L. C. Hsieh-Wilson. 2004. Exploring the *O*-GlcNAc proteome: Direct identification of *O*-GlcNAc-modified proteins from the brain. *Proceedings of the National Academy of Sciences of the United States of America* 101(36): 13132-13137.

Kiessling, L. L., and R. A. Splain. 2010. Chemical approaches to glycobiology. *Annual Review of Biochemistry* 79:619-653.

Klemm, D., B. Heublein, H. P. Fink, and A. Bohn. 2005. Cellulose: Fascinating biopolymer and sustainable raw material. *Angewandte Chemie—International Edition* 44(22):3358-3393.

Klemm D., F. Kramer, S. Moritz, T. Lindstrom, M. Ankerfors, D. Gray, and A. Dorris. 2011. Nanocelluloses: A new family of nature-based materials. *Angewandte Chemie International Edition* 50:5438-5466.

Kobayashi, M., M. Fukuda, and J. Nakayama. 2009a. Role of sulfated *O*-glycans expressed by high endothelial venule-like vessels in pathogenesis of chronic inflammatory gastrointestinal diseases. *Biological & Pharmaceutical Bulletin* 32(5):774-779.

Kobayashi, M., H. Lee, J. Nakayama, and M. Fukuda. 2009b. Roles of gastric mucin-type *O*-glycans in the pathogenesis of *Helicobacter pylori* infection. *Glycobiology* 19(5):453-461.

Kochlamazashvili, G., O. Senkov, S. Grebenyuk, C. Robinson, M. F. Xiao, K. Stummeyer, R. Gerardy-Schahn, A. K. Engel, L. Feig, A. Semyanov, V. Suppiramaniam, M. Schachner, and A. Dityatev. 2010. Neural cell adhesion molecule-associated polysialic acid regulates synaptic plasticity and learning by restraining the signaling through GluN2B-containing NMDA receptors. *Journal of Neuroscience* 30(11):4171-4183.

Koeller, K. M., and C. H. Wong. 2001. Enzymes for chemical synthesis. *Nature* 409(6817): 232-240.

Kolarich, D., B. Lepenies, and P. H. Seeberger. 2012. Glycomics, glycoproteomics and the immune system. *Current Opinion in Chemical Biology* 16(1-2):214-220.

Kornfeld, S. 2010. A fascination with sugars. *Molecular Biology of the Cell* 21(22):3773-3775.

Korpos, E., C. Wu, and L. Sorokin. 2009. Multiple roles of the extracellular matrix in inflammation. *Current Pharmaceutical Design* 15(12):1349-1357.

Krambeck, F. J., S. V. Bennun, S. Narang, S. Choi, K. J. Yarema, and M. J. Betenbaugh. 2009. A mathematical model to derive *N*-glycan structures and cellular enzyme activities from mass spectrometric data. *Glycobiology* 19(11):1163-1175.

Kremer, R. M. D., and C. Gallo-Rodriguez. 2004. Naturally occurring monosaccharides: Properties and synthesis. *Advances in Carbohydrate Chemistry and Biochemistry* 59:9-67.

Kuboyama, T., K. Hirata, K. Aoki-Kinoshita, H. Kashima, and H. Yasuda. 2006. A gram distribution kernel applied to glycan classification and motif extraction. *Genome Inform* 17(2):25-34.

Kudlow, J. E. 2006. Post-translational modification by *O*-GlcNAc: Another way to change protein function. *Journal of Cellular Biochemistry* 98(5):1062-1075.

Lachmann, R. H. 2003. Miglustat. Oxford GlycoSciences/Actelion. *Current Opinion in Investigational Drugs* 4(4):472-479.

Lairson, L. L., B. Henrissat, G. J. Davies, and S. G. Withers. 2008. Glycosyltransferases: Structures, functions, and mechanisms. *Annual Review of Biochemistry* 77:521-555.

Langer, H. F., and T. Chavakis. 2009. Leukocyte-endothelial interactions in inflammation. *Journal of Cellular and Molecular Medicine* 13(7):1211-1220.

Laubli, H., and L. Borsig. 2010. Selectins promote tumor metastasis. *Seminars in Cancer Biology* 20(3):169-177.

Laughlin, S. T., and C. R. Bertozzi. 2009. Imaging the glycome. *Proceedings of the National Academy of Sciences of the United States of America* 106(1):12-17.

Lazarus, B. D., D. C. Love, and J. A. Hanover. 2009. *O*-GlcNAc cycling: Implications for neurodegenerative disorders. *International Journal of Biochemistry & Cell Biology* 41(11):2134-2146.

Lee, S. U., A. Grigorian, J. Pawling, I. J. Chen, G. Gao, T. Mozaffar, C. McKerlie, and M. Demetriou. 2007. *N*-glycan processing deficiency promotes spontaneous inflammatory demyelination and neurodegeneration. *Journal of Biological Chemistry* 282(46): 33725-33734.

Lepenies, B., and P. H. Seeberger. 2010. The promise of glycomics, glycan arrays and carbohydrate-based vaccines. *Immunopharmacology and Immunotoxicology* 32(2):196-207.

Li, D., T. Mallory, and S. Satomura. 2001. AFP-L3: A new generation of tumor marker for hepatocellular carcinoma. *Clinica Chimica Acta* 313(1-2):15-19.

Li, M., L. J. Song, and X. Y. Qin. 2010. Glycan changes: Cancer metastasis and anti-cancer vaccines. *Journal of Biosciences* 35(4):665-673.

Liang, P. H., C. Y. Wu, W. A. Greenberg, and C. H. Wong. 2008. Glycan arrays: Biological and medical applications. *Current Opinion in Chemical Biology* 12(1):86-92.

Lindberg, B., and J. Lonngren. 1978. Methylation analysis of complex carbohydrates: General procedure and application for sequence analysis. *Methods in Enzymology* 50:3-33.

Liu, Q. P., G. Sulzenbacher, H. Yuan, E. P. Bennett, G. Pietz, K. Saunders, J. Spence, E. Nudelman, S. B. Levery, T. White, J. M. Neveu, W. S. Lane, Y. Bourne, M. L. Olsson, B. Henrissat, and H. Clausen. 2007. Bacterial glycosidases for the production of universal red blood cells. *Nature Biotechnology* 25(4):454-464.

Loss, A., P. Bunsmann, A. Bohne, E. Schwarzer, E. Lang, and C. W. von der Lieth. 2002. SWEET-DB: An attempt to create annotated data collections for carbohydrates. *Nucleic Acids Research* 30(1):405-408.

Loss, A., R. Stenutz, E. Schwarzer, and C. von der Lieth. 2006. GlyNest and CASPER: Two independent approaches to estimate 1H and 13C NMR shifts of glycans available through a common Web interface. *Nucleic Acids Research* 34(Web server issue):W733-W737.

Louthrenoo, W., N. Kasitanon, R. Wichainun, S. Wangkaew, W. Sukitawut, S. Kuwata, and F. Takeuchi. 2010. Anti-agalactosyl IgG antibodies in Thai patients with rheumatoid arthritis, systemic lupus erythematosus, and systemic sclerosis. *Clinical Rheumatology* 29(3):241-246.

Lowe, J. B., and J. D. Marth. 2003. A genetic approach to mammalian glycan function. *Annual Review of Biochemistry* 72:643-691.

Lu, Y. P., Y. H. P. Zhang, and L. R. Lynd. 2006. Enzyme-microbe synergy during cellulose hydrolysis by Clostridium thermocellum. *Proceedings of the National Academy of Sciences of the United States of America* 103(44):16165-16169.

Lütteke, T., A. Bohne-Lang, A. Loss, T. Goetz, M. Frank, and C. von der Lieth. 2006. GLYCOSCIENCES.de: An Internet portal to support glycomics and glycobiology research. *Glycobiology* 16(5):71R-81R.

Lux, A., and F. Nimmerjahn. 2011. Impact of differential glycosylation on IgG activity. *Advances in Experimental Medicine and Biology* 780:113-124.

Lynd, L. R. 2010. Bioenergy: In search of clarity. *Energy & Environmental Science* 3(9):1150-1152.

Lynd, L. R., and C. H. D. Cruz. 2010. Make way for ethanol. *Science* 330(6008):1176-1176.

Lynd, L. R., M. S. Laser, D. Brandsby, B. E. Dale, B. Davison, R. Hamilton, M. Himmel, M. Keller, J. D. McMillan, J. Sheehan, and C. E. Wyman. 2008. How biotech can transform biofuels. *Nature Biotechnology* 26(2):169-172.

Macnaughtan, M. A., F. Tian, S. Liu, L. Meng, S. Park, P. Azadi, K. W. Moremen, and J. H. Prestegard. 2008. (13)C-sialic acid labeling of glycans on glycoproteins using ST6Gal-I. *Journal of the American Chemical Society* 130(36):11864-11865.

Magnani, J. L., and B. Ernst. 2009. Glycomimetic drugs—a new source of therapeutic opportunities. *Discovery Medicine* 8(43):247-252.

Marino, K., J. Bones, J. J. Kattla, and P. M. Rudd. 2010. A systematic approach to protein glycosylation analysis: A path through the maze. *Nature Chemical Biology* 6(10):713-723.

Marth, J. D., and P. K. Grewal. 2008. Mammalian glycosylation in immunity. *Nature Reviews Immunology* 8(11):874-887.

Martin-Pastor, M., A. Canales-Mayordomo, and J. Jimenez-Barbero. 2003. NMR experiments for the measurement of proton-proton and carbon-carbon residual dipolar couplings in uniformly labelled oligosaccharides. *Journal of Biomolecular NMR* 26(4):345-353.

McEver, R. P. 2010. Rolling back neutrophil adhesion. *Nature Immunology* 11(4):282-284.

McEver, R. P., and C. Zhu. 2010. Rolling cell adhesion. *Annual Review of Cell and Developmental Biology* 26:363-396.

McGowan E. D., and K. Bowman. 2010. Background Paper on Glycosciences and Glycomics in the United States. Board on Chemical Sciences and Technology and Board on Life Sciences, National Research Council.

McLellan J. S., M. Pancera, C. Carrico, J. Gorman, J. P. Julien, R. Khayat, R. Louder, R. Pejchal, M. Sastry, K. Dai, S. O'Dell, N. Patel, S. Shahzad-ul-Hussan, Y. Yang, B. Zhang, T. Zhou, J. Zhu, J. C. Boyington, G. Y. Chuang, D. Diwanji, I. Georgiev, Y. D. Kwon, D. Lee, M. K. Louder, S. Moquin, S. D. Schmidt, Z. Y. Yang, M. Bonsignori, J. A. Crump, S. H. Kapiga, N. E. Sam, B. F. Haynes, D. R. Burton, W. C. Koff, L. M. Walker, S. Phogat, R. Wyatt, J. Orwenyo, L. X. Wang, J. Arthos, C. A. Bewley, J. R. Mascola, G. J. Nabel, W. R. Schief, A. B. Ward, I. A. Wilson, and P. D. Kwong. 2011. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* 480(7377):336-343.

Meany, D. L., and D. W. Chan. 2011. Aberrant glycosylation associated with enzymes as cancer biomarkers. *Clinical Proteomics* 8(1):7.

Mendonca, S., R. B. Cole, J. Zhu, Y. Cai, A. D. French, G. P. Johnson, and R. A. Laine. 2003. Incremented alkyl derivatives enhance collision induced glycosidic bond cleavage in mass spectrometry of disaccharides. *Journal of the American Society for Mass Spectrometry* 14(1):63-78.

Moon, R. J., A. Martini, J. Nairn, J. Simonsen, and J. Youngblood. 2011. Cellulose nanomaterials review: Structure, properties and nanocomposites. *Chemical Society Reviews* 40(7):3941-3994.

Morelle, W., and J. C. Michalski. 2005. The mass spectrometric analysis of glycoproteins and their glycan structures. *Current Analytical Chemistry* 1(1):29-57.

Morelli, L., L. Poletti, and L. Lay. 2011. Carbohydrates and immunology: Synthetic oligosaccharide antigens for vaccine formulation. *European Journal of Organic Chemistry* 29:5723-5777.

Morgan, W. T. J., and W. M. Watkins. 1969. Genetic and biochemical aspects of human blood group A- B- H- LEA- and LEB- specificity. *British Medical Bulletin* 25(1):30.

Muntoni, F., S. Torelli, D. J. Wells, and S. C. Brown. 2011. Muscular dystrophies due to glycosylation defects: Diagnosis and therapeutic strategies. *Current Opinion in Neurology* 24(5):437-442.

Narimatsu, H. 2006. Human glycogene cloning: Focus on beta 3-glycosyltransferase and beta 4-glycosyltransferase families. *Current Opinion in Structural Biology* 16(5):567-575.

National Research Council. 2011. *Life Science and Related Fields: Trends Relevant to the Biological Weapons Convention*. Washington, DC: National Academies Press.

Newburg, D. S., G. M. Ruiz-Palacios, and A. L. Morrow. 2005. Human milk glycans protect infants against enteric pathogens. *Annual Review of Nutrition* 25:37-58.

Nieuwdorp, M., M. C. Meuwese, H. Vink, J. B. L. Hoekstra, J. J. P. Kastelein, and E. S. G. Stroes. 2005. The endothelial glycocalyx: A potential barrier between health and vascular disease. *Current Opinion in Lipidology* 16(5):507-511.

Nyame, A. K., Z. S. Kawar, and R. D. Cummings. 2004. Antigenic glycans in parasitic infections: Implications for vaccines and diagnostics. *Archives of Biochemistry and Biophysics* 426(2):182-200.

Ohtsubo, K., and J. D. Marth. 2006. Glycosylation in cellular mechanisms of health and disease. *Cell* 126(5):855-867.

Ohtsubo, K., M. Z. Chen, J. M. Olefsky, and J. D. Marth. 2011. Pathway to diabetes through attenuation of pancreatic beta cell glycosylation and glucose transport. *Nature Medicine* 17(9):1067-1075.

Okahisa Y., A. Yoshida, S. Miyaguch, and H. Yano. 2009. Optically transparent wood cellulose nanocomposites as a base substrate for flexible organic light-emitting diode displays. *Composites Science and Technology* 69:1958-1961.

Olson, D. G., J. E. McBride, A. J. Shaw, and L. R. Lynd. 2011. Recent progress in consolidated bioprocessing. *Current Opinion in Biotechnology* 23(3):396-405.

Olsson, M. L., and H. Clausen. 2008. Modifying the red cell surface: Towards an ABO-universal blood supply. *British Journal of Haematology* 140(1):3-12.

O'Reilly, M. K., and J. C. Paulson. 2009. Siglecs as targets for therapy in immune-cell-mediated disease. *Trends in Pharmacological Sciences* 30(5):240-248.

Packer, N., C. von der Lieth, K. Aoki-Kinoshita, C. Lebrilla, J. Paulson, R. Raman, P. Rudd, R. Sasisekharan, N. Taniguchi, and W. York. 2008. Frontiers in glycomics: Bioinformatics and biomarkers in disease. *Proteomics* 8(1):8-20.

Palcic, M. M. 2011. Glycosyltransferases as biocatalysts. *Current Opinion in Chemical Biology* 15(2):226-233.

Park, Y. M., and C. B. Lebrilla. 2005. Application of Fourier transform ion cyclotron resonance mass spectrometry to oligosaccharides. *Mass Spectrometry Reviews* 24(2):232-264.

Pejchal R., K. J. Doores, L. M. Walker, R. Khayat, P. S. Huang, S. K. Wang, R. L. Stanfield, J. P. Julien, A. Ramos, M. Crispin, R. Depetris, U. Katpally, A. Marozsan, A. Cupo, S. Maloveste, Y. Liu, R. McBride, Y. Ito, R. W. Sanders, C. Ogohara, J. C. Paulson, T. Feizi, C. N. Scanlan, C. H. Wong, J. P. Moore, W. C. Olson, A. B. Ward, P. Poignard, W. R. Schief, D. R. Burton, and I. A. Wilson. 2011. A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science* 334(6059):1097-1103.

Perlack, R. D., L. L. Wright, A. F. Turhollow, R. L. Graham, B. J. Stokes, and D. C. Erbach. *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply* (U.S. Department of Energy and U.S. Department of Agriculture, April 2005; Available at feedstockreview.ornl.gov/pdf/billion_ton_vision.pdf [accessed August 1, 2012]).

Petitou, M., and C. A. van Boeckel. 2004. A synthetic antithrombin III binding pentasaccharide is now a drug! What comes next? *Angewandte Chemie, International Edition in English* 43(24):3118-3133.

Pieters, R. J. 2011. Carbohydrate mediated bacterial adhesion. *Bacterial Adhesion: Chemistry, Biology and Physics* 715:227-240.

Pinho, S. S., R. Seruca, F. Gartner, Y. Yamaguchi, J. G. Gu, N. Taniguchi, and C. A. Reis. 2011. Modulation of E-cadherin function and dysfunction by *N*-glycosylation. *Cellular and Molecular Life Sciences* 68(6):1011-1020.

Porter, A., T. Yue, L. Heeringa, S. Day, E. Suh, and B. B. Haab. 2010. A motif-based analysis of glycan array data to determine the specificities of glycan-binding proteins. *Glycobiology* 20(3):369-380.

Price, J. L., D. Shental-Bechor, A. Dhar, M. J. Turner, E. T. Powers, M. Gruebele, Y. Levy, and J. W. Kelly. 2010. Context-dependent effects of asparagine glycosylation on pin WW folding kinetics and thermodynamics. *Journal of the American Chemical Society* 132(43):15359-15367.

Qua, E. H., P. R. Hornsby, H. S. S. Sharma, and G. Lyons. 2011. Preparation and characterisation of cellulose nanofibres. *Journal of Materials Science* 46(18): 6029-6045.

Rabinovich, G. A., and M. A. Toscano. 2009. Turning "sweet" on immunity: Galectin-glycan interactions in immune tolerance and inflammation. *Nature Reviews Immunology* 9(5): 338-352.

Rabinovich, G. A., M. A. Toscanol, S. S. Jackson, and G. R. Vasta. 2007. Functions of cell surface galectin-glycoprotein lattices. *Current Opinion in Structural Biology* 17(5):513-520.

Raju, T. S. 2008. Terminal sugars of Fc glycans influence antibody effector functions of IgGs. *Current Opinion in Immunology* 20(4):471-478.

Rambaruth, N. D. S., and M. V. Dwek. 2011. Cell surface glycan-lectin interactions in tumor metastasis. *Acta Histochemica* 113(6):591-600.

Ramirez-Correa, G. A., W. H. Jin, Z. H. Wang, X. Zhong, W. D. Gao, W. B. Dias, C. Vecoli, G. W. Hart, and A. M. Murphy. 2008. *O*-Linked GlcNAc modification of cardiac myofilament proteins. *Circulation Research* 103(12):1354-1358.

Ramya, T. N. C., E. Weerapana, L. Liao, Y. Zeng, H. Tateno, L. Liao, J. R. Yates III, B. F. Cravatt, and J. C. Paulson. 2010. In situ trans ligands of CD22 identified by glycan-protein photocross-linking-enabled proteomics. *Molecular and Cellular Proteomics* 9(6):1339-1351.

Ranzinger, R., S. Herget, C. W. von der Lieth, and M. Frank. 2011. GlycomeDB—a unified database for carbohydrate structures. *Nucleic Acids Research* 39(Database issue):D373-D376.

Raska, M., and J. Novak. 2010. Involvement of envelope-glycoprotein glycans in HIV-1 biology and infection. *Archivum Immunologiae et Therapiae Experimentalis* 58(3):191-208.

Rehm, B. H. A. 2010. Bacterial polymers: Biosynthesis, modifications and applications. *Nature Reviews Microbiology* 8(8):578-592.

Reis, C. A., H. Osorio, L. Silva, C. Gomes, and L. David. 2010. Alterations in glycosylation as biomarkers for cancer detection. *Journal of Clinical Pathology* 63(4):322-329.

Rexach, J. E., P. M. Clark, D. E. Mason, R. L. Neve, E. C. Peters, and L. C. Hsieh-Wilson. 2012. Dynamic *O*-GlcNAc modification regulates CREB-mediated gene expression and memory formation. *Nature Chemical Biology* 8(3):253-261.

Rich, J. R., and S. G. Withers. 2009. Emerging methods for the production of homogeneous human glycoproteins. *Nature Chemical Biology* 5(4):206-215.

Rillahan, C. D., and J. C. Paulson. 2011. Glycan microarrays for decoding the glycome. *Annual Review of Biochemistry* 80:797-823.

Rillahan, C. D., S. J. Brown, A. C. Register, H. Rosen, and J. C. Paulson. 2011. High-throughput screening for inhibitors of sialyl- and fucosyltransferases. *Angewandte Chemie, International Edition* 50(52):12534-12537.

Roy, D., M. Semsarilar, J. T. Guthrie, and S. Perrier. 2009. Cellulose modification by polymer grafting: A review. *Chemical Society Reviews* 38(7):2046-2064.

Rubin, E. M. 2008. Genomics of cellulosic biofuels. *Nature* 454:841-845.

Rudd, P. M., and R. A. Dwek. 1997. Glycosylation: Heterogeneity and the 3D structure of proteins. *Critical Reviews in Biochemistry and Molecular Biology* 32(1):1-100.

Sackstein, R. 2009. Glycosyltransferase-Programmed Stereosubstitution (GPS) to create HCELL: Engineering a roadmap for cell migration. *Immunological Reviews* 230:140-163

Sahoo, S., C. Thomas, A. Sheth, C. Henson, and W. York. 2005. GLYDE—an expressive XML standard for the representation of glycan structure. *Carbohydrate Research* 340(18):2802-2807.

Satoh, T., Y. Kanai, M. H. Wu, H. Yokozeki, R. Kannagi, J. B. Lowe, and K. Nishioka. 2005. Synthesis of alpha(1,3) fucosyltransferases IV- and VII-dependent eosinophil selectin ligand and recruitment to the skin. *American Journal of Pathology* 167(3):787-796.

Satoh, M., S. Iida, and K. Shitara. 2006. Non-fucosylated therapeutic antibodies as next-generation therapeutic antibodies. *Expert Opinion on Biological Therapy* 6(11):1161-1173.

Sauerzapfe, B., K. Krenek, J. Schmiedel, W. W. Wakarchuk, H. Pelantova, V. Kren, and L. Elling. 2009. Chemo-enzymatic synthesis of poly-*N*-acetyllactosamine (poly-LacNAc) structures and their characterization for CGL2-galectin-mediated binding of ECM glycoproteins to biomaterial surfaces. *Glycoconjugate Journal* 26(2):141-159.

Schaffer, R. 1972. Occurrence, properties, and preparation of naturally occurring monosaccharides (including 6-deoxy sugars). Pp. 69-111 in *The Carbohydrates, Chemistry and Biochemistry*. W. Pigman and D. Horton, eds. New York: Academic Press.

Schauer, R. 2009. Sialic acids as regulators of molecular and cellular interactions. *Current Opinion in Structural Biology* 19(5):507-514.

Schmaltz, R. M., S. R. Hanson, and C. H. Wong. 2011. Enzymes in the synthesis of glycoconjugates. *Chemical Reviews* 111(7):4259-4307.

Schulze, H., and K. Sandhoff. 2011. Lysosomal lipid storage diseases. *Cold Spring Harbor Perspectives in Biology* 3(6).

Seeberger, P. H. 2008. Automated oligosaccharide synthesis. *Chemical Society Reviews* 37(1):19-28.

Seeberger, P. H., and D. B. Werz. 2007. Synthesis and medical applications of oligosaccharides. *Nature* 446(7139):1046-1051.

Sethuraman, N., and T. A. Stadheim. 2006. Challenges in therapeutic glycoprotein production. *Current Opinion in Biotechnology* 17(4):341-346.

Sharon, N., and H. Lis. 2004. History of lectins: From hemagglutinins to biological recognition molecules. *Glycobiology* 14(11):53R-62R.

Shriver, Z., S. Raguram, and R. Sasisekharan. 2004. Glycomics: A pathway to a class of new and improved therapeutics. *Nature Reviews Drug Discovery* 3(10):863-873.

Silva, Z., K. Konstantopoulos, and P. A. Videira. 2012. The role of sugars in dendritic cell trafficking. *Annals of Biomedical Engineering* 40(4):777-789.

Skorobogatko, Y. V., J. Deuso, J. Adolf-Bergfoyle, M. G. Nowak, Y. S. Gong, C. F. Lippa, and K. Vosseller. 2011. Human Alzheimer's disease synaptic *O*-GlcNAc site mapping and iTRAQ expression proteomics with ion trap mass spectrometry. *Amino Acids* 40(3):765-779.

Skrisovska, L., M. Schubert, and F. H. T. Allain. 2010. Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. *Journal of Biomolecular NMR* 46(1):51-65.

Slawson, C., and G. W. Hart. 2011. *O*-GlcNAc signalling: Implications for cancer cell biology. *Nature Reviews Cancer* 11(9):678-684.

Slawson, C., R. J. Copeland, and G. W. Hart. 2010. *O*-GlcNAc signaling: A metabolic link between diabetes and cancer? *Trends in Biochemical Sciences* 35(10):547-555.

Sletten, E. M., and C. R. Bertozzi. 2011. From mechanism to mouse: A tale of two bioorthogonal reactions. *Accounts of Chemical Research* 44(9):666-676.

Smith, D. F., X. Z. Song, and R. D. Cummings. 2010. Use of glycan microarrays to explore specificity of glycan-binding proteins. *Methods in Enzymology* 480:417-444.

Solomon, S. S., G. Majumdar, A. Martinez-Hernandez, and R. Raghow. 2008. A critical role of Sp1 transcription factor in regulating gene expression in response to insulin and other hormones. *Life Sciences* 83(9-10):305-312.

Sorokin, L. 2010. The impact of the extracellular matrix on inflammation. *Nature Reviews Immunology* 10(10):712-723.

Sperandio, M., C. A. Gleissner, and K. Ley. 2009. Glycosylation in immune cell trafficking. *Immunological Reviews* 230(1):97-113.

St Hill, C. A. 2011. Interactions between endothelial selectins and cancer cells regulate metastasis. *Frontiers in Bioscience-Landmark* 16:3233-3251.

Stanford, K. I., J. R. Bishop, E. M. Foley, J. C. Gonzales, I. R. Niesman, J. L. Witztum, and J. D. Esko. 2009. Syndecan-1 is the primary heparan sulfate proteoglycan mediating hepatic clearance of triglyceride-rich lipoproteins in mice. *Journal of Clinical Investigation* 119(11):3236-3245.

Stevens, J., O. Blixt, J. C. Paulson, and I. A. Wilson. 2006. Glycan microarray technologies: Tools to survey host specificity of influenza viruses. *Nature Reviews Microbiology* 4(11):857-864.

Stockert, R. J. 1995. The asialoglycoprotein receptor—relationships between structure, function, and expression. *Physiological Reviews* 75(3):591-609.

Tallent, M. K., N. Varghis, Y. Skorobogatko, L. Hernandez-Cuebas, K. Whelan, D. J. Vocadlo, and K. Vosseller. 2009. In vivo modulation of *O*-GlcNAc levels regulates hippocampal synaptic plasticity through interplay with phosphorylation. *Journal of Biological Chemistry* 284(1):174-181.

Tannock, L. R., and V. L. King. 2008. Proteoglycan mediated lipoprotein retention: A mechanism of diabetic atherosclerosis. *Reviews in Endocrine & Metabolic Disorders* 9(4):289-300.

Tao, N. A., S. A. Wu, J. Kim, H. J. An, K. Hinde, M. L. Power, P. Gagneux, J. B. German, and C. B. Lebrilla. 2011. Evolutionary glycomics: Characterization of milk oligosaccharides in primates. *Journal of Proteome Research* 10(4):1548-1557.

Taylor, M. E., and K. Drickamer. 2007. Paradigms for glycan-binding receptors in cell adhesion. *Current Opinion in Cell Biology* 19(5):572-577.

Taylor, M. E., and K. Drickamer. 2009. Structural insights into what glycan arrays tell us about how glycan-binding proteins interact with their ligands. *Glycobiology* 19(11):1155-1162.

Teo, C. F., E. E. Wollaston-Hayden, and L. Wells. 2010. Hexosamine flux, the *O*-GlcNAc modification, and the development of insulin resistance in adipocytes. *Molecular and Cellular Endocrinology* 318(1-2):44-53.

Tillett, W. S., and T. Francis. 1929. Cutaneous reactions to the polysaccharides and proteins of pneumococcus in lobar pneumonia. *Journal of Experimental Medicine* 50(5):687-701.

Tilman, D., R. Socolow, J. A. Foley, J. Hill, E. Larson, L. Lynd, S. Pacala, J. Reilly, T. Searchinger, C. Somerville, and R. Williams. 2009. Beneficial biofuels—the food, energy, and environment trilemma. *Science* 325(5938):270-271.

Tumpey, T. M., T. R. Maines, N. Van Hoeven, L. Glaser, A. Solorzano, C. Pappas, N. J. Cox, D. E. Swayne, P. Palese, J. M. Katz, and A. Garcia-Sastre. 2007. A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science* 315(5812):655-659.

Van de Vyver S., J. Geboers, P. A. Jacobs, and B. F. Sels. 2011. Recent advances in the catalytic conversion of cellulose. *ChemCatChem* 3:82-94.

van Kooyk, Y., and G. A. Rabinovich. 2008. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nature Immunology* 9(6):593-601.

Varki, A. 1993. Biological roles of oligosaccharides—all of the theories are correct. *Glycobiology* 3(2):97-130.

Varki, A. 2006. Nothing in glycobiology makes sense, except in the light of evolution. *Cell* 126(5):841-845.

Varki, A. 2007. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* 446(7139):1023-1029.

Varki, A. 2011. Since there are PAMPs and DAMPs, there must be SAMPs? Glycan "self-associated molecular patterns" dampen innate immunity, but pathogens can mimic them. *Glycobiology* 21(9):1121-1124.

Varki, A., R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler. 2009. *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Vartiainen J., T. Pohler, K. Sirola, L. Pylkkanen, H. Alenius, J. Hokkinen, U. Tapper, P. Lahtinen, A. Kapanen, K. Putkisto, P. Hiekkataipale, P. Eronen, J. Ruokolainen, and A. Laukkanen 2011. Health and environmental safety aspects of fiction grinding and spray drying of microfibrillated cellulose. *Cellulose* 18:775-786.

Vasiliu, D., N. Razi, Y. Zhang, N. Jacobsen, K. Allin, X. Liu, J. Hoffmann, O. Bohorov, and O. Blixt. 2006. Large-scale chemoenzymatic synthesis of blood group and tumor-associated poly-*N*-acetyllactosamine antigens. *Carbohydrate Research* 341(10):1447-1457.

Voet, D., and J. G. Voet. 2010. *Biochemistry*, 4th Edition. John Wiley & Sons

Viswanathan, K., A. Chandrasekaran, A. Srinivasan, R. Raman, V. Sasisekharan, and R. Sasisekharan. 2010. Glycans as receptors for influenza pathogenesis. *Glycoconjugate Journal* 27(6):561-570.

Wagner R., R. J. Moon, and A. Raman. 2011. Uncertainty quantification in nanomechanical measurements using the atomic force microscope. *Nanotechnology* 22(45):455703.

Walker L. M., D. Sok, Y. Nishimura, O. Donau, R. Sadjadpour, R. Gautam, M. Shingai, R. Pejchal, A. Ramos, M. D. Simek, Y. Geng, I. A. Wilson, P. Poignard, M. A. Martin, and D. R. Burton. 2011. Rapid development of glycan-specific, broad, and potent anti-HIV-1 gp120 neutralizing antibodies in an R5 SIV/HIV chimeric virus infected macaque. *Proceedings of the National Academy of Sciences of the United States of America* 108(50):20125-20129.

Walsh, G., and R. Jefferis. 2006. Post-translational modifications in the context of therapeutic proteins. *Nature Biotechnology* 24(10):1241-1252.

Wang, C. C., J. C. Lee, S. Y. Luo, S. S. Kulkarni, Y. W. Huang, C. C. Lee, K. L. Chang, and S. C. Hung. 2007. Regioselective one-pot protection of carbohydrates. *Nature* 446(7138): 896-899.

Wang, A. M., C. de la Motte, M. Lauer, and V. Hascall. 2011. Hyaluronan matrices in pathobiological processes. *FEBS Journal* 278(9):1412-1418.

Wang, L.-X., and J. V. Lomino. 2012. Emerging technologies for making glycan-defined glycoproteins. *ACS Chemical Biology* 7(1):110-122.

Wetterstrand, K. A. 2012. DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program. Available at www.genome.gov/sequencingcosts [accessed August 16, 2012].

Werz, D. B., R. Ranzinger, S. Herget, A. Adibekian, C. W. von der Lieth, and P. H. Seeberger. 2007. Exploring the structural diversity of mammalian carbohydrates ("Glycospace") by statistical databank analysis. *ACS Chemical Biology* 2(10):685-691.

Williams, N. R., and J. D. Wander. 1980. Deoxy and branched-chain sugars. Pp. 761-798 in *The Carbohydrates, Chemistry and Biochemistry*. W. Pigman, D. Horton, and J. D. Wander, eds. New York: Academic Press.

Wu, C. Y., P. H. Liang, and C. H. Wong. 2009. New development of glycan arrays. *Organic & Biomolecular Chemistry* 7(11):2247-2254.

Xu, Y., S. Masuko, M. Takieddin, H. Xu, R. Liu, J. Jing, S. A. Mousa, R. J. Linhardt, and J. Liu. 2011. Chemoenzymatic synthesis of homogeneous ultralow molecular weight heparins. *Science* 334(6055):498-501.

Yamada, T. 2011. Therapeutic monoclonal antibodies. *The Keio Journal of Medicine* 60(2):37-46.

Yu, F., and J. H. Prestegard. 2006. Structural monitoring of oligosaccharides through 13C enrichment and NMR observation of acetyl groups. *Biophysical Journal* 91(5):1952-1959.

Yu, H., S. Huang, H. Chokhawala, M. Sun, H. Zheng, and X. Chen. 2006. Highly efficient chemoenzymatic synthesis of naturally occurring and non-natural alpha-2,6-linked sialosides: A *P. damsela* alpha-2,6-sialyltransferase with extremely flexible donor-substrate specificity. *Angewandte Chemie, International Edition* 45(24):3938-3944.

Yuzwa, S. A., X. Shan, M. S. Macauley, T. Clark, Y. Skorobogatko, K. Vosseller, and D. J. Vocadlo. 2012. Increasing *O*-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nature Chemical Biology* 8:393-399.

Zaia, J. 2004. Mass spectrometry of oligosaccharides. *Mass Spectrum Reviews* 23(3):161-227.

Zarbock, A., K. Ley, R. P. McEver, and A. Hidalgo. 2011. Leukocyte ligands for endothelial selectins: Specialized glycoconjugates that mediate rolling and signaling under flow. *Blood* 118(26):6743-6751.

Zhang, H., S. Singh, and V. N. Reinhold. 2005. Congruent strategies for carbohydrate sequencing. 2. FragLib: An MSn spectral library. *Analytical Chemistry* 77(19):6263-6270.

Zhang, N., L. Liu, C. D. Dumitru, R. N. Cummings, M. Cukan, Y. Jiang, Y. Li, F. Li, T. Mitchell, M. R. Mallem, Y. Ou, R. N. Patel, K. Vo, H. Wang, I. Burnina, B. K. Choi, H. E. Huber, T. A. Stadheim, and D. Zha. 2011. Glycoengineered Pichia produced anti-HER2 is comparable to trastuzumab in preclinical study. *mAbs* 3(3):289-298.

Zheng, T., H. Jiang, M. Gros, D. S. del Amo, S. Sundaram, G. Lauvau, F. Marlow, Y. Liu, P. Stanley, and P. Wu. 2011. Tracking *N*-acetyllactosamine on cell-surface glycans in vivo. *Angewandte Chemie, International Edition* 50(18):4113-4118.

Zhu, X. M., and R. R. Schmidt. 2009. New principles for glycoside-bond formation. *Angewandte Chemie, International Edition* 48(11):1900-1934.

Zijlstra, J. M., G. Lindauer-van der Werf, O. S. Hoekstra, L. Hooft, I. I. Riphagen, and P. C. Huijgens. 2006. 18F-fluoro-deoxyglucose positron emission tomography for post-treatment evaluation of malignant lymphoma: A systematic review. *Haematologica—The Hematology Journal* 91(4):522-529.

Appendix A

Committee Member Biographies

David R. Walt is Robinson Professor of Chemistry and professor of biomedical engineering at Tufts University and is a Howard Hughes Medical Institute professor. Dr. Walt served as chemistry department chairman at Tufts from 1989 to 1996. His laboratory applies micro- and nanotechnologies to urgent biological problems (such as the analysis of genetic variation and the behavior of single cells), single-molecule detection, and the practical application of arrays for diagnostics and the detection of explosives, chemical and biological warfare agents, and food and waterborne pathogens. Dr. Walt is the founding scientist of Illumina, Inc., and has been a director and chairman of its Scientific Advisory Board since 1998. He is also the founding scientist of Quanterix Corporation and is a director and the chairman of its Scientific Advisory Board. He serves on many government advisory panels and boards and on the editorial advisory board of numerous journals. He is a member of the Defense Sciences Research Council, a high-level advisory group for the U.S. Department of Defense and is a member of the Board on Chemical Sciences and Technology of the National Academy of Sciences. From 1996 to 2003 he served as executive editor of *Applied Biochemistry and Biotechnology*. Dr. Walt has published over 250 papers and holds more than 60 patents. He has received numerous national and international awards and honors for his fundamental and applied work in the field of optical sensors and arrays, including the American Chemical Society's 2010 National Award for Creative Invention. He is a member of the National Academy of Engineering, a fellow of the American Institute for Medical and Biological

Engineering, and a fellow of the American Association for the Advancement of Science. He received a B.S. in chemistry from the University of Michigan and a Ph.D. in chemical biology from the State University of New York at Stony Brook.

Kiyoko F. Aoki-Kinoshita received her B.S. and M.S. degrees in computer science from Northwestern University simultaneously in 1996. She received her doctorate in computer engineering from Northwestern in 1999. She was employed at BioDiscovery, Inc., in Los Angeles as a senior software engineer before moving to Kyoto, Japan, to work as a post-doctoral researcher at the Bioinformatics Center, Institute of Chemical Research, Kyoto University. There she developed various algorithmic and data-mining methods for analyzing glycan structure data accumulated in the KEGG GLYCAN database, which have been published in numerous journal papers. She then joined the Department of Bioinformatics in the Faculty of Engineering at Soka University in Tokyo and is now an associate professor of bioinformatics. She is also involved in several research projects pertaining to glycan functions based on their structure as well as recognition patterns of glycan structures by other proteins and even viruses. One of these projects is the development of a Web resource called RINGS (Resource for INformatics of Glycomes at Soka), which is intended to freely provide on the Internet many of the informatics algorithms and methods that have been published in the literature. These and other methods have been summarized in her book *Glycome Informatics: Methods and Applications* (CRC Press, 2009). Dr. Aoki-Kinoshita is a board member of the Japanese Society for Bioinformatics and the Japanese Society for Carbohydrate Research.

Brad Bendiak is an associate professor at the University of Colorado School of Medicine, where he teaches cell and developmental biology. He received his Ph.D. from the University of Cambridge in 1983. Dr. Bendiak's laboratory focuses on understanding the enzymes that synthesize cell surface carbohydrates, the glycosyltransferases. In addition, characterization of the carbohydrate structures themselves and development of new methods for elucidation of these molecules are ongoing. This includes new methods in higher-dimensional nuclear magnetic resonance (NMR) spectroscopy and fundamental studies in the fragmentation of carbohydrate molecules by mass spectrometry, with the overall goal being to assign the detailed structures of these complex molecules unambiguously. His laboratory is also interested in a series of glycosyltransferases involved in synthesis and branching of novel core structures of glycoprotein oligosaccharides and in better understanding the control of expression and the role of these enzymes in different tissues. For structural elu-

cidation of glycoprotein oligosaccharides, his laboratory uses high-field NMR, mass spectrometry, and gas chromatography-mass spectrometry, in addition to specific methods of chemical degradation that also are topics of research by the lab. Recent work has dealt with developments of gas-phase methods for separation and differentiation of oligosaccharide isomers.

Carolyn R. Bertozzi is the T. Z. and Irmgard Chu Distinguished Professor of Chemistry and a professor of molecular and cell biology at the University of California, Berkeley, an investigator at the Howard Hughes Medical Institute, and senior faculty scientist at the Lawrence Berkeley National Laboratory. She completed her undergraduate degree in chemistry from Harvard University in 1988 and her Ph.D. in chemistry from UC Berkeley in 1993. After completing postdoctoral work at the University of California, San Francisco, in the field of cellular immunology, she joined the UC Berkeley faculty in 1996. Dr. Bertozzi's research interests span the disciplines of chemistry and biology, with an emphasis on studies of cell surface glycosylation pertinent to disease states. Her laboratory focuses on developing chemical tools to probe changes in cell surface glycosylation associated with cancer, inflammation, and bacterial infection and on exploiting this information for development of diagnostic and therapeutic approaches. In addition, her group develops nanoscience-based technologies for probing cell function and methods for protein engineering. Dr. Bertozzi has been recognized with many honors and awards for both her research and teaching accomplishments. She is a member of the National Academy of Sciences, the Institute of Medicine, the American Academy of Arts and Sciences, and the German Academy of Sciences Leopoldina. Some awards of note include the Lemelson–Massachusetts Institute of Technology award for inventors, the Whistler Award, the Ernst Schering Prize, a MacArthur Foundation Fellowship, the American Chemical Society Award in Pure Chemistry, the Tetrahedron Young Investigator Award, and the Irving Sigal Young Investigator Award of the Protein Society. Her efforts in undergraduate education have earned her a UC Berkeley Distinguished Teaching Award and the Donald Sterling Noyce Prize for Excellence in Undergraduate Teaching.

Geert-Jan Boons received his M.Sc. in chemistry in 1987 and his Ph.D. in synthetic carbohydrate chemistry in 1991 from the State University of Leiden (The Netherlands). Prior to joining the faculty at the Complex Carbohydrate Research Center at the University of Georgia in 1998, he spent 7 years in the United Kingdom, first as a postdoctoral fellow at Imperial College and the University of Cambridge and then as a lecturer and professor at the University of Birmingham. In 2003, Dr. Boons was awarded

the Carbohydrate Research Award for Creativity in Carbohydrate Science by the European Carbohydrate Association. Also in 2003 he was elected chairman for the 2005 Gordon Research Conference on Carbohydrates. He serves on the editorial boards of the *Journal of Carbohydrate Chemistry*, *Advances in Carbohydrate Chemistry and Biochemistry*, *Glycoconjugate Journal*, and the *European Journal of Organic Chemistry*. In 2004, Dr. Boons received the Horace Isbell Award by the Division of Carbohydrate Chemistry of the American Chemical Society and was appointed Franklin Professor of Chemistry at the College of Arts and Sciences, University of Georgia. Research by the Boons Group deals with the synthesis and biological functions of carbohydrates and glycoconjugates. The diverse topics to which the group has made significant contributions include the development of new and better methods for synthesizing exceptionally complex molecules, the use of new methods in the synthesis and study of properties of complex carbohydrates of increasing size and complexity, the development of synthetic cancer and bacterial vaccines, the design and synthesis of glycosidase inhibitors, and the use of synthetic compounds for the study of innate immunity.

Alan Darvill received his B.S. in plant biology in 1973 from Wolverhampton Polytechnic (England) and his Ph.D. in plant physiology in 1976 from the University College of Wales, Aberystwyth. He founded the Complex Carbohydrate Research Center (CCRC), at the University of Georgia, with Peter Albersheim in September 1985. Dr. Darvill is currently director of the CCRC, director of the U.S. Department of Energy (DOE)–funded Center for Plant and Microbial Complex Carbohydrates, and the University of Georgia's lead in the DOE-funded BioEnergy Science Center. In 2003, Dr. Darvill was appointed Regents Professor of Biochemistry and Molecular Biology and became a senior faculty fellow. He was elected chairman for 1994-1995 of the Carbohydrate Division of the American Chemical Society and was appointed a member in 1993 and chairman in 1996 of the Martin Gibbs Medal Committee of the American Society of Plant Physiologists. He served on the editorial boards of *Glycobiology* and *Plant Journal for Cell and Molecular Biology*. Dr. Darvill received the Outstanding Faculty Award of the University of Georgia Chapter of the Golden Key National Honor Society in 1995 and in 2010 was named a fellow of the American Association for the Advancement of Science.

Gerald Hart is the DeLamar Professor and Director of Biological Chemistry at the Johns Hopkins University School of Medicine. He received his Ph.D. in developmental biology from Kansas State University in 1977. His laboratory studies the cross talk between dynamic GlcNAcylation and phosphorylation of nucleocytoplasmic proteins in signaling, transcrip-

tion, and cellular metabolism and the roles of abnormal GlcNAcylation in diabetes, neurodegenerative disease, and cancer (oncogene and tumor suppressor proteins, in particular). The laboratory is also focused on developing improved methods (e.g., mass spectrometry and site-specific antibodies) for the study of *O*-GlcNAc modification, some of which may have diagnostic value. The lab described a major new form of protein glycosylation (termed "*O*-GlcNAc") that is found in all multicellular organisms, including plants, animals, and viruses that infect them. A major research theme is to elucidate the biosynthesis, removal, attachment sites, and functions of this novel posttranslational modification. In 2010, Dr. Hart was named an honorary professor of Shanghai Medical College. From 2009 to 2011 he was president of the International Glycoconjugate Organization. He has received many honors and is a member of many scientific organizations. He was also the founding editor-in-chief of the journal *Glycobiology*.

Laura L. Kiessling received her B.S. in chemistry from the Massachusetts Institute of Technology and her Ph.D. in chemistry from Yale University. After carrying out postdoctoral training in chemical biology at the California Institute of Technology, she returned in 1991 to Wisconsin, where she was born, to begin her independent career at the University of Wisconsin-Madison. Currently, she is a Hilldale Professor in the Departments of Chemistry and Biochemistry and also the Laurens Anderson Professor of Biochemistry. She serves as director of the Keck Center for Chemical Genomics and as program director for the Chemistry-Biology Interface Predoctoral Training Program. Her interdisciplinary research interests focus on elucidating and exploiting the biological roles of oligosaccharides and oligosaccharide conjugates in biological systems. Some examples of her contributions include a new approach to inhibiting cell wall biosynthesis in Mycobacterium tuberculosis to devising sugar-binding surfaces to grow human embryonic stem cells. Dr. Kiessling serves on several editorial boards and is editor-in-chief of the American Chemical Society's *Chemical Biology*. She is a founder of Quintessence Biosciences, Inc., in Madison, Wisconsin. Her honors and awards include Guggenheim and MacArthur Foundation fellowships. She is a fellow of the American Academy of Arts and Sciences and a member of the National Academy of Sciences.

John Lowe joined Genentech, Inc., in 2008 as senior director of pathology. Previously, he was a faculty member at Washington University in St. Louis and the University of Michigan. Most recently, at Case Western Reserve University, Dr. Lowe was chair of a large pathology department whose missions included providing diagnostic laboratory and surgical pathol-

ogy services in support of a 985-bed tertiary care hospital and several other hospitals; educating medical school students, Ph.D. students, postdoctoral fellows, and pathology residents; and managing a comprehensive set of National Institutes of Health–funded basic research programs in immunity and neuroscience. His own research efforts prior to joining Genentech focused primarily on discovering functions for cell surface glycans in mammalian organisms, with particular relevance to the immune system. His role as senior director of pathology at Genentech includes opportunities to continue this research in an outstanding, disease-focused scientific environment while also leading the growth and development of scientific discovery and research support activities of the pathology department at Genentech. These efforts will help Genentech continue to make a major positive difference to the health and well-being of a large number of people who have cancer, autoimmune syndromes, neurodegenerative diseases, and other illnesses for which therapies are unsatisfactory or nonexistent.

Robert J. Moon is a materials research engineer with the Performance-Enhanced Biopolymers Group of the U.S. Forest Service Forest Products Laboratory in Madison, Wisconsin, and an adjunct associate professor at the School of Materials Engineering and a member of the Brick Nanotechnology Center at Purdue University. He received a B.S. in metallurgy from the University of Wisconsin (1994) and an M.S. (1996) and a Ph.D. (2000) in materials engineering from Purdue University. He completed his postdoctoral research (2000-2005) at the School of Materials Science and Engineering, University of New South Wales, Australia. His specialty is in processing-structure-property relationships of layered, gradient, and hierarchical structured materials and composites. In 2005, Dr. Moon joined the Forest Products Laboratory and in 2007 was selected by the laboratory to lead a collaborative research program with Purdue University that aims to advance nanoscale science and engineering of forestry-based materials. Dr. Moon has applied his expertise to the study of the role of hierarchical structures and interfaces on the mechanisms that dictate properties at the nano-, meso-, and macrolength scales of cellulose nanomaterials and their composites.

James C. Paulson is a professor in the Department of Chemical Physiology and the Department of Molecular Biology at Scripps Research Institute. He is also a principal investigator for the Consortium of Functional Glycomics; a member of the scientific advisory board for the Boston University Mass Spectrometry Resource; a co-chair of the Human Glycomics/Proteomics Initiative; and a scientific advisor to Nexbio, Institute for Biological Sciences, Neose Technologies, Inc., and the Alberta Ingenu-

ity Center for Carbohydrate Science. He is an honorary member of the American Society for Clinical Investigation; a member of the editorial board of *Glycobiology*; and a member of the American Chemical Society, the American Society of Biological Chemists, and the Society for Complex Carbohydrates. Before joining Scripps, Dr. Paulson worked at Cytel Corporation (1990-1999) and the University of California at Los Angeles School of Medicine (1978-1990). He received his Ph.D. in biochemistry from the University of Illinois, Urbana-Champagne, in 1974. He holds numerous patents and has published over 230 scientific papers. His current research focuses on the roles of glycan-binding proteins that mediate cellular processes central to immune regulation and human diseases. He works at the interface of biology and chemistry to understand how the interaction of glycan-binding proteins with their ligands mediates cell-cell interactions, endocytosis, and cell signaling.

Ram Sasisekharan is Alfred H. Caspary Professor of Biological Engineering in the David H. Koch Institute for Integrative Cancer Research at the Massachusetts Institute of Technology and a principal investigator in the Infectious Diseases Interdisciplinary Research Group of the SMART Centre in Singapore. In addition to developing analytical tools to study glycans, Dr. Sasisekharan's group was the first to conduct detailed studies of a class of glycan-degrading enzymes that were revealed to be critical tools to uncover fundamental biological roles of glycans in diseases including cancer, cardiovascular biology, and infectious diseases. Dr. Sasisekharan has published over 150 manuscripts and filed 70 United States patents and patent applications. He was a founder of Momenta Pharmaceuticals and served as a Director through September 2010. The company was founded in 2001 based on Dr. Sasisekharan's glycan sequencing platform, and the company has since leveraged this technology to produce the first biosimilar low molecular weight heparin. In 2005 he founded Cerulean Pharmaceuticals, which focuses on combination therapy using nanotechnology. In 2008 he founded Visterra Inc. with a focus on infectious diseases, and he currently serves on the board of this early-stage, venture-backed company. Dr. Sasisekharan is a consultant to and serves on the advisory boards of multiple biotechnology companies, venture funds, and nonprofit institutions involved in the translation of life-sciences innovation. He obtained his bachelor's degree in physical sciences from Bangalore University and Ph.D. in medical sciences from Harvard Medical School.

Ajit P. Varki received basic training in physiology, medicine, biology, and biochemistry at Christian Medical College (Vellore, India), the University of Nebraska, and Washington University in St. Louis. He also has formal training and certification in internal medicine, hematology, and oncology.

He is now distinguished professor of medicine and cellular and molecular medicine and co-director of the Glycobiology Research and Training Center at the University of California, San Diego. Dr. Varki is also executive editor of the textbook *Essentials of Glycobiology*. He is a founder and co-director of the UCSD Center for Academic Research and Training in Anthropogeny. He has served as chief editor of the *Journal of Clinical Investigation*. Dr. Varki is an elected member of the American Academy of Arts and Sciences, Institute of Medicine, American Society for Clinical Investigation, and Association of American Physicians. Dr. Varki has received a MERIT Award from the National Institutes of Health, an American Cancer Society Faculty Research Award, the Karl Meyer Award of the Society for Glycobiology, and the International Glycoconjugate Organization Award. He serves on the National Chimpanzee Observatory Working Group and on the editorial board of *Glycobiology*. He is a specialist advisor to the Human Gene Nomenclature Committee. His research interests currently focus on the family of sugar molecules called sialic acids and their roles in biology, evolution, and disease. Active projects are relevant to the roles of sialic acids in viral and bacterial infectivity, regulation of the immune response, initiation and progression of tumors, and unique aspects of human evolution. The lab is particularly intrigued to find multiple differences in sialic acid biology between humans and our closest evolutionary cousins, the great apes. These differences are a signature of the multiple cellular and molecular events that occurred during the past few million years of human evolution and are relevant to understanding several aspects of the current human condition, both in health and disease.

Chi-Huey Wong received his Ph.D. in chemistry from the Massachusetts Institute of Technology and completed a postdoctoral fellowship at Harvard University in 1983. Currently, Dr. Wong is president of Academia Sinica (Taipei, Taiwan) and a professor of chemistry at the Scripps Research Institute. He is a member of the National Academy of Sciences and the American Academy of Arts and Sciences. He has served on the National Research Council Board on Chemical Sciences and Technology and has held advisory positions in industry and academia. He has received more than 20 awards for his scientific work. His main research interests are in chemical biology and synthetic chemistry, including synthesis of complex carbohydrates, glycoproteins, and small-molecule probes for the study of posttranslational glycosylation and carbohydrate-mediated biological recognition.

Appendix B

The Landscape of Current Research in Glycoscience: Additional Information

As a starting point to inform its deliberations, the committee sought to better understand the current landscape of major U.S. and international glycoscience efforts. This appendix presents an overview of some of the research, funding, and industry initiatives the committee identified during its data-gathering process. It is not meant to be comprehensive but rather to provide a baseline of current investments in the field, including centers of research activity and funding, in this country and abroad. The examples provided are not meant to imply endorsement by the committee or the National Research Council (NRC).

B.1 AN OVERVIEW OF GLYCOSCIENCE IN THE UNITED STATES

B.1.1 Federal Agency Interests in Glycoscience

Glycoscience research in the United States is conducted within and supported by a number of federal departments and agencies, including the National Institutes of Health (NIH; through multiple individual institutes), the National Science Foundation (NSF), the U.S. Department of Energy (DOE), the Food and Drug Administration (FDA), the U.S. Department of Agriculture (USDA), and the National Institute of Standards and Technology (NIST).

NIH is one of the principal U.S. funding agencies for glycoscience research, reflecting the myriad ways in which carbohydrate glycans are linked to physiology and disease. Despite broad interest in the field, only

a relatively small percentage of the total NIH portfolio is devoted to specifically identified glycoscience awards. For example, "in an unofficial search of applications submitted to NIH using an R mechanism for [2010], 107 had glycomic in its abstract/summary statement and 400 had glycan" (McGowan and Bowman, 2010). A search of the NIHReporter system for fiscal year 2011 research grants having the term "glycan" in their project title or abstract produced 159 results (129 projects and 30 subprojects), with total funding of approximately $59 million.[1] The National Institute of General Medical Sciences (NIGMS) is a central contributor to many NIH glycoscience efforts, although investments exist for other institutes, such as the National Cancer Institute (NCI), the National Heart, Lung, and Blood Institute (NHLBI), and the National Institute of Allergy and Infectious Diseases (NIAID).

In addition to funding research individually, the institutes support collaborative initiatives. A long-term glycoscience special-interest group (GlycoSIG) brings together interested researchers from across NIH and FDA. While not meant to be an exhaustive list, key glycoscience programs within NIH include:

- support for the Consortium of Functional Glycomics (CFG), described in more detail below, through a Glue Grant mechanism that has now ended (NIGMS);
- continued support to enable researcher access to resources developed by the CFG through Legacy Community-Wide Scientific Resources funding (NIGMS);
- the Alliance of Glycobiologists for Detection of Cancer and Cancer Risk, currently in its second phase of funding for 2012-2017 (NCI);
- support of glycomics laboratories through the Early Detection Research Network (NCI);
- support of national centers for research resources (recently transferred to NIGMS) emphasizing glycoscience and glycomics;
- programs of excellence in glycoscience, supporting six awards for 2011-2018 and including a requirement for both glycoscience research and skills development components (NHLBI);

[1] Including "Research Project Grants (both SBIR/STTR and non-SBIR/STTR)" and "Other Research Related" but not including "Research Centers." The total funding of approximately $59 million broken down by agency included approximately $12 million at NIGMS, $16 million at NHLBI, $11 million at NIAID, and $8 million at NCI, with other (multiple) institutes making up the remainder (www.projectreporter.nih.gov; search conducted on June 7, 2012). A similar search for fiscal year 2012 projects produced 84 results (68 projects and 16 subprojects) with total funding of approximately $25 million. It should be noted that fiscal year 2012 was ongoing at the time of the search.

- Functional Glycomics in HIV Vaccine Design, an RO1 program anticipated to start in 2013 (NIAID); and
- SBIR contract mechanisms for such areas as production of specific monoclonal antibodies or analytical technologies (various).

Glycoscience research is also relevant to the mandates of NSF directorates and divisions, including biological sciences, chemistry, and materials research. Although NSF has not developed funding solicitations specifically dedicated to glycoscience, it receives relevant proposals under a range of programs. An approximate and unofficial portfolio estimate indicates that approximately 5 percent of chemical synthesis submissions address carbohydrate synthesis strategies, whereas approximately 5 to 10 percent of biomaterials submissions involve polysaccharides or sugars in some fashion (personal communication from Kelsey Cook, Tingyu Li, and David Berkowitz, NSF Division of Chemistry, and David Brant, NSF Division of Materials Research, phone conversation, 11/18/2010).

FDA and NIST both maintain interests in the development of measurement technologies and standards relevant to health care products and therapeutics. Research conducted by these agencies includes efforts to better understand physiological interactions and trafficking of glycosylated protein therapeutics and carbohydrate-based vaccines, methods to monitor glycosylation in cell cultures for biologics production, and characterization and quantification of glycans and glycoproteins. These agencies are also interested in the creation of reference standards and materials, such as standards for the interpretation of glycan mass spectral data.

Finally, DOE and USDA are engaged in glycoscience, particularly as it relates to nonmammalian systems. Glycoscience efforts at DOE emphasize areas related to biofuel and bioproduct development, such as cell wall chemistry, synthesis, and deconstruction with a particular focus on plants and on nonmedical aspects of microorganisms. Three bioenergy research centers are supported by the DOE Office of Biological and Environmental Research (BER); the DOE Basic Energy Sciences (BES) office supports a network of 46 Energy Frontier Research Centers, including the Center for Lignocellulose Structure and Formation at Pennsylvania State University. BER also supports the DOE-Michigan State University Plant Research Laboratory and provides funding to one center within the Complex Carbohydrate Research Center (CCRC) at the University of Georgia—the DOE Center for Plant and Microbial Complex Carbohydrates—while BER supports a component of the DOE BioEnergy Science Center housed within the CCRC (Greene, 2011). Within USDA, glycoscience efforts are focused on improvement to bioconversion processes and on the creation of value-added agricultural materials and products.

B.1.2 Collaborative Glycoscience Initiatives and Research Clusters

Two centers or consortia consistently mentioned as hubs that address multiple challenges in glycoscience research are the Consortium for Functional Glycomics (CFG) and the CCRC.

The CFG provides glycomics resources directed toward studying the impact of glycans and glycan-binding proteins on human health and disease (www.functionalglycomics.org). The consortium is guided by a steering committee that sets its scientific direction; it also contains two scientific cores (the Bioinformatics Core and the Protein-Glycan Interaction Core), along with hundreds of participating investigators located at institutions in this country and around the world organized into subgroups based on research interests. These groups include glycan synthesis and microarrays, glycans in immune recognition and function, glycans in development and physiology, structural glycobiology, glycans in cancer biology, and glycomics and glycoinformatics. The CFG maintains the functional glycomics gateway (http://www.functionalglycomics.org) and acts as the central hub and clearinghouse for glycomics, offering resources that include glycomics profiling, carbohydrate compounds and reagents, microarray analysis, mouse phenotyping, glycan array screening, and databases to investigators whose projects are approved for scientific relevance by the consortium's steering committee. Although NIH funding to support the CFG through its Glue Grant mechanism has now ended, the consortium continues and mechanisms have been established to enable ongoing access to the resources and tools it has developed.

The CCRC, housed at the University of Georgia, seeks to foster "cooperation and collaboration among disciplines (biomedical, plant, and microbial glycosciences, synthetic and analytical chemistry) both within the CCRC and with scientists worldwide and to foster analytical service and training" in carbohydrate research (http://www.ccrc.uga.edu). It is funded by a combination of federal, state, industry, foundation, and research funds and is made up of a cluster of centers each addressing different aspects of carbohydrate research. These include the Center for Plant and Microbial Complex Carbohydrates (funded by DOE's BES); Integrated Technology Resource for Biomedical Glycomics (funded by NIH); Research Resource for Integrated Glycotechnology (funded by NIH); Southeast Collaboratory for Biomolecular NMR (funded by NIGMS and the Georgia Research Alliance; research initiatives on GLYCAM/AMBER Modeling Tools for Glycoscience; Monoclonal Antibodies for Plant Cell Walls (funded by NSF); and Plant Cell Walls and Biomass Recalcitrance (both components of the Bioenergy Science Center [BESC], funded by the DOE Office of Biological and Environmental Research).

In addition to the two multidisciplinary organizations described above, a variety of other U.S. research programs and clusters of research

expertise were noted by respondents who provided input through the committee's website or were identified during other data-gathering efforts. Several of these are listed here to help underscore that research is occurring nationwide. It should be emphasized that the list below (in alphabetical order by university and/or geographical "cluster") does *not* reflect the only places in which important and useful glycoscience research is under way:

- Baltimore-Washington cluster, including Johns Hopkins University, University of Maryland, and intramural research at NIH, such as the Carbohydrate Section of the National Institute of Diabetes and Digestive and Kidney Diseases;
- Cornell University;
- Emory University School of Medicine;
- Harvard University;
- Memorial Sloan-Kettering Cancer Center;
- National Center for Glycomics and Glycoproteomics, Indiana University;
- National Renewable Energy Laboratory;
- C3Bio, Purdue University;
- San Diego, California, cluster, including the Glycobiology Research and Training Center; University of California, San Diego, which incorporates researchers at Salk, Scripps, and Sanford-Burnham in La Jolla; and glycoscientists elsewhere throughout California;
- University of California, Davis;
- Glycomics Center, University of New Hampshire;
- USDA-ARS Laboratory and USDA-Forest Service Forest Products Laboratory; and
- Virginia Polytechnic Institute and State University.

B.2 AN OVERVIEW OF GLYCOSCIENCE OUTSIDE THE UNITED STATES

Active clusters of glycoscience research are located around the world, and a few of these regions are highlighted below. Different research organizations or even different countries may have greater expertise in certain aspects of glycoscience, such as medical applications, biofuels and biomass, mammalian glycoscience, bacterial glycoscience, or enzyme research. The data gathering conducted by the committee was intended to be fairly broad in scope in order to identify centers of expertise beyond the intersection of glycoscience with human health. Again, the list below should not be considered comprehensive; it is meant to provide a sense of major ongoing efforts.

B.2.1 Canada

Glycoscience programs in Canada include the Alberta Glycomics Centre (www.glycomicscentre.ca), a partnership between the University of Alberta and the University of Calgary that conducts multidisciplinary research in glycoscience and carbohydrate chemistry, such as the development of analytical methods, structural studies, and research on carbohydrate therapeutics. Other ongoing efforts in Canada include those at the University of Toronto and the National Research Council Canada Institute for Biological Sciences in Ottawa.

B.2.2 United Kingdom and Europe

Within the United Kingdom, glycoscience research is supported by such organizations as the Wellcome Trust, the Engineering and Physical Sciences Research Council, the Biotechnology and Biological Sciences Research Council, and the Medical Research Council; centers of activity include Imperial College London and the University of Oxford and University of Dundee. Collaborative efforts include the UK Glycoarrays Consortium (http://www.glycoarrays.org.uk), which draws on scientists from the universities of Dundee, Liverpool, Manchester, Oxford, East Anglia, and Imperial College London to develop carbohydrate microarrays.

In Europe glycoscience research is supported by both national funding organizations such as the Deutsch Forschungsgemeinschaft in Germany and through programs such as the European Commission (EC). One current EC-supported initiative is the Euroglycosciences Forum (http://www.egsf.org), supported through the European Science Foundation Research Network Programme for 2009-2014. The forum seeks to provide access to glycoscience tools and supports meetings and workshops to link members of the glycoscience community. A variety of database and bioinformatics resources are housed in Europe, including EuroCarbDB (http://www.eurocarbdb.org), funded under the 6th Research Framework Program of the European Union, and subsequent related efforts such as UniCarb-DB (http://unicarb-db.biomedicine.gu.se), supported as a partnership by the University of Gothenburg (Sweden), Macquarie University, Swiss Institute for Bioinformatics, and the National Institute for Bioprocessing Research and Training (Ireland), along with support from the Swedish Foundation for International Cooperation in Research and Higher Education and the Australian Research Council. Other resources include Glycosciences.de (http://www.glycosciences.de); the CASPER tool for analysis of nuclear magnetic resonance spectra developed at Stockholm University; the GlycoWorkbench suite of tools for mass spectrometry-based glycomics (http://www.glycoworkbench.org) developed at Imperial Col-

lege London; the Carbohydrate-Active EnZymes database (http://www.cazy.org), developed by CNRS and Aix-Marseille University (France); the EuroGlycoArrays training network (http://www.euroglycoarrays.eu), supported under the Seventh Framework Programme of the EC; and the Euroglycanet network for congenital disorders of glycosylation (http://www.euroglycanet.org).

A range of additional research centers and programs were noted by participants who provided input through the committee's website or were identified through the Web of Science citation search. In Europe these included:

- Carlsberg Laboratory, Denmark, which researches barley, yeast, brewing, and fermenting technology;
- Centre de Recherches sur les Macromolecules Vegetales of the Centre National de la Recherche Scientifique (CNRS), France;
- Ghent University, Belgium;
- Institut d'Investigació Biomèdica de Bellvitge, Spain;
- Institut National de la Recherche Agronomique, France; and
- Carbohydrate Competence Center, The Netherlands.

B.2.3 Asia

Asia is a center of glycoscience research, and there is significant and emerging interest in such countries as Japan, South Korea, and China.

In Japan glycoscience research is supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the Japan Society for the Promotion of Science, as well as by such organizations as the Mitzutani Foundation for Glycoscience. A recently established initiative is the Japan Consortium for GlycoBiology and Glycotechnology (http://www.jcgg.jp/index_e.html), supported by MEXT, which has developed the JCGGDB database to house data, including the GlycoGeneDataBase, the Glycan Mass Spectral DataBase, the Lectin Frontier DataBase, and the Glycoprotein Database. MEXT is also supporting a neuroglycobiology project (2011-2015) with the goal of deciphering sugar chain–based signals that regulate integrative neuronal functions. Glycoscience research centers in Japan include the National Institute of Advanced Industrial Science and Technology, Kyoto University, University of Tokyo, and RIKEN, which includes the RIKEN Omics Science Center, the RIKEN Bioinformatics and Systems Engineering Division, and the System Glycobiology Research Group. Through a program on exploratory research for advanced technology, RIKEN is leading a 5-year glycotrilogy project linking molecular biology, synthetic chemistry, and analytical chemistry to study glycans and glycan derivatives. RIKEN recently partnered with the Max Planck Insti-

tute of Colloids and Interfaces in Germany to establish a joint research center in systems chemical biology on areas such as disease glycomics and oligosaccharide synthesis.

In China glycoscience research is funded under the National Basic Research Program of the Ministry of Science and Technology and by the National Natural Science Foundation of China, which manages the National Natural Science Fund and promotes basic and applied research. Active glycoscience research is carried out in laboratories within the Chinese Academy of Sciences.

Glycoscience research is also conducted in Taiwan—for example, at the Academia Sinica and at several national universities. The Taiwan GlycoForum (https://sites.google.com/site/taiwanglycoforum/home) was established in 2011 to provide a mechanism for communication and collaboration.

The Australia and New Zealand Glycosciences Group seeks to provide a forum for the glycoscience community in these countries. Well-known programs include the Institute for Glycomics at Griffith University (http://www.griffith.edu.au/science-aviation/institute-glycomics), which focuses on understanding the roles of glycans in disease through investigations of both mammalian and microbial glycomics and on development of enabling technologies such as new synthesis and analytical tools. Glycomics research is also conducted at Macquarie University in Sydney, and established programs that analyze plant cell walls include those at the University of Melbourne and the ARC Centre of Excellence in Plant Cell Walls at the University of Adelaide (http://www.adelaide.edu.au/plant-cell-walls).

B.2.4 Latin America

Glycoscience research is supported in Brazil by the Conselho Nacional de Desenvolvimento Científico e Tecnológico, the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, and the Fundação de Amparo à Pesquisa do Estado de São Paulo. Respondents to the committee's website included scientists from Chile and Mexico, indicating ongoing research efforts in these countries as well.

The international nature of glycoscience research and the existence of significant centers of excellence outside the United States are important factors to keep in view as the U.S. research community and federal agencies look toward the future development and transformative potential of the field. Key challenges to advance glycoscience may be shared, and the development of many types of resources, such as common standards for interpretation and sharing of glycan structural data, will necessarily need to be achieved within the framework of a global community to be truly effective.

B.3 INDUSTRY INTEREST IN GLYCOBIOLOGY

This very brief summary is not meant to capture the totality of industry interest across the field but rather to highlight just a few of the existing investments in research, development, and application of glycoscience. Mention of specific companies is made solely for illustrative purposes and reflects information obtained during the committee's data gathering or from committee members' knowledge of the field. Such mention does not in any way imply committee, NRC, or study sponsor endorsement of any commercial product or service.[2]

Companies with interests in therapeutic glycoproteins include Amgen (e.g., erythropoietin) and Genentech (e.g., antibodies and the antiinfluenza drug Tamiflu, originally involving Gilead Sciences). Genzyme's drug Cerezyme, approved in 1994, provides enzyme replacement for the treatment of Type 1 Gaucher disease and uses carbohydrate targeting. Other companies such as GlaxoSmithKline (manufacturer of the antiinfluenza drug Relenza, originally developed in part with Biota Holdings), Baxter (with its drug heparin), and Wyeth (now Pfizer, for glycoconjugate vaccines) have major programs in glycoscience or aspects of glycoscience. GlycoFi, established in 2000 to develop yeast-based production systems for protein-based drugs, was acquired by Merck in 2006. Novo Nordisk conducts clinical trials for the hemophilia treatment GlycoPEGylated factor IX, which contains polyethylene glycol groups linked to glycan chains through carbohydrate engineering to prolong the pharmacokinetics. Smaller companies active in aspects of glycoscience include GlycoMimetics, Inc. (focused on compounds that target cellular adhesion molecules such as selectins); ProtAffin (development of glycan-binding pharmaceuticals); GlycoMira Therapeutics (development of heparin derivatives); Sialix, Inc. (focused on pathological effects and a nonhuman sialic acid); Momenta (heparin and heparan sulfate); Biomarin (carbohydrate-based targeting for enzyme replacement therapy); Hyalose (tissue engineering and therapeutics based on hyaluronic acid), Zacharon (therapeutics and diagnostics for rare storage diseases); Ancora (carbohydrate-based vaccines and custom synthesis); and Selexys Pharmaceuticals (antibodies to P-selectin and its ligand). International companies include Dextra (synthetic oligosaccharides) and Seikagaku (oligosaccharides and enzymes). Companies in the energy industry, including large corporations such as BP and ExxonMobil, are interested in cellulose-based biofuels, an area that will rely on glycoscience as it develops.

[2]Several members of the committee are associated with companies or have received funding from companies identified above, including Genentech, Inc. (Bertozzi; Lowe); GlaxoSmithKline (Bertozzi); GlycoMimetics, Inc. (Bertozzi); and Sialix, Inc. (Varki).

Appendix C

Workshop on the Future of Glycoscience: Agenda and Participants

AGENDA

Thursday, January 12, 2012

8:00–8:20 am **Plenary: Workshop Opening**

- Welcome and introduction to the goals and context of the workshop
  David Walt, Tufts University

8:20–9:50 am **Plenary: Challenges and Opportunities for Glycoscience in Human Health**

Chair: Ajit Varki, University of California, San Diego

- Hudson Freeze, Sanford Burnham Medical Research Institute
- Pamela Stanley, Albert Einstein College of Medicine
- Robert Sackstein, Harvard University
- John Magnani, GlycoMimetics, Inc.
- Michael Rosenblatt, Merck & Co., Inc.

9:50–10:10 am **Break**—*Move to breakout sessions*

10:10–11:40 am **Breakout Sessions: Health**

11:50–12:30 pm **Plenary: Feedback from the Breakout Sessions**

12:30–1:30 pm	**Lunch Talk** *(20-minute lecture begins at 1:00 pm)* • George Whitesides, Harvard University
1:30–2:15 pm	**Plenary: Challenges and Opportunities for Glycoscience in Materials Science** Chair: Robert Moon, U.S. Forest Service • Nathan Mosier, Purdue University • Orlando Rojas, North Carolina State University • John Simonsen, Oregon State University
2:15–3:00 pm	**Plenary: Challenges and Opportunities for Glycoscience in Energy** Chair: Alan Darvill, University of Georgia • Lee Lynd, Dartmouth University (*speaking remotely*) • Markus Pauly, University of California, Berkeley • Michael Himmel, National Renewable Energy Laboratory
3:00–3:20 pm	**Break**—*Move to breakout sessions*
3:20–4:50 pm	**Breakout Sessions: Materials Science and Energy**
5:00–6:15 pm	**Plenary: Feedback from the Breakout Sessions and Day 1 Wrap-up**
6:15 pm	**Adjourn Day 1**

Friday, January 13, 2012

8:00–8:10 am **Plenary: Day 2 Welcome**—*David Walt, Tufts University*

8:10–10:10 am **Plenary: Overview Talks on Key Scientific Challenges**
Chair: James Paulson, The Scripps Research Institute

- Synthesis
    - Peter Seeberger, Max Planck Institute of Colloids and Interfaces
    - Robert Linhardt, Rensselaer Polytechnic Institute
- Chemical Analysis
    - Milos Novotny, Indiana University
    - Norman Dovichi, University of Notre Dame
- Biological Analysis
    - Richard Cummings, Emory University
    - Jeffrey Esko, University of California, San Diego
- Informatics and Databases
    - Nicolle Packer, Macquarie University, Australia
    - William York, Complex Carbohydrate Research Center, University of Georgia

10:10–10:30 am **Break**—*Move to breakout sessions*

10:30–12:00 pm **Breakout Discussion Sessions**

- Group #1: Synthesis
- Group #2: Chemical Analysis
- Group #3: Biological Analysis
- Group #4: Informatics and Databases

12:00–1:00 pm **Plenary: Feedback from the Breakout Sessions**

1:00 pm **Adjourn**

PARTICIPANTS

Kiyoko Aoki-Kinoshita
Department of Bioinformatics
Soka University

William Balch
Department of Cell Biology
The Scripps Research Institute

Roger Beachy
Department of Biology
Washington University in St. Louis and Donald Danforth Plant Science Center

Brad Bendiak
Department of Cell and Development Biology
University of Colorado, Denver

Carolyn Bertozzi
Department of Chemistry and Department of Molecular and Cell Biology
University of California, Berkeley

Bruce Bochner
Department of Medicine
Johns Hopkins School of Medicine

Geert-Jan Boons
Complex Carbohydrate Research Center
University of Georgia

Wendy Boss
Division of Molecular and Cellular Biosciences
National Science Foundation

Ralph Bradshaw
Department of Pharmaceutical Chemistry
University of California, San Francisco

David Brant
Division of Materials Research
National Science Foundation

Penny Wung Burgoon
National Institute of Dental and Craniofacial Research
National Institutes of Health

Kevin Campbell
Department of Molecular Physiology and Biophysics
University of Iowa

Russell Carlson
Complex Carbohydrate Research Center
University of Georgia

John Cipollo
Center for Biologics Evaluation and Research
Food and Drug Administration

Catherine Costello
Mass Spectrometry Resource
Boston University School of Medicine

Richard Cummings
Department of Biochemistry
Emory University

Alan Darvill
Complex Carbohydrate Research Center
University of Georgia

Norman Dovichi
Department of Chemistry and Biochemistry
University of Notre Dame

Stephen Eichhorn
Department of Physics
University of Exeter

Jeffrey Esko
Department of Cellular and Molecular Medicine
University of California, San Diego

Daron Freedberg
Center for Biologics Evaluation and Research
Food and Drug Administration

Hudson Freeze
Stanford Burnham Medical Research Institute

Gilman Grave
Eunice Kennedy Shriver National Institute of Child Health and Human Development
National Institutes of Health

Derek Gray
Department of Chemistry
McGill University

Robert Haltiwanger
Department of Biochemistry and Cell Biology
Stony Brook University

Gerald Hart
Department of Biological Chemistry
Johns Hopkins School of Medicine

Herbert Hill
Department of Chemistry
Washington State University

Michael Himmel
Energy Sciences Directorate
National Renewable Energy Laboratory

William Jacobs
Department of Microbiology and Immunology and Department of Genetics
Albert Einstein College of Medicine

Kenneth Keegstra
Department of Biochemistry and Molecular Biology
Michigan State University

Stephen Kelley
Department of Forest Biomaterials
North Carolina State University

Laura Kiessling
Department of Chemistry and Department of Biochemistry
University of Wisconsin-Madison

Amy Krafft
National Institute of Allergy and Infectious Diseases
National Institutes of Health

Donna Krasnewich
National Institute of General Medical Sciences
National Institutes of Health

Karl Krueger
National Cancer Institute
National Institutes of Health

Tingyu Li
Division of Chemistry
National Science Foundation

Robert Linhardt
Department of Chemistry and Chemical Biology
Rensselaer Polytechnic Institute

David Lipman
National Library of Medicine
National Institutes of Health

Lee Lynd
School of Engineering
Dartmouth College

John Magnani
GlycoMimetics, Inc.

Pamela Marino
National Institute of General Medical Sciences
National Institutes of Health

Debra Mohnen
Complex Carbohydrate Research Center
University of Georgia

Robert Moon
Forest Products Laboratory
U.S. Forest Service

Nathan Mosier
Department of Agricultural and Biological Engineering
Purdue University

Milos Novotny
Department of Chemistry
Indiana University

Thandi Onami
National Institute of Allergy and Infectious Diseases
National Institutes of Health

Nicolle Packer
Department of Chemistry and Biomolecular Science
Macquarie University

James Paulson
Department of Chemical Physiology and Department of Molecular Biology
The Scripps Research Institute

Markus Pauly
Department of Plant and Microbial Biology
University of California, Berkeley

Mercy PrabhuDas
National Institute of Allergy and Infectious Diseases
National Institutes of Health

Susan Pucie
National Heart, Lung, and Blood Institute
National Institutes of Health

Arthur Ragauskas
Department of Chemistry and Biochemistry
Georgia Institute of Technology

Richard Roberts
New England Biolabs, Inc.

David Rockcliffe
Division of Molecular and Cellular Biosciences
National Science Foundation

Michael Rogers
National Institute of General Medical Sciences
National Institutes of Health

Orlando Rojas
Department of Forest Biomaterials
North Carolina State University

Maren Roman
Department of Wood Science and Forest Products
Virginia Polytechnic Institute and State University

Michael Rosenblatt
Merck & Co., Inc.

Robert Sackstein
Department of Medicine and Department of Dermatology
Harvard Medical School

Neeraja Sathyamoorthy
National Cancer Institute
National Institutes of Health

Henrik Scheller
Feedstocks Division
Joint Bioenergy Institute

Ronald Schnaar
Department of Pharmacology and Molecular Sciences and Department of Neuroscience
Johns Hopkins School of Medicine

Peter Seeberger
Max Planck Institute of Colloids and Interfaces

Douglas Sheeley
National Center for Research Resources
National Institutes of Health

John Simonsen
College of Forestry
Oregon State University

Sudhir Srivastava
National Cancer Institute
National Institutes of Health

Robert Stack
Office of Basic Energy Sciences
U.S. Department of Energy

Pamela Stanley
Department of Cell Biology
Albert Einstein College of Medicine

Michael Tarlov
Material Measurement Laboratory
National Institute of Standards and Technology

Jon Thorson
School of Pharmacy
University of Wisconsin, Madison

Ajit Varki
Department of Medicine and Department of Cellular and Molecular Medicine
University of California, San Diego

David Walt
Department of Chemistry
Tufts University

Lance Wells
Complex Carbohydrate Research Center
University of Georgia

George Whitesides
Department of Chemistry
Harvard University

Carolyn Wilson
Center for Biologics Evaluation and Research
Food and Drug Administration

Ian Wilson
Department of Molecular Biology and Skaggs Institute for Chemical Biology
The Scripps Research Institute

William Winter
Department of Chemistry
State University of New York College of Environmental Science and Forestry

Chi-Huey Wong
President, Academia Sinica
Department of Chemistry
The Scripps Research Institute

Yu Xia
Department of Chemistry
Purdue University

William York
Complex Carbohydrate Research Center
University of Georgia

Hui Zhang
Department of Pathology
Johns Hopkins University

National Research Council Staff

Dorothy Zolandz
Director
Board on Chemical Sciences and Technology

Frances Sharples
Director
Board on Life Sciences

Katherine Bowman
Senior Program Officer
Board on Life Sciences

Douglas Friedman
Program Officer
Board on Chemical Sciences and Technology

India Hook-Barnard
Senior Program Officer
Board on Life Sciences

Sheena Siddiqui
Senior Program Associate
Board on Chemical Sciences and Technology

Rachel Yancey
Senior Program Assistant
Board on Chemical Sciences and Technology

Joe Alper
Consulting Writer

Appendix D

Input Received Online and Through Other Data Gathering

A website for the committee's study (http://glyco.nas.edu) invited members of the community to provide input by addressing several questions related to the committee's Statement of Task. These questions are listed below:

1. What do you view as the most significant opportunities for glycoscience and glycomics to forge new roads of discovery, particularly opportunities that build on advances made in other fields (e.g., genomics and proteomics) and/or opportunities for glycoscience knowledge to significantly transform other areas of biology and chemistry?
2. What do you view as key challenges to growth and development of the field of glycoscience?
3. What research or technological achievements are necessary to significantly advance glycoscience and glycomics?
4. Are there other significant research barriers or roadblocks that must be overcome?
5. Are there particularly noteworthy research centers, programs, or investments (in the United States or internationally) that the committee should be aware of as it examines the baseline of current glycoscience research?

Responses to the questions were received from 115 people in 16 countries.

The committee also held a data-gathering session at the annual meeting of the Society for Glycobiology on November 11, 2011, in Seattle, Washington, and conducted several additional data-gathering teleconferences with Sabine Flitsch, University of Manchester, UK; Anne Dell, Imperial College London, UK; Todd Lowary, University of Alberta, Canada; Bernard Henrisaat, CNRS, France; Jim Richards, NRC Ottawa, Canada; and Naoyuki Taniguchi, Japan.

Appendix E

Glossary[1]

Amino sugar: A monosaccharide in which an alcoholic hydroxyl group is replaced by an amino group.

Anomeric carbon: The carbon atom of a monosaccharide that bears the hemiacetal functionality (C-1 for most sugars; C-2 for sialic acids).

Anomers: Stereoisomers of a monosaccharide that differ only in configuration at the anomeric carbon of the ring structure.

Carbohydrate: Generic term used interchangeably in this report with sugar, saccharide, or glycan. This term includes monosaccharides, oligosaccharides, and polysaccharides as well as derivatives of these compounds.

Carbohydrate recognition domain: The domain of a polypeptide that is specifically involved in binding to a carbohydrate. In lectins it is often a highly evolutionarily conserved region of the polypeptide.

Cellulose: A repeating homopolymer of β1–4-linked glucose residues.

Chemoenzymatic synthesis: Glycan synthesis that uses both chemical and enzymatic transformations to obtain the desired product.

Chitin: A repeating homopolymer of β1–4-linked *N*-acetyl-glucosamine residues. It is the main component of the cell walls of fungi and the exoskeletons of arthopods, among other functions.

[1] Adapted from *Essentials of Glycobiology*, 2nd ed., A. Varki et al., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2009; used with permission.

Complex glycan: A glycan containing more than one type of monosaccharide.

Deoxy sugar: A monosaccharide in which a hydroxyl group is replaced by a hydrogen atom.

Epimers: Two isomeric monosaccharides differing only in the configuration of a single chiral carbon. For example, mannose is the C-2 epimer of glucose.

Furanose: Five-membered (four carbons and one oxygen; i.e., an oxygen heterocycle) ring form of a monosaccharide named after the structural similarity to the compound furan.

Galectins: S-type (sulfhydryl-dependent) β-galactoside-binding lectins, usually occurring in a soluble form, expressed by a wide variety of animal cell types and distinguishable by the amino acid sequence of their carbohydrate recognition domains.

Genome: The complete genetic sequence of one set of chromosomes.

Glycan: Generic term for any sugar or assembly of sugars, in free form or attached to another molecule, used interchangeably in this report with saccharide or carbohydrate.

Glycan array: A collection of glycans attached to a surface in a spatially addressed manner.

Glycan-binding proteins: Proteins that recognize and bind to specific glycans and mediate their biological function.

Glycobiology: Study of the structure, chemistry, biosynthesis, and biological functions of glycans and their derivatives.

Glycocalyx: The cell coat consisting of glycans and glycoconjugates surrounding animal cells that is seen as an electron-dense layer by electron microscopy.

Glycoconjugate: A molecule in which one or more glycan units are covalently linked to a noncarbohydrate entity.

Glycoforms: Different molecular forms of a glycoprotein, resulting from variable glycan structure and/or glycan attachment site occupancy.

Glycogen: A polysaccharide comprising α1–4- and α1–6-linked glucose residues that functions in short-term energy storage in animals; sometimes referred to as animal starch.

Glycolipid: General term denoting a molecule containing a glycan linked to a lipid aglycone. In higher organisms most glycolipids are glycosphingolipids, but glycoglycerolipids and other types exist.

Glycome: The total collection of glycans synthesized by a cell, a tissue, or an organism under specified conditions of time, space, and environment.

Glycomics: Systematic analysis of the glycome.

Glycomimetics: Noncarbohydrate compounds that mimic the properties of glycans.

Glycopeptide: A peptide having one or more covalently attached glycans.

Glycoprotein: A protein with one or more covalently bound glycans.

Glycoproteomics: The systems-level analysis of glycoproteins, including their protein identities, sites of glycosylation, and glycan structures.

Glycosaminoglycans: Polysaccharide side chains of proteoglycans or free complex polysaccharides composed of linear disaccharide repeating units, each composed of a hexosamine and a hexose or a hexuronic acid.

Glycosidase: An enzyme that catalyzes the hydrolysis of glycosidic bonds in a glycan.

Glycoside: A glycan containing at least one glycosidic linkage to another glycan or an aglycone.

Glycosidic linkage: Linkage of a monosaccharide to another residue via the anomeric hydroxyl group. The linkage generally results from the reaction of a hemiacetal with an alcohol (e.g., a hydroxyl group on another monosaccharide or amino acid) to form an acetal. Glycosidic linkages between two monosaccharides have defined regiochemistry and stereochemistry.

Glycosyl acceptor: The nucleophile in a glycosylation reaction, usually containing a free hydroxyl group.

Glycosylation: The enzyme-catalyzed covalent attachment of a carbohydrate to a polypeptide, lipid, polynucleotide, carbohydrate, or other organic compound, generally catalyzed by glycosyltransferases, utilizing specific sugar nucleotide donor substrates.

Glycosyl donor: The electrophile in a glycosylation reaction; the nucleotide sugar in an enzymatic glycosylation reaction.

Glycosyltransferase: The enzyme that catalyzes transfer of a sugar from a sugar nucleotide donor to a substrate.

Heparan sulfate: A glycosaminoglycan defined by the disaccharide unit (GlcNAcα1–4GlcAβ1–4/IdoAα1–4), containing *N*- and *O*-sulfate esters at various positions, and typically found covalently linked to a proteoglycan core protein.

Heparin: A type of heparan sulfate made by mast cells that has the highest amount of iduronic acid and *N*- and *O*-sulfate residues. Pharmaceutical heparin binds and activates antithrombin.

Heteropolysaccharide: A polysaccharide containing more than one type of monosaccharide.

Hexosamine: Hexose with an amino group in place of the hydroxyl group at the C-2 position. Common examples found in vertebrate glycans are the *N*-acetylated sugars, *N*-acetylglucosamine, and *N*-acetylgalactosamine.

Hexose: A six-carbon monosaccharide typically with an aldehyde (or potential aldehyde) at the C-1 position (aldo-hexose) and hydroxyl groups at all other positions. Common examples in vertebrate glycans are mannose, glucose, and galactose.

Homopolysaccharide: A polysaccharide composed of only one type of monosaccharide.

Lectin: A protein (other than a glycan-specific antibody) that specifically recognizes and binds to glycans without catalyzing a modification of the glycan.

Ligand: A molecule that is recognized by a specific receptor. In the case of lectins the ligands are partly or completely glycan based and are sometimes called counterreceptors.

Methylation analysis: A method for carbohydrate structure analysis based on the acid stability of methyl ethers and the acid lability of glycosidic linkages; used to determine the linkage positions of monosaccharide residues in an oligosaccharide chain.

Microheterogeneity: Structural variations in a glycan at any given glycosylation site on a protein (one source of glycoforms).

Monosaccharide: A carbohydrate that cannot be hydrolyzed into a simpler carbohydrate. It is the building block of oligosaccharides and polysaccharides. Simple monosaccharides are polyhydroxyaldehydes or polyhydroxyketones with three or more carbon atoms.

N-Glycan: A glycan covalently linked to an asparagine residue of a polypeptide chain in the consensus sequence: -Asn-X-Ser/Thr. Unless otherwise stated, the term *N*-glycan is used generically in this report to denote the most common linkage region, Manβ1–4GlcNAcβ1–4GlcNAcβ1-N-Asn.

Nucleotide sugars: Activated forms of monosaccharides, such as UDP-Gal, GDP-Fuc, and CMP-Sia, typically used as donor substrates by glycosyltransferases.

O-Glycan: A glycan glycosidically linked to the hydroxyl group of the amino acids serine, threonine, tyrosine, or hydroxylysine. Unless otherwise stated, the term *O*-glycan is used in this report to denote the common linkage GalNAcα1-O-Ser/Thr.

Oligosaccharide: A linear or branched chain of monosaccharides attached to one another via glycosidic linkages. The number of monosaccharide units can vary; the term polysaccharide is usually reserved for large glycans with repeating units.

Polysaccharide: A glycan composed of repeating monosaccharides, generally greater than 10 monosaccharide units in length.

Protecting group: A chemical moiety commonly used in glycan synthesis that masks hydroxyl groups in order to prevent them from reacting with other chemical reagents.

Proteome: The total collection of proteins in a cell, tissue, or organism, under specific conditions of time, space, and environment.

Pyranose: Six-membered (five carbons and one oxygen; i.e., an oxygen heterocycle) ring form of a monosaccharide; the most common form found for hexoses and pentoses. The name is based on the structural similarity to the compound "pyran."

Saccharide: Generic term for any carbohydrate or assembly of carbohydrates, in free form or attached to another molecule, used interchangeably in this report with carbohydrate and glycan.

Sialic acids: Family of acidic sugars with a nine-carbon backbone, of which the most common is *N*-acetylneuraminic acid, in vertebrates.

Siglecs: Sialic acid–binding proteins that are members of the I-type lectin family and have an amino-terminal V-set domain with typical conserved residues.

Sugar: Generic term often used to refer to any carbohydrate but most frequently to low molecular weight carbohydrates that are sweet in taste. Table sugar, sucrose, is a nonreducing disaccharide (Fruβ2-1αGlc). Oligosaccharides are sometimes called "sugar chains," and individual monosaccharides in a sugar chain are sometimes referred to as "sugar residues."

Transcriptome: The total collection of RNA transcripts in a cell, a tissue, or an organism, under specific conditions of time, space, and environment.