Springer Protocols

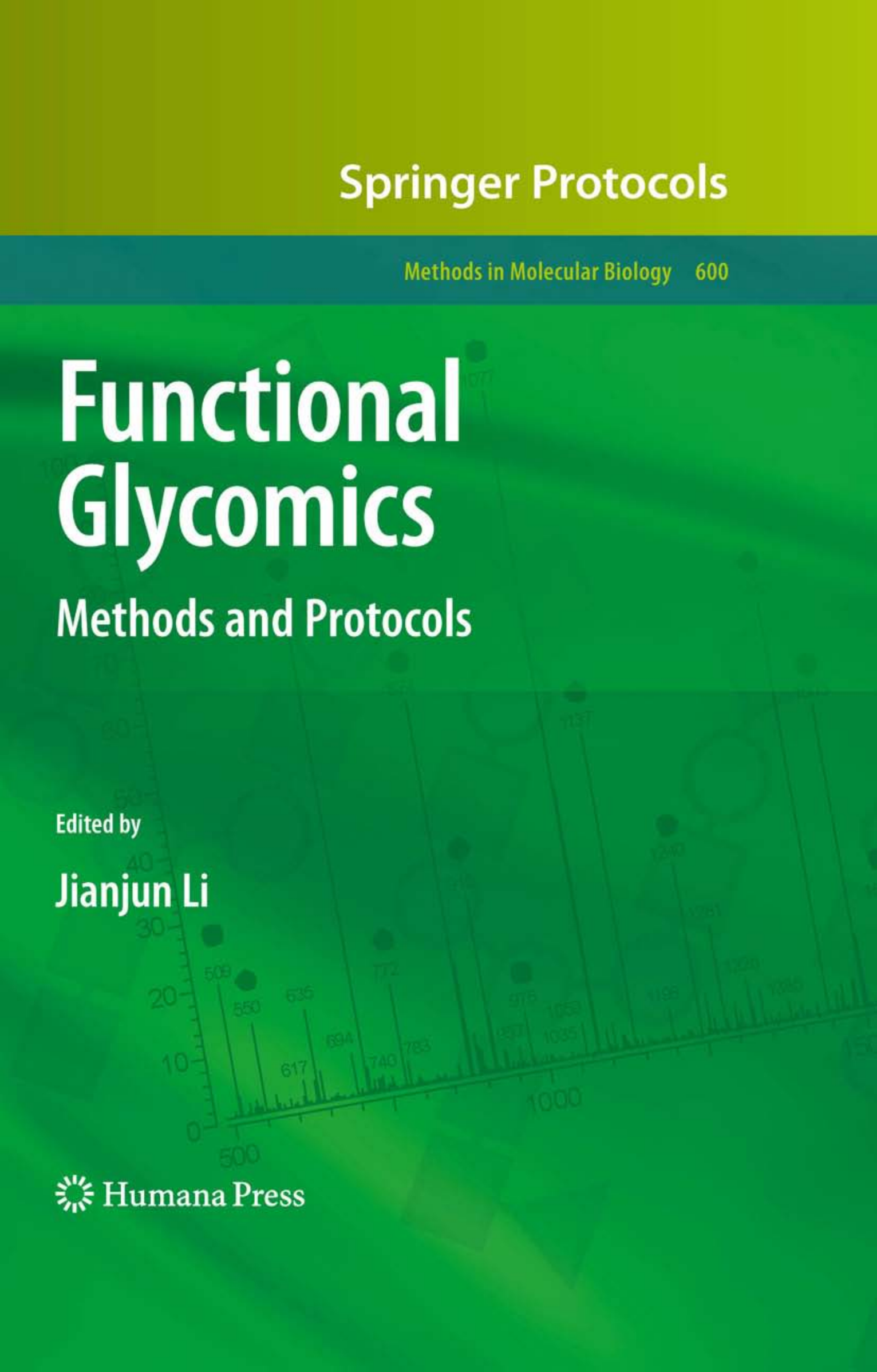
Methods in Molecular Biology 600

Functional Glycomics

Methods and Protocols

Edited by

Jianjun Li



Humana Press

METHODS IN MOLECULAR BIOLOGY™

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For other titles published in this series, go to
www.springer.com/series/7651

Functional Glycomics

Methods and Protocols

Edited by

Jianjun Li

National Research Council of Canada, Ottawa, ON, Canada

Humana Press

Editor
Jianjun Li
National Research Council
of Canada
Inst. Biological Sciences
100 Sussex Drive
Ottawa, ON, K1A OR6
Canada
jianjun.li@nrc-cnrc.gc.ca

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-60761-453-1 e-ISBN 978-1-60761-454-8
DOI 10.1007/978-1-60761-454-8

Library of Congress Control Number: 2009938350

Cover illustration: Cover image adapted from Figure 6 of Chapter 3 and Figure 2 of Chapter 14

Cover design: Cover sketch created by Karen Schulz

Printed on acid-free paper

springer.com

Preface

Glycans in proteins and lipids play key roles in modulating their structures and functions in living organisms. Functional glycomics is the study of structure–function relationships of glycans in a given biosystem. It requires the continuous development of rapid and sensitive methods for the identification of glycan structures and integration to structure–function relationships. In the recent years, a broad range of technologies is evolving rapidly to meet the needs of this field. These methods, including sample preparation, pre-fractionation, analytical separations, quantitation, mass spectrometry, microarray, bioinformatics, have advanced rapidly to support the explosive growth of application in functional glycomics. It is the aim of this book to provide new developments and emerging glycomics techniques to the researchers in the functional glycomics field. It is not solely a simple collection of methods; instead, detailed protocols have been written by world leaders in the field and demonstrated with respect to the fundamental challenges and novel applications.

A review on carbohydrate analysis is introduced at the beginning of this book (Chapters 1–3), including the nomenclatures, traditional analytical techniques, modern mass spectrometry strategies, NMR, and microarray techniques. Glycosphingolipids are the major components of the plasma membrane, which play key roles in cellular trafficking, signaling functions, and interactions with external agents. These glycolipids are important for the development of the nervous system. Lipopolysaccharides (LPS) and lipooligosaccharides (LOS), on the other hand, have important structural and functional roles in the life of a bacterial cell. In the context of functional glycomics, Chapter 4 is thus devoted to a review of human gangliosides and bacterial LOS in the development of autoimmune neuropathies. Chapter 5 introduces a synthetic approach to polyacrylamide-based glycoconjugates with biotin tag, with special emphasis on the development of carbohydrate biosensors and arrays. In general, permethylation increases the MS response by several orders of magnitude, and sequence information is readily obtained. Thus, Chapter 6 will discuss the details for using this approach to identify all isomeric glycoforms in very heterogeneous LPS preparations. In the past decade, much effort has been put into elucidating the biosynthetic pathway of O antigens. Chapter 7 deals with the development of technologies that will permit in vitro reconstitution of the O-repeating unit via glycosyltransferases. New strategies for the rapid determination of N-glycosylation sites and site heterogeneity are introduced in Chapters 8 and 9. These methods are rapid, reliable, and applicable to all organisms capable of producing N-linked glycans. The protein–carbohydrate interactions are discussed in Chapter 10, with an emphasis on a microarray strategy. This emerging technique presents an opportunity for high-throughput tools to speed research in glycomics. NMR spectroscopy has grown enormously in importance over the past decade in the field of functional glycomics. A highlight of its application is presented in Chapter 11. Glycosylation involves several multi-enzymatic pathways that generate a diverse and complex range of structures. One approach is the development of new enzymatic and analytical tools for structural characterization of carbohydrates and the integration with other functional "-omics", e.g., metabolomics and lipidomics. Chapters 12 and 13, therefore, focus on new developments in the characterization of intermediates in biosynthesis

pathways. Glycans in serum are of particular interest as approximately half of all proteins are glycosylated. Chapter 14 will highlight the methods for profiling the glycans in human serum to identify glycan biomarkers. Chapter 15 describes a method for connective tissue glycosaminoglycan extraction, derivatization, and workup for subsequent capillary electrophoretic and/or mass spectrometric analysis. Mass spectrometry provides a rapid and reliable method for characterization of bacterial total glycomes. The comprehensive characterizations of bacterial protein glycosylation are presented in Chapter 16, while the direct analysis of bacterial cells allows for rapid analysis of cell surface polysaccharide antigens and provides a basis for serological typing and epidemiological surveillance studies of human and animal pathogens (Chapter 17). Finally, a case study against permethylation strategy, which has been widely used in glycomics community, is demonstrated with the analysis of fish serum N-glycans in Chapter 18. Given the importance of information integration in glycomics, the last chapter focuses on the bioinformatics platform to integrate the diverse data obtained from different technologies to allow a systems approach to glycan structure–function relationships.

Functional Glycomics: Methods and Protocols is aimed at both beginners and experienced researchers in the field of glycobiology, including biochemistry, molecular biology, cell biology, immunology, microbiology, and virology. Readers will find it very useful not just because of a variety of new techniques, but also the valuable success stories of the problem-resolving powers of a diverse set of analytical tools. This book then serves as a reference to guide researchers toward appropriate techniques and to design achievable research plans.

Jianjun Li

Contents

Contributors

Luis Afonso • *Institute for Marine Biosciences, National Research Council Canada, Halifax, Nova Scotia, Canada*

Eleonora Altman • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*

Hyun Joo An • *Department of Chemistry and Department of Biochemistry and Molecular Medicine, University of California, Davis, CA, USA*

Arun Apte • *PREMIER Biosoft International, Palo Alto, CA, USA*

Veer Bhatt • *Departments of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA*

Alicia M. Bielik • *Department of Biochemistry, Boston University School of Medicine, Boston, MA, USA*

Nicolai V. Bovin • *M. M. Shemyakin and Yu. A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russian Federation*

Jean-Robert Brisson • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*

Alexander A. Chinarev • *M. M. Shemyakin and Yu. A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russian Federation*

Ivan K. Chu • *Department of Chemistry, The University of Hong Kong, Hong Kong, China*

David M. Cox • *AB/MDS Sciex, Concord, Ontario, Canada*

Oxana E. Galanina • *M. M. Shemyakin and Yu. A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russian Federation*

Jingqing Gao • *State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, and Graduate School of the Chinese Academy of Sciences, Beijing, China*

Wanyi Guan • *Departments of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA; National Glycoengineering Research Center and State Key Laboratory of Microbial Technology, Shandong University, Shandong, China*

Weiqing Han • *Departments of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA*

Joseph P. M. Hui • *Institute for Marine Biosciences, National Research Council Canada, Halifax, Nova Scotia, Canada*

Harold Jarrell • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*

John F. Kelly • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*

Nam Huan Khieu • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*

Maggie P. Y. Lam • *Department of Chemistry, The University of Hong Kong, Hong Kong, China*

EDWARD LAU • *Department of Chemistry, The University of Hong Kong, Hong Kong, China*
CARLITO B. LEBRILLA • *Department of Chemistry and Department of Biochemistry and Molecular Medicine, University of California, Davis, CA, USA*
JIANJUN LI • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*
LEI LI • *Departments of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA; National Glycoengineering Research Center and State Key Laboratory of Microbial Technology, Shandong University, Shandong, China*
XIANWEI LIU • *Departments of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA; National Glycoengineering Research Center and State Key Laboratory of Microbial Technology, Shandong University, Shandong, China*
XIN LIU • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*
SUSAN LOGAN • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*
DAVID J. MCNALLY • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*
NINGOMBAM SANJIB MEITEI • *PREMIER Biosoft International, Palo Alto, CA, USA*
HARALD NOTHAFT • *Department of Biological Sciences, Alberta Ingenuity Centre for Carbohydrate Science, University of Alberta, Edmonton, Alberta, Canada*
RON ORLANDO • *Departments of Biochemistry and Molecular Biology, and Chemistry, Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA*
NICHOLAS PETTIT • *Departments of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA*
CHRISTOPHER W. REID • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*
JAMES C. RICHARDS • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*
IAN SCHOENHOFEN • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*
ELKE K. H. SCHWEDA • *Clinical Research Centre, Karolinska Institute and Södertörn University, Huddinge, Sweden*
S. O. SIU • *Department of Chemistry, The University of Hong Kong, Hong Kong, China*
JING KATHERINE SONG • *Center for Microbial Pathogenesis in The Research Institute at Nationwide Children's Hospital, Columbus, OH, USA*
EVELYN C. SOO • *Institute for Marine Biosciences, National Research Council Canada, Halifax, Nova Scotia, Canada*
JACEK STUPAK • *Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada*
CHRISTINE M. SZYMANSKI • *Department of Biological Sciences, Alberta Ingenuity Centre for Carbohydrate Science, University of Alberta, Edmonton, Alberta, Canada*
LUC TESSIER • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*
SUSAN M. TWINE • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*
EVGENY VINOGRADOV • *Institute for Biological Sciences, National Research Council Canada, Ottawa, Ontario, Canada*

PENG GEORGE WANG • *Departments of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA*
ZHENXIN WANG • *State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, and Graduate School of the Chinese Academy of Sciences, Beijing, China*
ROBERT WOODWARD • *Departments of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA*
WILLIAM S. B. YEUNG • *Department of Obstetrics and Gynaecology, The University of Hong Kong, Hong Kong, China*
WEN YI • *Departments of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA*
NOBUHIRO YUKI • *Departments of Neurology and Clinical Research, Niigata National Hospital, Niigata, Japan*
JOSEPH ZAIA • *Department of Biochemistry, Boston University School of Medicine, Boston, MA, USA*

Chapter 1

Functional Glycomics and Glycobiology: An Overview

Jianjun Li and James C. Richards

Abstract

Glycomics is the study of the biological role of glycans and glycoconjugates, including glycoproteins, glycolipids, proteoglycans, and of protein–glycan interactions. This chapter outlines the scope of functional glycomics, from biological/biomedical significance to technology development.

Key words: Glycomics, glycobiology, glycoconjugate, glycoprotein, glycolipid, glycosaminoglycan.

1. The Historical Background of Glycobiology

Coupling carbohydrate moieties to proteins and lipids, i.e., glycosylation, is a key factor in modulating their structures and biological functions. Glycoconjugates (termed glycoproteins and glycolipids) decorate the outer surfaces of all cells and serve as cellular identification tags to the surrounding environment. Despite the importance of carbohydrates, research on their functions has remained in the shadows of genomic and proteomic studies because glycan structures are complicated and not encoded directly from the genome. Indeed the complexity of carbohydrates is unique in that, unlike nucleic acids and proteins, glycans are linked together in more than one way, i.e., to one of three or four different positions of the neighboring monosaccharide unit and in either an α-glycosidic or β-glycosidic linkage. Nevertheless, over the last two decades, researchers have increasingly turned their attention toward understanding the role of carbohydrates in normal cellular function and in disease. This is to a large extent because of

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_1,

significant advances that have increased the ability to study their structures. These advances have opened up new research fronts in terms of probing glycans as targets in the design and development of novel drugs or new therapies for infectious disease, neurological disease, cancer, and metabolic disorders (1, 2).

Glycobiology is defined as the study of the structure, biosynthesis, and biological functions of glycans and glycoconjugates, including glycoproteins, glycolipids, proteoglycans, and on protein–glycan interactions (such as lectins, glycosyltransferases, and glycosidases). In eukaryotic systems, it has become clear that the carbohydrate substituents (termed glycoforms) provide a dynamic dimension during development and can change dramatically at different stages of cellular differentiation. Prokaryotes are also covered by carbohydrates. Lipopolysaccharides (LPS) or lipooligosaccharides (LOS) make up the outer leaflet of the outer membrane of Gram-negative bacteria, while capsulan polysaccharides (CPS) often envelope both Gram-negative and Gram-positive organisms. Moreover, bacterial glycoproteins have also been described.

1.1. Glycan and Disease

The surfaces of all vertebrate cells are decorated with a dense and complex array of sugar chains, which are mostly attached to proteins and lipids (3). Changes in glycosylation are often a hallmark of disease states. For example, cancer cells frequently display glycans at different levels or with fundamentally different structures than those observed on normal cells (4). Glycans are one of the four fundamental biomolecular components of all cells and are highly regulated in the immune system. On the other hand, glycans can function as antigenic markers revealing the presence of an unnatural or infectious agents. Carbohydrate differences in phylogeny and between species can allow for this type of molecular discrimination between self and non-self or invader. Therefore, glycan structures are activators of the innate immune system and pathogenic triggers for autoimmune disease (5). The connection between glycosylation and disease include mild as well as severe syndromes and can span from early neonatal to adult life (6).

The study of the carbohydrates that provide docking for bacteria and viruses opens the door to biomarker discovery. For example, the preferential recognition by influenza virus of specific linkages of sialic acid has significant implications for viral transmission and pathogenesis. The drugs Relenza and Tamiflu battle flu virus by inhibiting the viral neuraminidase, which allows newly produced virus to bud free of the cell surface binding sites to infect other cells. Bacterial pathogens are a major cause of human diseases worldwide, including gastric and respiratory tract infections some of which are potentially fatal such as cholera and meningitis. Bacteria are classified as Gram-positive and

Gram-negative depending on the structure of their outer membrane. Gram-positive bacteria possess a thick cell wall containing many layers of peptidoglycan and teichoic acids and are often encapsulated. In contrast, Gram-negative bacteria have a relatively thin cell wall consisting of a few layers of peptidoglycan surrounded by a second lipid membrane containing LPS/LOS, lipoproteins and some are encapsulated by CPS. LPS/LOS and CPS participate in many physiological processes and play a key role in the pathogenesis and manifestation of bacterial infection. The structural diversity of bacterial glycans and their expression by both pathogens and commensals inhabiting diverse host sites reveal the importance of the study of interactions of carbohydrates with the host immune system.

1.2. Glycoproteins

It is well known that glycans play key roles in protein folding, protein–protein interaction, cell–cell recognition, cancer metastasis, and the immune system. Glycosylation can substantially modify the structure and function of proteins by steric influences involving intermolecular and intramolecular interactions and can mediate interactions with other proteins termed lectins (5). In the glycoproteins of higher organisms, glycans typically attach to proteins via N-linkages and O-linkages, termed as N-glycans and O-glycans, respectively. N-Glycans are typically linked to asparagine (Asn) residues in the peptide chain having the codon of Asn-X-Thr/Ser (X is any amino acid except Pro); while O-glycans are attached to serine (Ser) or threonine (Thr) residues. The study of protein glycosylations remains at the forefront of glycomics research.

1.3. Glycolipids

Glycolipids typically form less than 5% of the lipid content of most mammalian tissues, but represent more than 25% of the lipid content of myelin sheaths that insulate axons in the nervous system (7). Gangliosides are sialic acid-containing glycolipids and highly abundant in the vertebrate nervous system. Gangliosides have been implicated as antigens in various autoimmune neurological diseases, such as Guillain–Barré syndrome (GBS). GBS is a post-infectious autoimmune-mediated neuropathy, which can be triggered by the display of LOS-bound ganglioside mimics by the bacterium *Campylobacter jejuni* (8).

1.4. Glycosaminoglycans

The glycosaminoglycans (GAG) consist of disaccharide repeating units and the polysaccharides include hyaluronan, heparan sulfate, keratan sulfate, chondroitin sulfate, and dermatan sulfate. The classification of GAG is based on the type of hexosamine, hexose, or hexuronic acid unit that they contain, as well as the geometry of the glycosidic linkage between these units. Thus, hyaluronic acid is composed of a disaccharide unit containing β-D-glucuronic acid (GlcA) and D-N-acetylglucosamine (GlcNAc)

residues, while heparin which also contains the C-6 epimer of GlcA, iduronic acid (IdoA) is a heterogonous mixture of disaccharide units. When one or more GAG chains are covalently attached to a protein through O-linkage, the glycoconjugate is defined as a proteoglycan. Proteoglycans can be found in tissues and extracellular matrices of tissues and exhibit great structural variation. Heparan sulfate proteoglycans have essential roles in all of these systems (9). The inactivity of specific lysosomal enzymes that are responsible for degrading glycosaminoglycans can result in the accumulation of proteoglycans within cells. Consequently, this deficiency leads to a variety of disease symptoms associated with the type of proteoglycan that is not degraded. Because this variation is a means whereby the functions of a limited number of protein gene products in animal genomes are elaborated, the phenotypic and functional assessment of GAG structures expressed spatially and temporally is an important goal in glycomics (10).

2. The Concept of Functional Glycomics

The cells of all organisms consist of four fundamental biomolecules, i.e., nucleic acids (DNA and RNA), proteins, lipids, and glycans. To define the research in these fields, the terms "genomics", "proteomics", "metabonomics", "lipidomics", and "glycomics" have emerged (10).

We are currently in the post-genomics era and proteomics is considered to be the most prominent, which aims to study all the proteins in a cell or organism including post-translational modification. However, most proteomics research has not adequately addressed the analytical and functional details of the glycosylation of proteins, lipids, and proteoglycans. Consequently, many opportunities for new discovery are missed. The emerging field of glycomics, defined as the study of the biological role of glycans, requires an arsenal of high-throughput technologies and expertise drawn from a myriad of disciplines including chemistry, bioanalysis, molecular and cell biology, medicine, and informatics (11). In the past several years, the field of glycomics has focused considerable attention on the development of enabling technologies. This is because of the inherent level of complexity in the field of glycomics compared to other omics. While there is considerable effort being placed on the development of new technologies, the primary aspect of glycomics research is focused on the functional study of glycosylation. Advances in glycomics are anticipated to be benefited by large-scale collaborative networks throughout the world. The National Institute of General Medical Sciences (NIGMS) has been a pioneer in finding innovative ways

to fund team science, exemplified by the institute's "glue grant" program. One such grant, the Consortium for Functional Glycomics (CFG), was launched in October 2001 as a large-scale, collaborative effort to systematically map interactions between glycans and proteins on cell surfaces. The scientific focus and goal of the CFG is the elucidation of the functions of glycan-binding proteins that mediate biology at the surface of a mammalian cell (http://www.functionalglycomics.org/static/index.shtml). In May 2004, the Glycoarrays Consortium was funded by the Basic Technology Programme of Research Councils UK (http://www.glycoarrays.org.uk/). The aim of this consortium is to develop carbohydrate arrays (glycoarrays) as innovative tools to map highly specific interactions between carbohydrate and protein partners. Another major collaboration initiative is The Human Disease Glycomics/Proteome Initiative (HGPI) under the Human Proteome Organization (HUPO). The overall plan of this initiative is to identify disease-related glycosylation changes in blood and urine that represent potentially useful glycobiomarkers for early diagnosis, monitoring, and treatment of common diseases such as cancer, inflammation, life style-related diseases, neurodegenerative disease, and congenital disorders of glycosylation (CDG) (http://www.hupo.org/research/hgpi/). A meeting entitled "The Frontiers in Glycomics; Bioinformatics and Biomarkers in Disease" was jointly organized by CFG, HGPI, National Cancer Institute, National Institute of General Medical Sciences, and Japan Society for the Promotion of Science and National Center for Research Resources in September 2006. The mandate of this meeting was to describe and identify the state and the potential of glycans as biomarkers for diseases and to recommend the tools that need to be supported so as to develop and enhance biomarker discoveries that involve the glycome (12). More recently, the Division of Blood Diseases and resources of the National Heart, Lung, and Blood Institute (NHLBI) convened a working group of scientific investigators in February 2008, in Bethesda, MD, to identify scientific opportunities and priorities emerging from the recent explosion of technological and biological advances in the glycosciences (11).

3. The Emerging Technologies in Functional Glycomics

Carbohydrate sequencing and the synthesis of defined oligosaccharides are two key technologies that have contributed to progress in glycomics research (1). Advances in glycomics are anticipated to be driven by improvements in the development of techniques to synthesize complex glycan structures,

analysis of the interactions of glycans with other molecules, carbohydrate sequencing, and bioinformatics. Advances in analytical techniques, including MS and NMR, are having a significant impact on the field of glycomics. These techniques are extremely useful for the characterization of glycan structures and their quantities. More recently, there has been increasing interest in developing microarray strategies, including glycan microarrays and lectin microarrays.

3.1. Mass Spectrometry

Mass spectrometry has been widely used in the analysis of biomolecules, such as mass spectrometry-based approaches for proteomics, lipidomics, and metabonomics (13). It has become a highly sensitive and powerful technique for glycomics analysis. By coupling mass spectrometry with database structures, automated analysis of the carbohydrate structures of glycoprotein has become possible. These emerging techniques have been applied to the study of the role of glycans in infectious disease, cancer, and age-related disease.

3.2. Microarray

A major area of technology development has been the recent emergence of microarrays, including glycan microarrays and lectin microarrays. Carbohydrate microarrays allow the rapid screening of interactions between glycans and other molecules. For example, the application of Globo H and its truncated analogs in microarray format has been reported to profile the binding specificity of monoclonal and polyclonal antibodies to the relevant carbohydrate epitopes (14). The glycan microarray thus provides a new platform to monitor the immune response to carbohydrate epitopes after vaccine therapy or during the course of cancer progression (14). Continuing efforts to develop carbohydrate-based vaccines for pathogens and cancer have highlighted the further benefit from microarrays. Alternatively, lectin microarrays were developed for rapid profiling of glycosylation, although their use was mainly restricted to glycoproteins from cell lysates, and thus unable to profile intact cell surface glycans (15). Recently, this technology has been further developed for direct analysis of the dynamic bacterial glycome (16) and live mammalian cell-surface glycome (17). In addition, antibody arrays for glycan detection were found to be highly effective for profiling variation in specific glycans on multiple proteins and should be useful in diverse areas of glycobiology research (18).

3.3. Bioinformatics

The enormous size of data generated from the different platforms and/or different laboratories has resulted in the development of bioinformatics, with the goal of building a systems biology approach to glycan structure–function relationships (19). The past few years have seen the arrival of many commercial programs and algorithms that assign glycan structures based on MS spectra

(2). Among them, SimGlycan™ is a comprehensive desktop tool designed to address the challenges of glycan structure prediction from glycan mass spectrometry data. Not only does it predict the glycan structure using an MS/MS database searching technique, but it also facilitates predicting novel glycans by providing the ability to generate the glycan structure, predicate fragmentation patterns to compare with observed spectrum.

4. Future Perspective

The field of glycomics is vast and diverse in nature. Carbohydrates are important in many biological processes, but the full extent of their distribution and function remains unclear (2). Despite recent technological advances, glycomics research is still in its infancy compared to genomics and proteomics. Further innovation and development in analytical methods are essential to advancing the biomedical applications of glycomics, including sample preparation, separations, MS, NMR, microarrays, and other chemical technologies. Another key area of intense activity is the design of bioinformatics tools for glycomics. The development of new bioinformatics tools, algorithms, and data collections is an absolute requirement to manage and analyze successfully the large amount of data that will be produced by upcoming glycomics projects (20).

References

1. Seeberger, P. H. and Werz, D. B. (2007) Synthesis and medical applications of oligosaccharides. *Nature* **446**, 1046–1051.
2. Blow, N. (2009) Glycobiology: a spoonful of sugar. *Nature* **457**, 617–620.
3. Varki, A. (2008) Sialic acids in human health and disease. *Trends Mol. Med.* **14**, 351–360.
4. Dube, D. H. and Bertozzi, C. R. (2005) Glycans in cancer and inflammation–potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **4**, 477–488.
5. Marth, J. D. and Grewal, P. K. (2008) Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* **8**, 874–887.
6. Ohtsubo, K. and Marth, J. D. (2006) Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855–867.
7. Kanter, J. L., Narayana, S., Ho, P. P., Catz, I., Warren, K. G., Sobel, R. A., Steinman, L., and Robinson, W. H. (2006) Lipid microarrays identify key mediators of autoimmune brain inflammation. *Nat. Med.* **12**, 138–143.
8. Yuki, N., Susuki, K., Koga, M., Nishimoto, Y., Odaka, M., Hirata, K., Taguchi, K., Miyatake, T., Furukawa, K., Kobata, T., and Yamada, M. (2004) Carbohydrate mimicry between human ganglioside GM1 and *Campylobacter jejuni* lipooligosaccharide causes Guillain-Barre syndrome. *Proc. Natl. Acad. Sci. U. S. A* **101**, 11404–11409.
9. Bishop, J. R., Schuksz, M., and Esko, J. D. (2007) Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* **446**, 1030–1037.
10. Zaia, J. (2008) Mass spectrometry and the emerging field of glycomics. *Chem. Biol.* **15**, 881–892.
11. Varki, A. P., Baum, L. G., Bellis, S. L., Cummings, R. D., Esko, J. D., Hart, G. W., Linhardt, R. J., Lowe, J. B., McEver, R. P., Srivastava, A., and Sarkar, R. (2008) Working group report: the roles of glycans in hemostasis, inflammation and vascular biology. *Glycobiology* **18**, 747–749.

12. Packer, N. H., von der Lieth, C. W., oki-Kinoshita, K. F., Lebrilla, C. B., Paulson, J. C., Raman, R., Rudd, P., Sasisekharan, R., Taniguchi, N., and York, W. S. (2008) Frontiers in glycomics: bioinformatics and biomarkers in disease. An NIH white paper prepared from discussions by the focus groups at a workshop on the NIH campus, Bethesda MD (September 11–13, 2006). *Proteomics* **8**, 8–20.
13. Dell, A. and Morris, H. R. (2001) Glycoprotein structure determination by mass spectrometry. *Science* **291**, 2351–2356.
14. Wang, C. C., Huang, Y. L., Ren, C. T., Lin, C. W., Hung, J. T., Yu, J. C., Yu, A. L., Wu, C. Y., and Wong, C. H. (2008) Glycan microarray of Globo H and related structures for quantitative analysis of breast cancer. *Proc. Natl. Acad. Sci. U. S. A* **105**, 11661–11666.
15. Kuno, A., Uchiyama, N., Koseki-Kuno, S., Ebe, Y., Takashima, S., Yamada, M., and Hirabayashi, J. (2005) Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nat. Methods* **2**, 851–856.
16. Hsu, K. L., Pilobello, K. T., and Mahal, L. K. (2006) Analyzing the dynamic bacterial glycome with a lectin microarray approach. *Nat. Chem. Biol.* **2**, 153–157.
17. Tateno, H., Uchiyama, N., Kuno, A., Togayachi, A., Sato, T., Narimatsu, H., and Hirabayashi, J. (2007) A novel strategy for mammalian cell surface glycome profiling using lectin microarray. *Glycobiology* **17**, 1138–1146.
18. Chen, S., LaRoche, T., Hamelinck, D., Bergsma, D., Brenner, D., Simeone, D., Brand, R. E., and Haab, B. B. (2007) Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. *Nat. Methods* **4**, 437–444.
19. Raman, R., Raguram, S., Venkataraman, G., Paulson, J. C., and Sasisekharan, R. (2005) Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat. Methods* **2**, 817–824.
20. von der Lieth, C. W., Bohne-Lang, A., Lohmann, K. K., and Frank, M. (2004) Bioinformatics for glycomics: status, methods, requirements and perspectives. *Brief. Bioinform.* **5**, 164–178.

Chapter 2

Historical Overview of Glycoanalysis

Alicia M. Bielik and Joseph Zaia

Abstract

More than half of all human proteins are glycosylated. Glycosylation defines the adhesive properties of glycoconjugates and it is largely through glycan–protein interactions that cell–cell and cell–pathogen contacts occur. Not surprisingly, considering the central role they play in molecular encounters, glycoprotein and carbohydrate-based drugs and therapeutics represent a greater than $20 billion market. Glycomics, the study of glycan expression in biological systems, relies on effective analytical techniques for correlation of glycan structure with function. This overview summarizes techniques developed historically for glycan characterization as well as recent trends. Derivatization methods key to both traditional and modern approaches for glycoanalysis are described. Monosaccharide compositional analysis is fundamental to any effort to understand glycan structure–function relationships. Chromatographic and electrophoretic separations are key parts of any glycoanalytical workflow. Mass spectrometry and nuclear magnetic resonance are complementary instrumental techniques for glycan analysis. Finally, microarrays are emerging as powerful new tools for dynamic analysis of glycan expression.

Key words: Glycomics, glycoconjugate, carbohydrate, glycan, mass spectrometry, chromatography, capillary electrophoresis, nuclear magnetic resonance.

1. Introduction

The central paradigm of modern molecular biology is that information flows from DNA to RNA to protein to cell. Proteins are vital parts of living organisms, as they are the main components of the physiological pathways of cells. Proteomics is the study of proteins in biological systems and has expanded the knowledge of protein expression, modification, interaction, and function. Most animal proteins are post-translationally modified by glycosylation (1). The glycan or carbohydrate portion of these glycoconjugates

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_2,

can serve as signaling molecules, energy storage molecules, and structural components (2).

The major classes of animal glycans (shown in **Fig. 2.1**) include N-linked glycans, O-linked glycans, glycolipids, and glycosaminoglycans (GAGs). N-Linked glycans are attached to the nitrogen of an asparagine in the core protein. N-Linked glycans have been found to be important in proper protein folding mechanisms in eukaryotic cells and can also play an important role in cell–cell interactions (3). O-Linked glycans are created by the addition of a GalNAc residue to a serine or threonine residue of a core protein and are less branched than most N-glycans (4). Hyper-O-glycosylation can result in the formation of mucin-type molecules that coat mucosal surfaces (3, 5). Glycolipids are amphipathic molecules that consist of two components: an outer glycan portion attached to an inner lipid tail (6). Glycolipids are significant contributors to the structure of most eukaryotic cell membranes and can serve as markers for cellular recognition (7). GAGs are linear polysaccharides attached to proteoglycan core proteins, whose disaccharide building blocks consist of an amino sugar and an uronic acid (3). GAGs play significant roles in the modulation of cellular signals through interactions with proteins, including growth factors and growth factor receptors (8–10). O-GlcNAc is one of only a few O-linked monosaccharides and is a major form of intracellular glycosylation on the hydroxyl groups of serine and threonine, which produces O-GlcNAcylated proteins (O-GlcNAc-modified proteins) (11). O-GlcNAcylation of proteins has been shown to regulate trafficking of these molecules into and out of the nucleus (12). These proteins are known to form reversible multimeric complexes with other polypeptides through regulation by phosphorylation (13, 14).

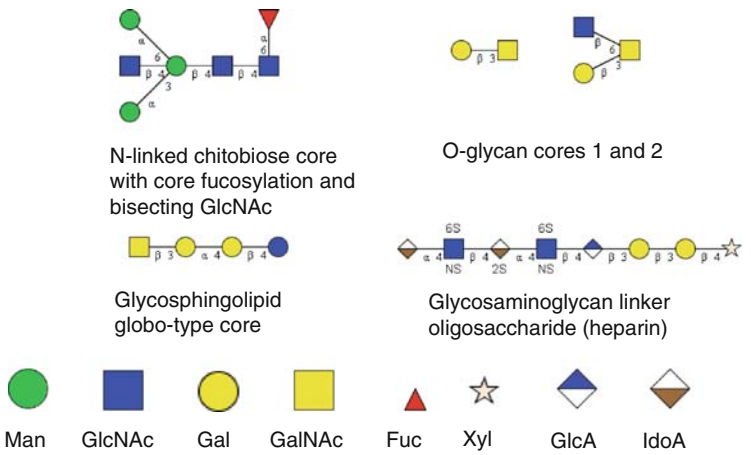


Fig. 2.1. Common animal glycoconjugate core structures.

Glycomics is the study of glycan expression in biological systems. A glycomics approach may combine three main aspects: the structural characterization of glycans, detailed understanding of glycan–protein interactions, and the study of in vitro and in vivo model systems for the determination of function as it correlates to specific glycan structures (2). Despite the numerous known biological roles of glycans, the field of glycomics is much less established in comparison to the field of proteomics. This results from the structural complexity of glycans due to the high degree of glycosylation that is possible and the heterogeneity of glycans at each site. The difficulties associated with the field of glycomics are compounded by the problems with determining glycan sequence and the fact that glycan biosynthesis cannot be predicted from a DNA template. The complex structural characteristics and glycosylation patterns, along with the lack of an overall structural template, create a huge challenge in the analysis of glycans. However, in recent years the field of glycomics has begun to expand and develop.

Structure–function relationships of glycans are important because of the numerous roles they play in critical biological processes. The number of roles implicated for glycans in biological processes has dramatically increased in recent years, due to the expansion of the field of glycobiology. Carbohydrates have vital functions in the body; in particular, carbohydrates can serve as signaling molecules, as intermediates to generate energy, and as structural components (3).

The intent of this review is to summarize the analytical techniques used in modern glycomics research. It is not possible to cover comprehensively all developments in glycoanalysis, and only a few examples are given to illustrate each of the approaches. We apologize to those authors whose references it was not possible to include.

2. Glycoanalytical Methods

Traditional approaches to glycan analysis, including specific exoglycosidases, derivatization, monosaccharide analysis, chromatography, and electrophoresis, remain important in a modern glycoanalytical workflow. Instrumental techniques including mass spectrometry (MS) and nuclear magnetic resonance (NMR) are evolving rapidly to meet the emerging needs of glycobiology. New microarray approaches are enabling dynamic profiling of glycan–protein interactions. These approaches are described below.

2.1. Specific Exoglycosidases in Glycan Analysis

Several techniques discussed in this review to identify and sequence oligosaccharides require exoglycosidases to sequentially remove sugar residues from glycoconjugates. Using fluorophore-labeled oligosaccharides, researchers have made it possible to screen, purify, and characterize a wide array of exoglycosidases (15). At present, more than 40 exoglycosidases are available and, when used in combination with one another, can be used to modify and sequence oligosaccharides. For example, exoglycosidases can confirm the difference between GalNAc and GlcNAc residues and can identify specific linkages of sialic acids, α-fucose, α-mannose, and β-galactose. Analysis of exoglycosidase-digested glycoconjugates by high pH anion exchange chromatography with pulsed amperometric detection (16) normal-phase and reversed-phase chromatography (17–19), thin layer chromatography (20), fluorophore-assisted carbohydrate electrophoresis (FACE) (21), or MS (22, 23) can lead to more detailed sequence and linkage information.

2.2. Derivatization Methods

2.2.1. Reductive Amination

As shown in **Fig. 2.2A**, reductive amination is often used in the analysis of released glycans to add a chromophore or fluorophore and to incorporate a means for stable isotopic labeling (24, 25). This addresses one of the main problems in the chemical analysis of glycans – the inability to detect the sugars during chromatographic processes. A highly sensitive means of oligosaccharide detection is needed due to the limited sample size that is often available. The addition of a chromophore or fluorophore via reductive amination reactions is one of the most widely practiced labeling techniques, because this step improves chromatographic, electrophoretic, and mass spectrometric detection (26, 27). The reducing end residue of an oligosaccharide is always

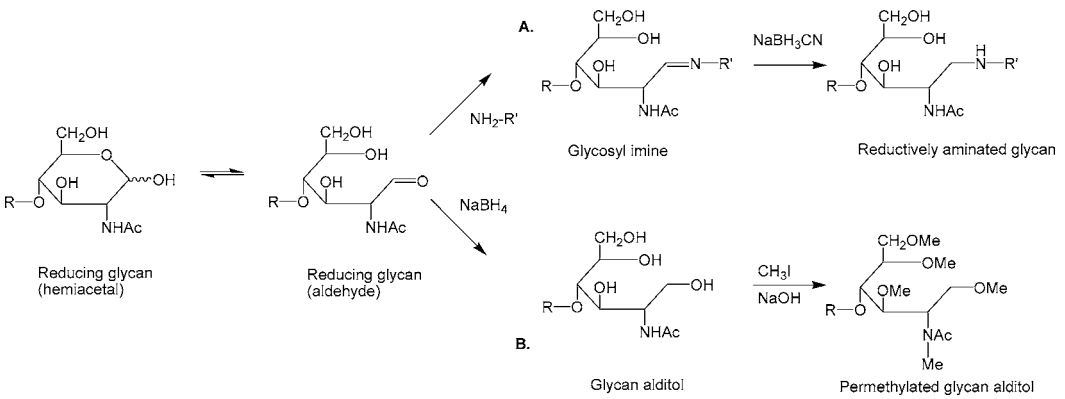


Fig. 2.2. Derivatization reactions for reducing glycans. R = glycan moiety or H.

in equilibrium between its closed ring (hemi-acetal) form and its open ring (aldehyde) form. When the reducing end aldehyde of the oligosaccharide is treated with a primary amine under acidic conditions, an imine is formed. Acidic conditions create a favorable path for the glycan to form an imine, otherwise known as a Schiff base, with the chromophore (or fluorophore). The Schiff base is unstable, and thus is usually reduced with sodium cyanoborohydride to form a stable alkyl amine in which the label is permanently attached to the reducing end of the oligosaccharide, as shown in **Fig. 2.2A**.

Many aromatic, primary or secondary amine derivatives exist and are available for use with reductive amination, including 2-anthranilic acid (2-AA) (28), 2-aminobenzamide (2-AB) (29), 2-aminopyridine (2-AP) (30), 2-aminoacridone (2-AMAC) (31), 8-aminonaphthalene-1,3,6-trisulfonic acid (ANTS) (32, 33), 8-aminopyrene-1,3,6-trisulfonic acid (APTS) (34), and 6-aminoquinoline (6-AQ) (35). 2-AA is an advantageous label for the characterization of carbohydrates because it efficiently labels the reducing end of oligosaccharides (28, 36); it has a specific absorbance at 310 nm; and it is commercially available in both normal and deuterated forms. The AMAC and APTS groups are used to derivatize glycans for chromatographic and electrophoretic detection.

2.2.2. Permethylation

Permethylation, shown in **Fig. 2.2B**, is a means first used to increase the volatility and stability of carbohydrates for gas chromatography–mass spectrometry analysis. Conversion of carbohydrate OH and NH groups to O-methyl and N-methyl groups continues to be used widely in modern soft MS ionization methods, *see* Section 2.6. The derivatization method of Ciucanu and Kerek (37) has been widely used, updated (38), and adapted for high-throughput microscale workflows (39, 40). Permethylation increases glycan signal strength and maximizes the structural information produced using a tandem mass spectrometry experiment (41).

2.3. Monosaccharide and Linkage Analysis

High pH anion exchange chromatography with pulsed amperometric detection allows direct detection of monosaccharides and identification using retention time correlation (42–44). Capillary electrophoresis (CE) methods for monosaccharides have been developed in which separate hydrolysis conditions are used for sialic acids, neutral sugars, and amino sugars (45). Sialic acids are reductively aminated with 2-aminoacridone for subsequent CE separation and fluorescence detection. Neutral and amino sugars are reductively aminated with a sulfated aminopyrine group to facilitate their mobility in the electric field.

The most powerful approach for linkage analysis of carbohydrates is GC–MS analysis of partially methylated alditol acetates

(46). The sample is permethylated and then acid hydrolyzed. The resulting partially methylated monosaccharides are reduced to alditols and then peracetylated. The acetylation pattern allows the determination of the linkages present in the carbohydrate samples, as quantified using GC–MS. This approach to linkage analysis remains in wide use today.

2.4. Chromatography

Mixtures of released oligosaccharides may be separated using chromatographic modes including size exclusion chromatography (SEC), hydrophilic interaction chromatography (HILIC) (47), reversed phase (27) and reversed-phase ion-pairing (RP-IP) chromatography (48), and porous graphitized carbon (PGC) chromatography.

2.4.1. Reversed-Phase Chromatography

Reductive amination of carbohydrates with a hydrophobic tag allows for retention of oligosaccharides on reversed-phase stationeries. Several tags have been used (27) to increase the reversed-phase retention and improve the mass spectrometric ionization responses by adding a chromophore or fluorophore (49). Reversed-phase separation is usually carried out with bulky hydrophobic tags whereby carbohydrates are resolved based on hydrophobicity, which is determined primarily from the tag and size (50). To achieve high resolution, solvent gradients are required.

RP-IP is a technique in which charged additives are used as ion-pairing agents to increase the retention of oppositely charged analytes to the reversed phase (51). RP-IP with direct mass spectrometric detection has been used for the characterization of acidic glycans including glycosaminoglycans (52, 53). Desalting steps, such as on-line cation exchange, are often used prior to on-line mass spectrometric analysis (48, 54).

2.4.2. Porous Graphitized Carbon (PGC) Chromatography

PGC columns were developed for high-performance liquid chromatography (HPLC) separation of carbohydrate compounds and can be used with higher organic solvent concentrations than reversed-phase chromatography columns (55). Retention by PGC chromatography occurs by an adsorption mechanism in which planar molecules exhibit more retention than non-planar ones. The graphite surface exhibits physical and chemical stability, allowing for use over the entire pH range. The mobile phase does not require salts, creating an interface adequate for mass spectrometry (56). PGC chromatography has been used with mass spectrometric detection for the analysis of mono-sulfated N-glycans (57, 58) and O-glycans from glycoproteins (59).

2.4.3. Size Exclusion Chromatography (SEC)

High-performance SEC liquid chromatography is a simple and reliable chromatographic method that can be used as a separa-

tion technique for glycans prior to mass spectrometric detection (47, 60). SEC is a technique for separating biological macromolecules, on the basis of molecular size (61). High-performance SEC columns are available in sizes that operate at or below the 100 μL/min flow rate range, allowing the effluent to be infused with or without a post-column splitter into the mass spectrometer ion source. The SEC column is packed with a stationary phase that is a composite matrix of highly cross-linked agarose and dextran. This medium is typically resistant to most chemicals and pH changes. The separation medium also has a low non-specific interaction, which allows for high selectivity and recovery of materials.

Referring to the Van Deemter plot (62), a low flow rate is chosen for separation of molecules the size of typical oligosaccharides to reduce band broadening within a chromatogram. Therefore, one limitation of on-line SEC–LC–MS is that the chromatographic resolution depends on the column volume; thus scaling down the chromatography dimensions will decrease the resolution. MS sensitivity is limited due to the high flow rate and salt concentration that SEC requires (60).

2.4.4. Hydrophilic Interaction Chromatography (HILIC)

HILIC is normal-phase chromatography in which water is employed as the eluting solvent and is a widely useful technique for the retention and separation of polar analytes, such as peptides and carbohydrates (63, 64). It offers complementary selectivity and greater polar compound retention than does reversed-phase chromatography. This mode of chromatography separates compounds using a highly polar stationary phase, such as functionalized silica with a hydrophobic coating, and a nonpolar mobile phase, such as an acetonitrile/water mixture. The polar or hydrophilic sample experiences more attraction toward the stationary phase; therefore, using a gradient of increasing water concentrations will elute solutes in the order of increasing hydrophilicity. It offers the ability to bind and separate both charged and uncharged carbohydrates using a gradient from high to low organic content. Retention times are reproducible and predictable based on glycan size and polarity. This has allowed retention times for unknown glycans to be correlated with those of standards for the purpose of partial or full identification based on retention times (17, 65).

2.5. Capillary Electrophoresis (CE)

CE is a useful separation system for various biomolecules including glycans (66) and proteins (67). Acidic oligosaccharides may be separated directly using CE due to their negative charge. It is common, however, to add a chromophore or fluorophore via reductive amination to aid in optical detection using CE. A CE instrument consists of a fused silica separation capillary, a high-voltage power supply, an auto-sampler, and a detector. The silica capillary has at its surface silanol groups (Si–OH). These groups

are slightly acidic, with a pKa of approximately 3.5, and ionize to SiO^- at a pH higher than this pKa value. To detect the glycans within a sample, the capillary ends are placed in buffer reservoirs and a high voltage (~30 kV) is applied to the buffer through platinum electrodes via a power supply. The glycan samples to be analyzed are then injected using pressure. This process can be operated in either forward or reversed polarity mode. The forward polarity mode is executed with a basic buffer system, causing the inner surface of the capillary to become negatively charged (68). At the same time, protons migrate toward the negatively charged cathode creating an electroosmotic flow (EOF). Analytes, whether charged or not, will be swept through the capillary by the EOF. Samples are then injected at the positively charged anode and migrate to the negatively charged cathode. The more negatively charged analytes migrate fastest, while the less negatively charged samples will migrate later. In reversed polarity mode, the EOF is minimized using an acidic buffer system (pH =3.50) (69). The acidic buffer system causes the silanol groups to become protonated, and only acidic glycans retain enough anionic character to migrate toward the anode for detection. In reversed polarity mode, the more negatively charged analytes elute first, and the less negatively charged disaccharides elute later. Capillary electrophoresis with laser-induced fluorescence (CE–LIF) can be achieved by reductive amination of the disaccharides to incorporate a fluorophore, such as AMAC or APTS (70–72). CE–LIF significantly improves the sensitivity of detection as compared to CE with UV detection.

2.6. Mass Spectrometry (MS)

MS ionization techniques for the analysis of biomolecules include fast atom bombardment (FAB) (73, 74), matrix-assisted laser desorption ionization (MALDI) (75), and electrospray ionization (ESI) (76, 77). FAB was used to develop fundamental mass spectrometric principles for the analysis of carbohydrates (78); however, the FAB ionization process is significantly more energetic and less sensitive than MALDI or ESI. As a result, fragmentation of glycans was observed in the ion source, in both positive and negative modes. The fragmentation of N-linked and O-linked glycans created an avenue for the principles of carbohydrate fragmentation to be elucidated and a carbohydrate fragmentation nomenclature to be created (79). FAB is now used infrequently for glycan analysis.

2.6.1. Matrix-Assisted Laser Desorption Ionization Mass Spectrometry (MALDI MS)

Using the MALDI MS technique a sample is mixed with a chemical matrix that facilitates the production of intact gas-phase ions from large, non-volatile, and thermally labile compounds. Analytes include proteins, oligonucleotides, synthetic polymers, large inorganic compounds, and carbohydrates. The MALDI matrix plays a key role by absorbing the laser light energy and causing a small part of the target substrate to ionize. Traditional

vacuum MALDI creates higher internal energies, as compared to ESI, which cause metastable fragmentation of fragile molecules, such as sialylated glycopeptides (80). A mechanistic implication of MALDI is that singly charged ions predominate (81), creating a disadvantage for the analysis of sulfated glycans. The disadvantage exists because acidic groups (sialic acid, sulfate, and phosphate) are most labile in their protonated forms and tend to undergo fragmentation in the MALDI source under vacuum (82, 83).

2.6.2. Electrospray Ionization (ESI)

ESI is an atmospheric pressure ionization technique widely used with peptides, proteins, and carbohydrates (84, 85). It is often termed a "soft ionization" technique because it imparts relatively low energy to the molecule during desolvation. ESI produces a continuous beam of gas-phase ions directly from solution, typically an aqueous or aqueous/organic solvent system, by creating a fine spray of highly charged droplets in the presence of an electric field (typically 3–5 kV). As a droplet evaporates and decreases in size, the electric charge density on its surface increases. The mutual repulsion between like charges on this surface becomes so great that it exceeds the forces of surface tension, known as the "Rayleigh" limit (86, 87), and the droplets divide (88). Solutions that are predominantly aqueous are desolvated with the addition of nitrogen gas, which aids in the drying process and improves ion formation. Separation of analytes occurs during the uneven droplet division process and has a significant impact on the analysis of heterogeneous samples that contain components varying in their chemical and/or physical properties. Nano-ESI produces ion droplets 10-fold smaller in size (87), thus eliminating stages of droplet division required for traditional ESI. The smaller droplets formed in nano-ESI yield a higher surface area-to-volume ratio and therefore greater access to the surface of the droplets (87). Therefore, a greater percentage of the actual sample is ionized in the mass spectrometer, yielding a stronger signal for carbohydrates. Multiply charged ions result from ESI, and this is advantageous in that molecular ions can be detected by mass spectrometers having a limited mass-to-charge (m/z) range. In the negative-ion mode, oligosaccharides are ionized as $[M–nX]^{n–}$, where X is a cation. If X is an H, a proton will be taken from the oligosaccharide to form an $[M–nH]^{n–}$ ion.

2.6.3. Instrumentation of Mass Spectrometry

Mass spectrometric measurements are carried out in the gas phase on ionized analytes. A mass spectrometer has three main components: an ionization source, a mass analyzer to separate ions based on their m/z ratios, and an ion detector that counts and records the number of ions as a function of m/z. The ionization source is the place where ionization occurs. MALDI (89) and ESI (86, 90) are now the most common ionization sources for biomolecular mass spectrometry, because both are soft ionization techniques.

Ion traps typically have ion sources with a long heated capillary, which provides gentle desolvation conditions for sulfated glycans and reduces in-source fragmentation. Sources with angled ion trajectories prior to the vacuum region entrance have fairly gentle desolvation properties, while sources in which ions are sprayed directly through an orifice into the vacuum region containing skimmers are the least gentle with regard to desolvation of sulfated oligosaccharides (82). The mass analyzer is used to separate sample ions and is central to the function of a mass spectrometer in that its key properties are sensitivity, resolution, and mass accuracy. Mass resolution is a dimensionless ratio of the mass of a peak divided by its width; it is the amount of separation of two ions of similar m/z. The mass resolution that is achievable by a mass spectrometer depends on both the type of analyzer and the experimental conditions. Several different types of mass analyzers exist including the ion trap, the time-of-flight, quadrupole, and Fourier transform ion cyclotron resonance. Each analyzer is different in design and capability and can be used alone or in tandem with another. The ion trap analyzer captures or "traps" ions for a specific amount of time and then subjects them to MS and MS/MS analysis. Ion traps are robust and sensitive; however, they have relatively low mass resolutions and accuracy due to the limited number of ions that can accumulate in a trap at once before space-charging effects distort their measurements.

2.6.4. Quadrupole Ion Trap (QIT)

The quadrupole ion trap mass spectrometer (QITMS) was invented by Wolfgang Paul and co-workers in the 1950s (91). However, it took breakthroughs in design at Finnigan in the 1980s to make the QIT a widely used commercial instrument (92). The QIT gets its name from its operation; it is a pulsing ion gate that traps ions. In an ion trap the ions are transmitted from the capillary inlet source into the trap by the action of three hyperbolic electrodes: the ring electrode and the entrance and exit endcap electrodes. Various voltages are applied to these electrodes resulting in the trapping of ions. An RF voltage applied to the ring electrode dynamically traps ions in three dimensions, in a stable oscillating trajectory. The exact motion of the ions is dependent on both the voltages applied and the individual m/z ratios. To detect the ions, potentials may be altered, by applying kinetic energy to desired ions, as a means of destabilizing the ion motions, thereby ejecting them from the trap through the exit endcap. Ions are typically ejected in the order of increasing m/z by a change in potentials. This flow of ions is focused into the detector to create a mass spectrum.

Desired ions may be isolated in the trap and subjected to subsequent collision-induced dissociation (CID). It is advantageous to operate with sufficient mass resolution to resolve the isotopic distribution in a CID spectrum. The mass resolution is depen-

dent on the scan speed of the mass analyzer. A slower scan speed will acquire more data points over time, allowing for a higher resolution of mass peaks. A primary advantage of QITs is that multiple sequential CID experiments can be performed without multiple analyzers. QIT analyzers are now capable of routinely analyzing an m/z value of up to 3000, which is useful because ESI of biomolecules commonly produces a charge distribution below m/z 3000. A disadvantage of the QITMS is the existence of space-charge effects. Space-charge effects are a consequence of the mutual repulsion between particles of like charge. The ion trap itself has a limited capacity for ions of all masses, thus resulting in possible ion suppression of polydisperse biomolecules.

2.6.5. Hybrid Ion-Trap FT and Ion-Trap Orbitrap MS

The Orbitrap mass spectrometer was developed in 2005 by Alexander Makarov (93) and is considered to be a modified "Kingdon" trap with an outer barrel-like electrode and an inner spindle-like electrode (94). The commercial hybrid LTQ Orbitrap™ mass spectrometer (Thermo Scientific, Waltham, MA) consists of a hybrid mass spectrometer that combines a linear ion trap (LTQ) and an Orbitrap FT mass analyzer (95). The hybrid LTQ Orbitrap consists of three main components: (1) a linear ion trap (LTQ) for sample ionization, selection, and fragmentation; (2) a radio frequency (RF) only C-trap, which is a curved linear trap that acts as an intermediate storage device for ions; and (3) an Orbitrap mass analyzer for FT-based analysis. The Orbitrap MS is a high-resolution, high-sensitivity, and high mass accuracy instrument that is advantageous for the analysis of proteins, peptides, carbohydrates, and small molecules.

The front part of this mass spectrometer is an LTQ ion trap that is capable of detecting MS and MS^n spectra at high sensitivity. It is consistent with most ion traps in that it has low resolution and low mass accuracy. The linear ion trap accumulates and isolates ions and fragments them in the mass range below m/z 2000. Ions accumulated in the LTQ can be transferred to the C-trap, which is an RF-only quadrupole that has the ability to accumulate and store ions. Ions are collisionally dampened by nitrogen in the C-trap, causing them to come to rest in the center of the trap. From this storage position, ions can be injected into the Orbitrap FT mass analyzer using high-voltage electric pulses. By lowering the electric potential of the inner electrode, ions are trapped in the Orbitrap and begin to cycle around the inner electrode in rings, while also moving back and forth along the axis of the electrode. Therefore, ions having a specific m/z move in rings which oscillate along the electrode spindle. The frequency of these harmonic oscillations is independent of the ion velocity and is instead dependent on the m/z ratio of the specific ions. The Orbitrap, an FT mass analyzer with a high dynamic range, can then be

used to detect ions with high mass accuracy (1–2 ppm), high resolution (up to 200,000), and high sensitivity (95).

Tandem mass spectrometry (MS/MS) enables both the structural elucidation and quantitative analysis of molecules, such as sulfated glycans, by fragmenting specific sample ions inside the mass spectrometer (96). Tandem MS of carbohydrates was first studied in detail from FAB mass spectra (97). Neutral carbohydrates ionized poorly, while acidic carbohydrates produced strong ion signal using FAB; however, permethylated and peracetylated oligosaccharides yielded increased signal (97). The FAB ionization process produced glycosidic and cross-ring cleavages for homogeneous molecules and became even more useful with the development of mass analyzers capable of selecting individual precursor ions and subsequent fragmentation. Subsequently, ESI increased ion signal strength of oligosaccharides in the tandem MS dimension allowing for the direct analysis of native oligosaccharides (98). ESI tandem MS of native glycans results in multiple charging and the formation of glycosidic bond cleavages.

2.6.6. Collision-Induced Dissociation (CID)

CID is a technique whereby selected precursor ions are made to undergo collision with a neutral gas to produce controlled fragmentations. CID is a slow heating process that causes the fragmentation of chemical bonds based on their lability. It is accomplished by selecting an ion of interest with a mass analyzer, such as a QIT analyzer, and increasing the internal energy by colliding it with inert gas molecules, typically argon or helium. The collisions result in an increase in internal energy to the point of bond scission. The resulting fragments are separated according to their m/z ratio and detected. Fragments will only be detected if they carry a minimum of one charge. Fragment ions are labeled with uppercase letters. Those that contain a non-reducing terminus are classified as A_n, B_n, or C_n. Fragment ions that contain the reducing end of the oligosaccharide are labeled with letters from the end of the alphabet and are classified as X_n, Y_n, or Z_n. The resulting tandem mass spectrum reveals the fragmentation patterns of a specific molecular ion. The QIT is particularly useful for the analysis of oligosaccharides because of its efficiency in acquiring multiple stages of tandem MS fragmentation.

2.6.7. Fragmentation Patterns

CID fragmentation patterns differ according to the oligosaccharide sequence, size and class of subunits and reflect the structural differences between compounds. Tandem mass spectra of GAGs result in fragment ions whose m/z values often correspond to either glycosidic bond cleavages or cross-ring cleavages. With low-energy CID, fragmentation of glycosidic linkages is most likely, whereas cross-ring fragmentation is less likely because two covalent bonds must be cleaved. Cross-ring cleavage does occur and provides important linkage information; however, the

more common of these two is glycosidic bond fragmentation (B_n, C_n, Y_n, Z_n), which yields losses corresponding to the elimination of monosaccharide residue masses from the precursor ion. It is indicative of sequence but not linkage and is charge-induced. Cross-ring cleavages (A_n and X_n), on the other hand, indicate linkages and are considered to be charge-remote fragmentations (99). The ion abundances of each fragment in the tandem mass spectrum may be calculated and used as quantitative tools when analyzing mixtures of glycans.

2.7. Nuclear Magnetic Resonance

NMR remains the most structurally informative spectroscopic technique for carbohydrate structural analysis, especially when combined with MS and monosaccharide analyses (92, 93). Carbohydrates have two natural NMR-active nuclei: ^{13}C and ^{1}H. Using classical NMR techniques ~1 mg of a trisaccharide is required for complete structural assignment using NMR. It is desirable to study smaller quantities, and much effort has been made in this direction over the past decade.

The classical approach for NMR carbohydrate analysis is summarized as follows (100). The number of monosaccharide residues may be determined by integrating the anomeric resonances occurring in the chemical shift range of 4.4–5.5 ppm. The monosaccharide residues may be identified using total correlation spectroscopy (TOCSY) and related techniques. Anomeric configurations may be determined using vicinal coupling constants between H1 and H2 and one bond ^{13}C–^{1}H coupling constants. Linkage and sequence information may be derived from ^{1}H and ^{13}C chemical shifts in cases where these values have been reported for specific linkages. Inter-residue nuclear Overhauser effects (NOEs) may also be used to provide information about glycosidic linkages. The position of substituent groups (acetate, methyl, sulfate, or phosphate) affects resonances of neighboring proton and carbon chemical shifts. Many of the carbohydrate chemical shifts are found in a narrow range, and software programs aid in the assignment of multidimensional NMR spectra. Enrichment of ^{13}C increases the sensitivity of multinuclear NMR to enable resolution of crowded regions of spectra. Such enrichment may be accomplished for HexNAc containing carbohydrates using chemical exchange of acetate groups (101). This approach has been applied to the conformational preferences of chondroitin sulfate oligomers (102).

The degree of conformational flexibility of carbohydrates is an important consideration that is amenable to analysis by NMR. Conformational models of oligosaccharides that are relatively rigid may be built using nuclear Overhauser effects (NOEs) spectroscopic data. Some oligosaccharides fail to produce a single conformational model using NOE (103). In these cases modeling based on scalar coupling offers a strategy to localize the

conformational flexibility in the oligosaccharide structure. This approach has been used to compare the conformation and dynamics of polysaccharides with those of an isolated oligosaccharide (104). For glycosaminoglycans, the presence of iduronic acid residues confers a degree of conformational flexibility relative to the glucuronic acid epimers. More than one ring conformation is possible for iduronic acid, changing dihedral angles and H–H distances that may be monitored using coupling constants and NOEs (105). This flexibility is key to understanding heparin–protein interactions. Dissolving biomolecules in oriented media, such as liquid crystals, serves to preserve information on dipolar couplings for which internuclear distances are known. This enables the use of such dipolar couplings for conformational analysis of oligosaccharides (106).

The biological activities of carbohydrates arise through their interactions with proteins in various contexts. Binding affinities for many carbohydrate classes range from the micromolar to millimolar range, although nanomolar affinities are possible for glycosaminoglycans (107). NMR experiments that take advantage of fast exchange have been used to reveal carbohydrate conformations in protein binding interactions, as described in recent reviews (108–111). NMR techniques often used for observation of high-affinity ligand–protein binding are typically not used for studies of interactions involving carbohydrates. In many cases, experiments focus on analysis of resonances associated with the carbohydrate ligand in both free and protein-bound forms. Two NMR experiment types have been widely applied to carbohydrate–protein interactions. Transferred nuclear Overhauser effects (tr-NOEs) produce data that describe the protein-bound conformations of carbohydrates. Saturation transfer difference (STD) NMR spectroscopy generates information about the thermodynamics and kinetics of carbohydrate–protein binding.

2.8. Microarrays

Carbohydrate microheterogeneity is a challenge to efforts to determine structures responsible for observed biological activities. Such activities include protein binding or binding inhibition. Therefore, there is substantial interest in probing the biological activities of well-characterized, monodisperse compounds for development of carbohydrate-based vaccines, drugs, and therapeutics (112). Until relatively recently, the only way to obtain a pure structurally homogeneous carbohydrate was through multistage isolation from a biological source. The effort required to achieve such isolation severely limited the amount of information available for establishing structure–function relationships. Advances in chemical synthesis over the past decade have now made it possible to produce targeted glycans (112, 113). These

approaches have enabled the development and use of carbohydrate microarrays (112, 114–118).

Immobilized carbohydrate microarrays serve as a means to probe the binding properties of candidate proteins. Many carbohydrate ligands cross-react with several binding proteins, and it is essential to have an array of immobilized ligands to use to develop conclusions based on multiple interactions. The carbohydrate immobilization chemistry and detection system are key to the effective application of microarrays to answer questions in glycobiology. Because of the relatively low affinity of many carbohydrate–protein interactions, high avidity is achieved through multivalent interactions. Monomeric oligosaccharides in solution may be used to inhibit multivalent interactions between a binding protein and surface-immobilized oligosaccharide ligands.

Reducing oligosaccharides may be reductively aminated using an aminolipid to form a neoglycolipid (NGL) with amphipathic properties (114, 119). Oxime ligation is an alternative chemistry for NGL formation (120). This approach may also be used on reduced glycan alditols after mild periodate oxidation. The hydrophobic lipid tail facilitates immobilization of the NGLs on silical gel surfaces, plastic microtiter plates, membranes, and microscope slides. NGLs may be separated using silica gel chromatography as needed. Oligosaccharides such as NGLs have been used to construct microarrays for determination of biotin-conjugated protein binding with detection using fluorescently labeled streptavidin.

Standard robotic printing technology has been used to create an array of greater than 200 synthetic and natural structurally defined glycoprotein and glycolipid glycans (115). The glycans were amine functionalized to permit immobilization onto glass slides functionalized with N-hydroxysuccinimide. Interactions with the microarray were detected either using fluorescently labeled glycan-binding protein or labeled secondary antibodies. This approach has been applied to some of the major classes of glycan-binding proteins, such as C-type lectins, galectins, siglecs, plant and viral lectins, and antibodies. A number of pathogen-derived lipopolysaccharide-derived carbohydrate antigens have been used to construct microarrays using a bifunctional linker to amine-reactive slides (117). The arrays provide a modern serological test for the presence of *salmonella* antibodies. An assay using similar technology has been developed to survey the host specificity of influenza viruses (121).

Surface-immobilized carbohydrate binding proteins of known specificities may also be used to inquire the carbohydrate epitopes present in a given biological samples. In particular, lectins (122) may be used for this purpose (123–125). There are approximately 60 commercially available lectins, the

majority of which interact with mammalian glycoconjugate glycans. The glycans are labeled with fluorescent groups to enable their detection. One advantage of the lectin array approach for glycomics analysis is that it is possible to follow dynamic changes to glycan expression as a function of biological variables. Toward these ends, a ratiometric approach has been developed in which glycans released from two samples are labeled with different fluorophores and then mixed prior to binding to the lectin array (126). This approach demonstrates high accuracy and reproducibility for the use of lectin microarrays. Microarrays have also been used to determine binding of whole cells to specific lectins (123). Using evanescent-field fluorescence detection, the excitation wavelength is applied only within a wavelength (100–200 nm) of the microarray surface. As a result, it is possible to detect ligand–receptor interactions under equilibrium conditions. This approach has been applied to a microarray of 39 lectins using fluorescently labeled glycoproteins (127).

3. Conclusions

The complexities of modern glycoanalytical techniques reflect an evolving appreciation for the roles of glycan expression in biological systems. While the number of proteins expressed in the genome is limited, their functions are elaborated through dynamic alterations to a diverse set of glycoconjugate glycans. Thus, to understand, for example, how cancer cells change their immunological and adhesive properties to evade normal growth controls, it is necessary to determine glycan expression at several stages. Researchers, therefore, must employ a combination of techniques to develop a comprehensive understanding of the roles of glycan expression in disease processes.

Use of exoglycosidase enzymes and/or chemical derivatization methods increases the effectiveness of instrumental glycoanalysis techniques. NMR, in combination with MS and compositional analysis, produces complete definition of the structure of a purified glycan. LC–MS analysis defines the compositions of the distribution of glycans in a biological sample at a given time, but does not determine all structural variables. Microarrays lend themselves to high sensitivity and throughput but of a limited set of glycan–protein interactions. It has been the intent of this overview to introduce and summarize these analytical approaches. The need for several techniques to be used in combination with one another may be required for complete glycoanalysis.

Acknowledgments

The authors were supported by NIH grant P41RR10888 and NIH contract NO1HV28178.

References

1. Spiro, R. G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R–56R.
2. Sasisekharan, R., Raman, R., and Prabhakar, V. (2006) Glycomics approach to structure-function relationships of glycosaminoglycans. *Annu. Rev. Biomed. Eng.* **8**, 181–231.
3. Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, G. (1999) *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
4. Abeijon, C. and Hirschberg, C. B. (1987) Subcellular site of synthesis of the N-acetylgalactosamine (alpha 1-0) serine (or threonine) linkage in rat liver. *J. Biol. Chem.* **262,** 4153–4159.
5. Amerongen, A. V., Bolscher, J. G., and Veerman, E. C. (1995) Salivary mucins: protective functions in relation to their diversity. *Glycobiology* **5,** 733–740.
6. Stults, C. L., Sweeley, C. C., and Macher, B. A. (1989) Glycosphingolipids: structure, biological source, and properties. *Methods Enzymol.* **179,** 167–214.
7. Hakomori, S., and Igarashi, Y. (1995) Functional role of glycosphingolipids in cell recognition and signaling. *J. Biochem. (Tokyo)* **118,** 1091–1103.
8. Kawashima, H., Hirose, M., Hirose, J., Nagakubo, D., Plaas, A. H., and Miyasaka, M. (2000) Binding of a large chondroitin sulfate/dermatan sulfate proteoglycan, versican, to L-selectin, P-selectin, and CD44. *J. Biol. Chem.* **275,** 35448–35456.
9. Lander, A. D. (1998) Proteoglycans: master regulators of molecular encounter? *Matrix Biol.* **17,** 465–472.
10. Trowbridge, J. M., Rudisill, J. A., Ron, D., and Gallo, R. L. (2002) Dermatan sulfate binds and potentiates activity of keratinocyte growth factor (FGF-7). *J. Biol. Chem.* **277,** 42815–42820.
11. Torres, C. R., and Hart, G. W. (1984) Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J. Biol. Chem.* **259,** 3308–3317.
12. Hart, G. W., Haltiwanger, R. S., Holt, G. D., and Kelly, W. G. (1989) Glycosylation in the nucleus and cytoplasm. *Annu. Rev. Biochem.* **58,** 841–874.
13. Hart, G. W., Kreppel, L. K., Comer, F. I., Arnold, C. S., Snow, D. M., Ye, Z., Cheng, X., DellaManna, D., Caine, D. S., Earles, B. J., Akimoto, Y., Cole, R. N., and Hayes, B. K. (1996) O-GlcNAcylation of key nuclear and cytoskeletal proteins: reciprocity with O-phosphorylation and putative roles in protein multimerization. *Glycobiology* **6,** 711–716.
14. Hart, G. W., Housley, M. P., and Slawson, C. (2007) Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446,** 1017–1022.
15. Wong-Madded, S. T., Landry, D., and Guthrie, E. P. (1997) Discovery and Uses of Novel Glycosidases, in *Techniques in Glycobiology* (Townsend, R., and Hotchkiss, A. T., Eds.) pp. 401–408, Marcel Dekker, Inc., New York.
16. Tyagarajan, K., Forte, J. G., and Townsend, R. R. (1996) Exoglycosidase purity and linkage specificity: assessment using oligosaccharide substrates and high-pH anion-exchange chromatography with pulsed amperometric detection. *Glycobiology* **6,** 83–93.
17. Rudd, P. M., Colominas, C., Royle, L., Murphy, N., Hart, E., Merry, A. H., Hebestreit, H. F., and Dwek, R. A. (2001) A high-performance liquid chromatography based strategy for rapid, sensitive sequencing of N-linked oligosaccharide modifications to proteins in sodium dodecyl sulphate polyacrylamide electrophoresis gel bands. *Proteomics* **1,** 285–294.
18. Garner, B., Merry, A. H., Royle, L., Harvey, D. J., Rudd, P. M., and Thillet, J. (2001) Structural elucidation of the N- and O-glycans of human apolipoprotein(a): role of o-glycans in conferring protease resistance. *J. Biol. Chem.* **276,** 22200–22208.
19. Royle, L., Mattu, T. S., Hart, E., Langridge, J. I., Merry, A. H., Murphy, N., Harvey, D. J., Dwek, R. A., and Rudd, P. M. (2002) An analytical and structural database provides

a strategy for sequencing O-glycans from microgram quantities of glycoproteins. *Anal. Biochem.* **304,** 70–90.
20. Wong-Madden, S. T., and Landry, D. (1995) Purification and characterization of novel glycosidases from the bacterial genus Xanthomonas. *Glycobiology* **5,** 19–28.
21. Starr, C. M., Irene Masada, R., Hague, C., Skop, E., and Klock, J. C. (1996) Fluorophore-assisted carbohydrate electrophoresis in the separation, analysis, and sequencing of carbohydrates. *J. Chromatogr. A* **720 ,** 295–321.
22. Harvey, D. J., Wing, D. R., Küster, B., and Wilson, I. B. (2000) Composition of N-linked carbohydrates from ovalbumin and copurified glycoproteins. *J. Am. Soc. Mass Spectrom.* **11,** 564–571.
23. Ninonuevo, M. R., Park, Y., Yin, H., Zhang, J., Ward, R. E., Clowers, B. H., German, J. B., Freeman, S. L., Killeen, K., Grimm, R., and Lebrilla, C. B. (2006) A strategy for annotating the human milk glycome. *J. Agric. Food Chem.* **54,** 7471–7480.
24. Anumula, K. R. (2000) High-sensitivity and high-resolution methods for glycoprotein analysis. *Anal. Biochem.* **283,** 17–26.
25. Anumula, K. R. (2006) Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates. *Anal. Biochem.* **350,** 1–23.
26. Hase, S. (1996) Precolumn derivatization for chromatographic and electrophoretic analyses of carbohydrates. *J. Chromatogr. A* **720,** 173–182.
27. Anumula, K. R. (2006) Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates. *Anal. Biochem.* **350,** 1–23.
28. Anumula, K. R. (1994) Quantitative determination of monosaccharides in glycoproteins by high-performance liquid chromatography with highly sensitive fluorescence detection. *Anal. Biochem.* **220,** 275–283.
29. Bigge, J. C., Patel, T. P., Bruce, J. A., Goulding, P. N., Charles, S. M., and Parekh, R. B. (1995) Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Anal. Biochem.* **230,** 229–238.
30. Hase, S. (1994) High-performance liquid chromatography of pyridylaminated saccharides. *Methods Enzymol.* **230,** 225–237.
31. Birrell, H., Charlwood, J., Lynch, I., North, S., Camilleri, P. (1999) A dual-detection strategy in the chromatographic analysis of 2-aminoacridone derivatized oligosaccharides. *Anal. Chem.* **71,** 102–108.
32. Jackson, P. (1994) The analysis of fluorophore-labeled glycans by high-resolution polyacrylamide gel electrophoresis. *Anal. Biochem.* **216,** 243–252.
33. Stefansson, M., and Novotny, M. (1994) Modification of the electrophoretic mobility of neutral and charged polysaccharides. *Anal. Chem.* **66,** 3466–3471.
34. Cheng, M. C., Lin, S.L., Wu, S.H., Inoue, S., Inoue, Y. (1998) High-Performance capillary electrophoretic characterization of different types of oligo- and polysialic acid chains. *Anal. Chem.* **260**, 154–159.
35. Nashabeh, W. and el Rassi, Z. (1992) Capillary zone electrophoresis of linear and branched oligosaccharides. *J. Chromatogr.* **600,** 279–287.
36. Anumula, K. R., and Dhume, S. T. (1998) High resolution and high sensitivity methods for oligosaccharide mapping and characterization by normal phase high performance liquid chromatography following derivatization with highly fluorescent anthranilic acid. *Glycobiology* **8,** 685–694.
37. Ciucanu, I., and Kerek, F. (1984) A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **131,** 209–217.
38. Ciucanu, I., and Costello, C. E. (2003) Elimination of oxidative degradation during the per-O-methylation of carbohydrates. *J. Am. Chem. Soc.* **125,** 16213–16219.
39. Kang, P., Mechref, Y., Klouckova, I., and Novotny, M. V. (2005) Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun. Mass Spectrom.* **19,** 3421–3428.
40. Kang, P., Mechref, Y., and Novotny, M. V. (2008) High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Commun. Mass Spectrom.* **22,** 721–734.
41. Ashline, D., Singh, S., Hanneman, A., and Reinhold, V. (2005) Congruent Strategies for Carbohydrate Sequencing. 1. Mining Structural Details by MS^n. *Anal. Chem.* **77,** 6250–6262.
42. Hardy, M. R. (1989) Monosaccharide analysis of glycoconjugates by high-performance anion-exchange chromatography with pulsed amperometric detection. *Methods Enzymol.* **179,** 76–82.
43. Gey, M. H., and Unger, K. K. (1996) A strategy for chromatographic and structural analysis of monosaccharide species from glycoproteins. *Anal. Bioanal. Chem.* **356,** 488–494.

44. Karlsson, N. G., and Hansson, G. C. (1995) Analysis of monosaccharide composition of mucin oligosaccharide alditols by high-performance anion-exchange chromatography. *Anal. Biochem.* **224,** 538–541.
45. Guttman, A. (1997) Analysis of monosaccharide composition by capillary electrophoresis. *J. Chromatogr. A* **763,** 271–277.
46. Lindberg, B., and Lonngren, J. (1978) Methylation analysis of complex carbohydrates: general procedure and application for sequence analysis. *Methods Enzymol.* **50,** 3–33.
47. Zaia, J., and Costello, C. E. (2001) Compositional analysis of glycosaminoglycans by electrospray mass spectrometry. *Anal. Chem.* **73,** 233–239.
48. Linhardt, R. J., Gu, K. N., Loganathan, D., and Carter, S. R. (1989) Analysis of glycosaminoglycan-derived oligosaccharides using reversed-phase ion-pairing and ion-exchange chromatography with suppressed conductivity detection. *Anal. Biochem.* **181,** 288–296.
49. Harvey, D. J. (2000) Electrospray mass spectrometry and fragmentation of N-linked carbohydrates derivatized at the reducing terminus. *J. Am. Soc. Mass Spectrom.* **11,** 900–915.
50. Charlwood, J., Birrell, H., Gribble, A., Burdes, V., Tolson, D., and Camilleri, P. (2000) A probe for the versatile analysis and characterization of N-linked oligosaccharides. *Anal. Chem.* **72,** 1453–1461.
51. Garcia, M. C. (2005) The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography-electrospray mass spectrometry. *J. Chromatogr. B* **825,** 111–123.
52. Kuberan, B., Lech, M., Zhang, L., Wu, Z. L., Beeler, D. L., and Rosenberg, R. (2002) Analysis of Heparan Sulfate Oligosaccharides with Ion Pair-Reverse Phase Capillary High Performance Liquid Chromatography-Microelectrospray Ionization Time-of-Flight Mass Spectrometry. *J. Am. Chem. Soc.* **124**, 8707–8718.
53. Thanawiroon, C., Rice, K. G., Toida, T., and Linhardt, R. J. (2004) Liquid chromatography–mass spectrometry sequencing approach for highly sulfated heparin-derived oligosaccharides. *J. Biol. Chem.* **279,** 2608–2615.
54. Henriksen, J., Roepstorff, P., and Ringborg, L. H. (2006) Ion-pairing reversed-phased chromatography–mass spectrometry of heparin. *Carbohydr. Res.* **341,** 382–387.
55. Koizumi, K. (1996) High-performance liquid chromatographic separation of carbohydrates on graphitized carbon columns. *J. Chromatogr. A* **720,** 119–126.
56. Davies, M., Smith, K. D., Harbin, A. M., and Hounsell, E. F. (1992) High-performance liquid chromatography of oligosaccharide alditols and glycopeptides on a graphitized carbon column. *J. Chromatogr.* **609,** 125–131.
57. Karlsson, N. G., Wilson, N. L., Wirth, H. J., Dawes, P., Joshi, H., and Packer, N. H. (2004) Negative ion graphitised carbon nano-liquid chromatography–mass spectrometry increases sensitivity for glycoprotein oligosaccharide analysis. *Rapid Commun. Mass Spectrom.* **18,** 2282–2292.
58. Thomsson, K. A., Karlsson, N. G., and Hansson, G. C. (1999) Liquid chromatography-electrospray mass spectrometry as a tool for the analysis of sulfated oligosaccharides from mucin glycoproteins. *J. Chromatogr. A* **854,** 131–139.
59. Schulz, B. L., Packer, N. H., and Karlsson, N. G. (2002) Small-scale analysis of O-linked oligosaccharides from glycoproteins and mucins separated by gel electrophoresis. *Anal. Chem.* **74,** 6088–6097.
60. Ziegler, A., and Zaia, J. (2006) Separation of heparin oligosaccharides by size-exclusion chromatography. *J. Chromatogr. B* **837,** 76–86.
61. Blom, K. F., Larsen, B. S., and McEwen, C. N. (1999) Determining affinity-selected ligands and estimating binding affinities by online size exclusion chromatography/liquid chromatography-mass spectrometry. *J. Comb. Chem.* **1,** 82–90.
62. Van Deemter, J. J., Zuiderweg, F.J., Klinkenberg, A. (1956) Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chem. Eng. Sci.* **5,** 271–289.
63. Alpert, A. (1990) Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A* **499,** 177–196.
64. Churms, S. C. (1996) Recent progress in carbohydrate separation by high-performance liquid chromatography based on hydrophilic interaction. *J. Chromatogr. A* **720,** 75–91.
65. Guile, G. R., Rudd, P. M., Wing, D. R., Prime, S. B., and Dwek, R. A. (1996) A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analyzing oligosaccharide profiles. *Anal. Biochem.* **240**, 210–226.
66. Zamfir, A., and Peter-Katalinic, J. (2004) Capillary electrophoresis-mass spectrometry for glycoscreening in biomedical research. *Electrophoresis* **25**, 1949–1963.

67. Gennaro, L. A., Salas-Solano, O., and Ma, S. (2006) Capillary electrophoresis-mass spectrometry as a characterization tool for therapeutic proteins. *Anal. Biochem.* **355,** 249–258.
68. al-Hakim, A. and Linhardt, R. J. (1991) Capillary electrophoresis for the analysis of chondroitin sulfate- and dermatan sulfate-derived disaccharides. *Anal. Biochem.* **195,** 68–73.
69. Pervin, A., al-Hakim, A., and Linhardt, R. J. (1994) Separation of glycosaminoglycan-derived oligosaccharides by capillary electrophoresis using reverse polarity. *Anal. Biochem.* **221,** 182–188.
70. Lamari, F. N., Militsopoulou, M., Mitropoulou, T. N., Hjerpe, A., and Karamanos, N. K. (2002) Analysis of glycosaminoglycan-derived disaccharides in biologic: samples by capillary electrophoresis and protocol for sequencing glycosaminoglycans. *Biomed. Chromatogr.* **16,** 95–102.
71. Militsopoulou, M., Lamari, F. N., Hjerpe, A., and Karamanos, N. K. (2002) Determination of twelve heparin- and heparan sulfate-derived disaccharides as 2-aminoacridone derivatives by capillary zone electrophoresis using ultraviolet and laser-induced fluorescence detection. *Electrophoresis* **23,** 1104–1109.
72. Militsopoulou, M., Lecomte, C., Bayle, C., Couderc, F., and Karamanos, N. K. (2003) Laser-induced fluorescence as a powerful detection tool for capillary electrophoretic analysis of heparin/heparan sulfate disaccharides. *Biomed. Chromatogr.* **17,** 39–41.
73. Carr, S. A., and Reinhold, V. N. (1984) Structural characterization of sulfated glycosaminoglycans by fast atom bombardment mass spectrometry: application to chondroitin sulfate. *J. Carbohydr. Chem.* **3**, 381–401.
74. Ii, T., Okuda, S., Hirano, T., and Ohashi, M. (1994) Tandem mass spectrometry for characterization of unsaturated disaccharides from chondroitin sulfate, dermatan sulfate and hyaluronan. *Glycoconj. J.* **11,** 123–132.
75. Juhasz, P., and Biemann, K. (1994) Mass spectrometric molecular-weight determination of highly acidic compounds of biological significance via their complexes with basic polypeptides. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 4333–4337.
76. Chai, W., Luo, J., Lim, C. K., and Lawson, A. M. (1998) Characterization of heparin oligosaccharide mixtures as ammonium salts using electrospray mass spectrometry. *Anal. Chem.* **70**, 2060–2066.
77. Kim, Y. S., Ahn, M. Y., Wu, S. J., Kim, D. H., Toida, T., Teesch, L. M., Park, Y., Yu, G., Lin, J., and Linhardt, R. J. (1998) Determination of the structure of oligosaccharides prepared from acharan sulfate. *Glycobiology* **8,** 869–877.
78. Peter-Katalinić, J. (1994) Analysis of glycoconjugates by fast atom bombardment mass spectrometry and related MS techniques. *Mass Spectrom. Rev.* **13**, 77–98.
79. Domon, B., and Costello, C. E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **5,** 397–409.
80. Huberty, M. C., Vath, J. E., Yu, W., and Martin, S. A. (1993) Site-specific carbohydrate identification in recombinant proteins using MALD-TOF MS. *Anal. Chem.* **65,** 2791–2800.
81. Karas, M., and Kruger, R. (2003) Ion formation in MALDI: the cluster ionization mechanism. *Chem. Rev.* **103,** 427–440.
82. Zaia, J. (2007) MassSpectrometric Ionization of Carbohydrates, in *Encyclopedia of Mass Spectrometry*, **vol 6** (Gross, M. G., Ed.) Elsevier, New York.
83. Sturiale, L., Naggi, A., and Torri, G. (2001) MALDI mass spectrometry as a tool for characterizing glycosaminoglycan oligosaccharides and their interaction with proteins. *Semin. Thromb. Hemost.* **27**, 465–472.
84. Fenn, J. B. (2003) *Angew. Chem. Int. Ed.* **42,** 3871–3874.
85. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246,** 64–71.
86. Cech, N. B., and Enke, C. G. (2001) Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.* **20,** 362–387.
87. Karas, M., Bahr, U., and Dulcks, T. (2000) Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius J. Anal. Chem.* **366,** 669–676.
88. Taylor, G. (1964) Disintegration of water droplets in an electric field. *Proc. R. Soc. London* **280,** 383–397.
89. Scott, C. T. J., Kosmidis, C., Jia, W. J., Ledingham, K. W. D., and Singhal, R. P. (1994) Formation of atomic-hydrogen in matrix-assisted laser-desorption ionization. *Rapid Commun. Mass Spectrom.* **8**, 829–832.
90. Yamashita, M., Fenn, J.B. (1984) Electrospray Ion-Source – Another Variation on the Free-Jet Theme. *J. Phy. Chem.* **88,** 4451–4459.
91. Wolfgang, P., et al. (1953) *Naturforsch A* 448–450.

92. Louris, J. N., Cooks, R. G., Syka, J. E. P., Taylor, D. M., Kelley, P. E., Stafford, G. C. Jr., and Todd, J. F. J. (1987) Instrumentation, applications, and energy deposition in quadrupole ion-trap tandem mass spectrometry. *Anal. Chem.* **59,** 1677–1685.
93. Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005) The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40,** 430–443.
94. Kingdon, K. H. (1923) A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phy. Rev.* **21,** 408–418.
95. Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4,** 2010–2021.
96. Desaire, H., and Leary, J. (2000) Detection and quantification of the sulfated disaccharides in chondroitin sulfate by electrospray tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **11**, 916–920.
97. Egge, H., and Peter-Katalinc, J. (1987) Fast atom bombardment mass spectrometry for structural elucidation of glycoconjugates. *Mass Spectrom. Rev.* **6,** 331–393.
98. Zaia, J. (2004) Mass spectrometry of oligosaccharides. *Mass Spectrom. Rev.* **23,** 161–227.
99. Domon, B., and Costello, C. E. (1988) Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry. *Biochemistry* **27**, 1534–1543.
100. Duus, J., Gotfredsen, C. H., and Bock, K. (2000) Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem. Rev.* **100,** 4589–4614.
101. Yu, F., and Prestegard, J. H. (2006) Structural monitoring of oligosaccharides through 13C enrichment and NMR observation of acetyl groups. *Biophys. J.* **91,** 1952–1959.
102. Yu, F., Wolff, J. J., Amster, I. J., and Prestegard, J. H. (2007) Conformational preferences of chondroitin sulfate oligomers using partially oriented NMR spectroscopy of 13C-labeled acetyl groups. *J. Am. Chem. Soc.* **129,** 13288–13297.
103. Martin-Pastor, M., and Bush, C. A. (1999) New strategy for the conformational analysis of carbohydrates based on NOE and 13C NMR coupling constants. Application to the flexible polysaccharide of Streptococcus mitis J22. *Biochemistry* **38,** 8045–8055.
104. Martin-Pastor, M., and Bush, C. A. (2000) Comparison of the conformation and dynamics of a polysaccharide and of its isolated heptasaccharide repeating unit on the basis of nuclear Overhauser effect, long-range C-C and C-H coupling constants, and NMR relaxation data. *Biopolymers* **54,** 235–248.
105. Mulloy, B., and Forster, M. J. (2000) Conformation and dynamics of heparin and heparan sulfate. *Glycobiology* **10,** 1147–1156.
106. Tian, F., Al-Hashimi, H. M., Craighead, J. L., and Prestegard, J. H. (2001) Conformational analysis of a flexible oligosaccharide using residual dipolar couplings. *J. Am. Chem. Soc.* **123,** 485–492.
107. Conrad, H. E. (1998) *Heparin Binding Proteins*, Academic Press, New York.
108. Angulo, J., Rademacher, C., Biet, T., Benie, A. J., Blume, A., Peters, H., Palcic, M., Parra, F., and Peters, T. (2006) NMR analysis of carbohydrate-protein interactions. *Methods Enzymol.* **416,** 12–30.
109. Kogelberg, H., Solis, D., and Jimenez-Barbero, J. (2003) New structural insights into carbohydrate-protein interactions from NMR spectroscopy. *Curr. Opin. Struct. Biol.* **13,** 646–653.
110. Brecker, L., Schwarz, A., Goedl, C., Kratzer, R., Tyl, C. E., and Nidetzky, B. (2008) Studying non-covalent enzyme carbohydrate interactions by STD NMR. *Carbohydr. Res.* **343**, 2153–2161.
111. Jones, C. (2005) NMR assays for carbohydrate-based vaccines. *J. Pharm. Biomed. Anal.* **38**, 840–850.
112. Seeberger, P. H., and Werz, D. B. (2007) Synthesis and medical applications of oligosaccharides. *Nature* **446 ,** 1046–1051.
113. Turnbull, J. E., and Linhardt, R. J. (2006) Synthetic sugars enhance the functional glycomics toolkit. *Nat. Chem. Biol.* **2**, 449–450.
114. Feizi, T., and Chai, W. (2004) Oligosaccharide microarrays to decipher the glyco code. *Nat. Rev. Mol. Cell. Biol.* **5**, 582–588.
115. Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., Bryan, M. C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D. J., Skehel, J. J., van Die, I., Burton, D. R., Wilson, I. A., Cummings, R., Bovin, N., Wong, C. H., and Paulson, J. C. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17033–17038.
116. Liang, P. H., Wu, C. Y., Greenberg, W. A., and Wong, C. H. (2008) Glycan arrays: bio-

logical and medical applications. *Curr. Opin. Chem. Biol.* **12,** 86–92.
117. Blixt, O., Hoffmann, J., Svenson, S., and Norberg, T. (2008) Pathogen specific carbohydrate antigen microarrays: a chip for detection of Salmonella O-antigen specific antibodies. *Glycoconj. J.* **25**, 27–36.
118. Feizi, T., Fazio, F., Chai, W., and Wong, C.-H. (2003) Carbohydrate microarrays – a new set of technologies at the frontiers of glycomics. *Curr. Opin. Struct. Biol.* **13,** 637–645.
119. Fukui, S., Feizi, T., Galustian, C., Lawson, A. M., and Chai, W. (2002) Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions. *Nat. Biotechnol.* **20**, 1011–1017.
120. Liu, Y., Chai, W., Childs, R. A., and Feizi, T. (2006) Preparation of neoglycolipids with ring-closed cores via chemoselective oxime-ligation for microarray analysis of carbohydrate-protein interactions. *Methods Enzymol.* **415,** 326–340.
121. Stevens, J., Blixt, O., Paulson, J. C., and Wilson, I. A. (2006) Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nat. Rev. Micro.* **4,** 857–864.
122. Cummings, R. (1994) Use of lectins in analysis of glycoconjugates. *Methods Enzymol.* **230**, 66–86.
123. Zheng, T., Peelen, D., and Smith, L. M. (2005) Lectin arrays for profiling cell surface carbohydrate expression. *J. Am. Chem. Soc.* **127,** 9982–9983.
124. Pilobello, K. T., and Mahal, L. K. (2007) Deciphering the glycocode: the complexity and analytical challenge of glycomics. *Curr. Opin. Chem. Biol.* **11**, 300–305.
125. Song, X., Xia, B., Lasanajak, Y., Smith, D., and Cummings, R. (2008) Quantifiable fluorescent glycan microarrays. *Glycoconj. J.* **25**, 15–25.
126. Pilobello, K. T., Slawek, D. E., and Mahal, L. K. (2007) A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 11534–11539.
127. Kuno, A., Uchiyama, N., Koseki-Kuno, S., Ebe, Y., Takashima, S., Yamada, M., and Hirabayashi, J. (2005) Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nat. Methods* **2**, 851–856.

Chapter 3

Quantitative Glycomics

Ron Orlando

Abstract

The ability to quantitatively determine changes is an essential component of comparative glycomics. Multiple strategies are available by which this can be accomplished. These include label-free approaches and strategies where an isotopic label is incorporated into the glycans prior to analysis. The focus of this chapter is to describe each of these approaches while providing insight into their strengths and weaknesses, so that glycomic investigators can make an educated choice of the strategy that is best suited for their particular application.

Key words: Comparative glycomics, relative quantitation, isotopic labeling, label free, mass spectrometry (MS).

1. Introduction

Mass spectrometry is a powerful tool for qualitative glycomics; however, several issues arise when this technique is used to obtain quantitative results. Because quantitative glycomics is a relatively new field, it is difficult to evaluate the benefits and limitations of the approaches that have been developed in this area. However, many of these techniques can be directly compared to strategies that have been used for many years in the proteomics field, which also relies heavily upon mass spectrometry. Thus, it seems appropriate to begin with a review of the approaches used in quantitative proteomics, including a summary of the lessons learned in proteomics to gain insight into the strategies used for quantitative glycomics. This will be followed by a description of the quantitative approaches developed for glycomics. This later section

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_3,

primarily focuses on the analysis of glycoprotein glycans; however, many of these strategies can be applied to other types of glycans.

It is important to note at the onset that this chapter focuses on relative quantitative glycomics, i.e., how do the levels of individual glycans change between samples. Determining the quantity of each individual glycan in a sample (absolute quantitation) has not yet been widely applied to glycomics. One reason that absolute quantitation is difficult if not impossible at the current time is that the response for any analyte is going to depend directly on the analyte's ionization efficiency, which in turn is related to numerous factors, such as molecular mass, proton/cation affinity, and surface activity. All of these are going to differ between analytes. Other issues encountered in quantitative MS experiments relate to variability in the sample matrix, in the instrument response over time, and in losses during sample preparations; all of these will be discussed in the following paragraph. For these reasons, absolute quantitation typically entails the addition of a known quantity of an isotopically labeled version of each analyte under investigation. Unfortunately, standard isotopically labeled glycans of known quantity are not widely available, which in turn has limited the ability to perform absolute quantitation in the glycomics field.

Strategies for relative quantitation with MS-based techniques address errors introduced by variability in the sample matrix, the instrument response, instrument to instrument performance, and/or the sample preparation process. Matrix effects, which are often attributed to phenomenon such as ion suppression from the presence of other compounds competing with or interfering with the ionization of the analyte, can alter the response from a particular analyte even when the analyte's concentration does not change. Matrix effects are one of the reasons that ion intensities often do not directly correlate with concentration. Variability in the instrument response further exacerbates the issue of relative quantitation, as it leads to differences in ion abundances between analyses. Of course, these differences are somewhat related to the time between experiments, and thus glycans that are observed in the same spectrum are less prone to this error than glycans that are observed at different time points, such as those that elute at different times in an LC separation. Variability in instrument performance is particularly problematic when long time periods are present between the times when samples are analyzed. Furthermore, different MS systems can produce different ion intensities and different ratios of ion abundances from the same sample. This is particularly true when different MS configurations from different vendors are compared, and thus this type of error can limit cross-laboratory reproducibility. One additional source of quantitative error results from differential analyte losses occurring when samples are treated separately in a parallel manner. The success of

the different approaches for relative quantitation depends on how well each of these sources of error is addressed.

The techniques used in proteomics for relative quantitation can be broadly subdivided into two general schemes – those that involve the use of labels and those methods that are label free. In the label-free approaches, various aspects of the peptides/proteins such as normalized ion intensities, spectral counts, mass, scan number and signal intensity, and accurate mass plus retention time have been successfully used to assign protein expression levels for comparative investigations (1–4). In some of these label-free approaches, the analytes themselves serve as standards. In this way the response of an analyte is compared with the response from one, several, or all other species present in the sample. With spectral counting, for instance, the number of MS/MS spectra from a particular protein can be expressed as a fraction of all MS/MS spectra acquired for all proteins in the sample. With other methods, one or several of the proteins are assigned as being constant between the samples analyzed, and all results are made relative to these reference species. In this manner, errors associated with changes in instrument performance are decreased since all proteins are analyzed in the same sample, and thus subjected to the same instrument performance. The net result is that this strategy improves the reproducibility of these experiments. However, this is not a perfect solution for variation in instrument performance since there is often sufficient time between acquisition of the analyte and standard signals to allow for altered instrument response, particularly when an LC separation is used. Errors larger than 25% have been reported when the analyte and reference only partially overlap chromatographically (5); larger errors can be expected as the time difference between analyte and reference elution increases. When the analytes serve as their own standards, it can be difficult to deduce if an increased response for a particular protein is due to it being up-regulated or if this increased prevalence is actually caused by a decrease in the abundance of another protein in the sample, since both of these lead to the same result. These approaches do not compensate for issues associated with matrix effects if analytes are separated chromatographically, variability in instrument to instrument performance, and errors resulting from differential losses from parallel sample preparations. For these reasons, many of these approaches are considered to be only semi-quantitative in nature and often are attributed to have the inability to reproducibly detect changes in protein expression that are smaller than 2-fold (6–8). A primary advantage of these approaches is their simplicity, as these do not alter the sample workflow; however, this does add additional data processing steps. Also, these strategies do not need stable isotopes, which elevate the cost of the other internal standard approaches that will be discussed. Label-free approaches thus offer a

straightforward manner to obtain quantitative results and are often used as a "screening process" to identify proteins of interest to serve as the subjects of more detailed studies using other methods.

The use of internal standards is the other general strategy to compensate for the problems associated with quantification by MS. However, the type of internal standard used and when it is introduced into the sample determines if it is capable of compensating for variability in the sample, instrument variability, and/or sample preparation. Consequently, not all internal standard approaches are created equal, which in turn has led to the development of multiple strategies, each of which has its own unique benefits and limitations. In general, the closer the chemical properties of an internal standard to its analyte, the better it compensates for the various sources of error, and thus the optimal internal standard for each analyte is typically an isotopomer of the analyte itself (9, 10). For example, a sample could be mixed with an internal standard consisting of an isotopically labeled (^{13}C, D, ^{15}N, etc.) form of each analyte followed by MS analysis. The mass analyzer resolves the isotopomers, permitting their relative abundances to be determined by comparing the signal intensity of each analyte ion to that from its isotopically labeled form.

It is extremely difficult and/or expensive to obtain isotopically labeled standards for all species observed in a high-throughput proteomic analysis of complex samples, hence a variety of isotopic labeling procedures have been developed where one of the samples is modified with a "light" tag while the other is derivatized with a "heavy" tag (11–18). For example, isotope-coded affinity tags (ICAT) chemically target specific amino acids, typically cysteine, in the peptide sequence for differential labeling (11). Additional chemical labeling approaches have been developed to target other functional groups of the polypeptides (12, 14–18). Using these procedures, an isotopically labeled internal standard can easily be created for all components in the mixture. The use of these approaches compensates for ion suppression and variability in instrument to instrument performance when both isotopic variants co-elute. However, since many of the strategies introduce the isotopic label after significant sample processing, variable losses during this sample processing can introduce errors. Typical standard deviations for these labeling procedures have been reported to be 20–25% (5), which is significantly better than the 100% standard deviations often reported for the label-free approaches. These strategies also improve the throughput of the experiment as two or more samples are analyzed simultaneously; however, these processes alter the discovery workflow as additional steps are needed to label and/or clean-up the sample prior to analysis and increase the cost due to the isotopic labeling reagent.

Stable isotopes can also be introduced into biological systems through metabolic labeling. For instance, stable isotope labeling with amino acids in cell culture (SILAC) provides a simple and straightforward method for the incorporation of an isotopic label into proteins prior to MS-based proteomics (13). In a SILAC experiment, two cell populations are grown in culture media that are identical except that one of them contains a "light" form and the other a "heavy" form of particular amino acids (^{12}C-labeled and ^{13}C-labeled lysine and arginine, for example). The labeled analogs of amino acids are supplied to cells in culture instead of the natural amino acids, and it becomes incorporated into all newly synthesized proteins. After a number of cell divisions, each instance of the particular amino acids is replaced by its isotope-labeled analog. An advantage of this approach over the chemical tagging approaches is that the cells are mixed together immediately after cell lysis. Thus, proteins from both cell types are subjected to the exact same experimental conditions during sample handling, digestion, purification, etc., eliminating the differential losses that can occur when the samples are treated separately in a parallel manner. For this reason, SILAC is often considered the "gold standard" for quantitative proteomic analyses (19). The primary limitation of these in vivo labeling procedures is that they are typically limited to cells grown in culture.

2. Quantitative Approaches for Glycomics

Many of these quantitative proteomic tools have been adapted for glycomic analysis, and can be broadly divided once again into those that are label free and those where an isotopic label is introduced. The latter of these can be subdivided based on how the isotope is incorporated, i.e., metabolically or via chemical derivatization. These three strategies are expected to share similarity with their proteomic counterpart as far as the quantitative issues which each compensates for, and thus the level of reproducibility each can attain. These approaches will be described in the following sections, with the overall purpose of this work being to familiarize glycomics researchers with these quantitative techniques.

2.1. Label-Free Approaches

The dominant method for quantitative glycomics currently used is some variation of a label-free approach. This process involves releasing the glycans from a sample, often followed by a derivatization step, then acquisition of a MALDI–MS or ESI–MS spectrum, or performing an LC–MS experiment, as shown in **Fig. 3.1**. In many of these studies, the response from any one glycan is reported as the percentage of the response for all

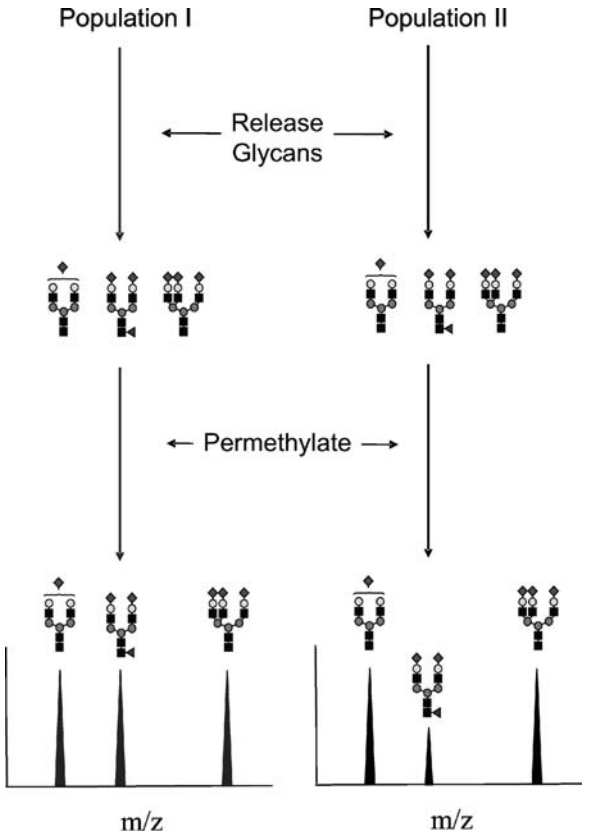


Fig. 3.1. Flow chart for quantitative glycan analysis using label-free approaches. Here, the glycans from two biological populations are released and permethylated in a parallel manner. The derivatized glycans are analyzed by MS and changes in glycan abundance are determined from the glycans' relative peak heights in the two spectra.

glycans in the sample. This follows the general procedure utilized for data processing in label-free proteomics, and thus compensates for instrument variability and matrix effects to some extent. The Human Proteome Organization (HUPO) recently published the results from a multi-institutional study on the label-free profiling of glycoprotein glycans, which highlights the strengths and weaknesses of this methodology (20). In general, this study revealed that MS quantitation was consistent with and comparable to chromatographic analysis of reductively aminated glycans, which is the generally accepted standard method for glycan quantitation (20). A second study reached the same conclusion when ESI-MS was compared with the more traditional LC-based methods with florescent detection (21). The HUPO study also noted several issues with quantitation, one of which is prompt, in-source, fragmentation of sialylated glycan ions produced by the MALDI process. ESI was found to be gentle enough so that the extent of sialylation was not perturbed. The presence of sialic acids

on a glycan had previously been reported to decrease the relative ionization efficiencies of sialylated glycans (22). Both of these factors contributed to significant errors in the quantification of sialylated glycans (20). Permethylation of glycans, which is commonly performed prior to MS analysis of glycans, was found to be critical to perform glycan quantitation by MALDI–MS (20), as this process stabilized the sialic acid residues and thus decreased the extent of in-source fragmentation. In addition, this derivatization process leads to more uniform ionization as it converts highly polar -OH and $-COO^-$ groups into non-polar, chemically homogeneous derivatives, which overcomes the issue associated with decreased ionization efficiency of sialylated glycans. The multitude of steps in the permethylation reaction does, however, introduce the possibility of differential losses, which in part has led to the development of a solid-phase procedure that is expected to minimize this potential problem (23, 24). These observations led the HUPO study to conclude that permethylation was needed for MALDI–MS glycan quantitation.

The simplicity of label-free glycan quantitation makes this an attractive procedure. The HUPO study concluded that MALDI–MS of permethylated glycans was as reliable as chromatographic methods of quantification. The interlaboratory reproducibility of this study on technical replicates (i.e., sample was prepared, then divided into three for three replicate MALDI–MS experiments) was very good, with coefficients of variation ranging between 1.3 and 8.8% for abundant glycans and 12–34% for a glycan of lower abundance. The good reproducibility of these experiments was expected as the response for each glycan was normalized to that from all glycans, all of the glycans were observed in the same MALDI–MS spectrum (eliminating instrument variation), and the technical replicate method negated changes in intensity from matrix effects. The cross-laboratory portion of this study also indicated very good reproducibility for all the major glycans, with coefficients of variation ranging between 6 and 12%; however, the variation was very large (>100%) for low-abundance glycans. This observation parallels those from label-free proteomic studies which have observed very good reproducibility for intense species, but significantly worse deviations for lower proteins of lower abundance (7). A limitation of MALDI–MS is that it cannot be used to quantify structurally distinct isomeric glycans when these are present in isomeric mixtures as this approach provides a measure of the total abundance of the collection of isomers at a particular mass, rather than the abundances of individual species in that collection. This is an advantage of the label-free LC–MS approach for quantitating glycans (25). In general, the label-free approaches offer a simple quantitative approach that does not alter the sample workflow. These strategies do not need stable isotopes, which elevate the cost of the other internal standard

approaches that will be discussed. Consequently, these approaches thus offer a straightforward manner to screen samples for glycan changes that can serve as the subjects of more detailed studies using an isotopic labeling strategy.

2.2. Isotopic Labeling Strategies

Several approaches have been developed that introduce a stable isotope to a glycan via chemical derivatization. Glycans are typically derivatized prior to analysis either by tagging the reducing terminus with a chromophore when subsequent analyses are chromatographic or by permethylation when the sample is to be analyzed by MS. Methods that introduce an isotopic label have also followed these two strategies and in essence have utilized the commercial availability of isotopically labeled reagents for these procedures. A typical workflow for these in vitro labeling approaches involves the parallel release of glycans from the sample populations under investigation, derivatization with an isotopic label after which the samples are mixed, followed by MS analysis, as shown in **Fig. 3.2**, where permethylation with $^{12}C/^{13}C$ methyl iodide was used as an example. Each of these procedures will be summarized in the following sections.

2.2.1. Labeling on the Reducing Terminus

Several of the reagents typically used to label the reducing terminus of glycans are available in their deuterated form, and thus provide a straightforward method to incorporate an isotopic label through reductive amination using standard protocols (26–28). Tags have also been synthesized that can modify oligosaccharides with four isotope-enriched variants allowing up to four samples to be analyzed simultaneously (29). One application of these tetraplex tags is to designate a reference glycan mixture, which is labeled with one isotopic variant of the tag, and used as an internal standard for all analyses. This approach enables the comparison of a large number of samples. In all of these cases, tags were added to the reducing terminus of the glycan, which is problematic for O-linked glycans since the reducing termini of these glycans are usually reduced as a result of β-elimination. The use of sodium borodeuteride in the place of sodium during β-elimination has been reported as a method for incorporation of an isotope for the quantification of O-linked glycans (30). The 1-Da shift introduced by this procedure is insufficient to shift the deuterated species away from the naturally occurring ^{13}C isotope peak of the species labeled with sodium borohydride, which in turn can lead to challenging quantitation. Overall, the commercial availability of these reagents makes this approach relatively easy to implement in labs that typically label the reducing terminus.

Each of these isotopic tags uses hydrogen/deuterium as the light/heavy isotopic species. This pair introduces the largest isotopic effect as the atomic mass of deuterium is twice that of hydrogen, and thus this issue should be addressed. For the reductive

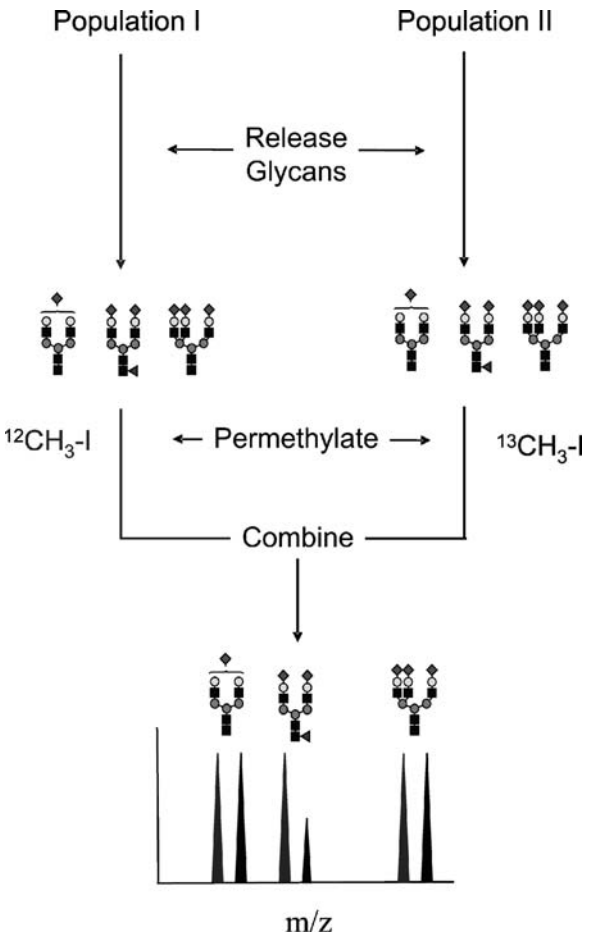


Fig. 3.2. Flow chart for quantitative glycan analysis using isotopic labeling during the permethylation step. Glycans from two biological samples are released followed by permethylation with $^{13}CH_3I$ or $^{12}CH_3I$ and mixed together prior to MS analysis. Changes in glycan abundance are determined by comparing the peak heights of the light and heavy labeled glycans.

amination tags, the kinetic isotope effect is expected to be fairly small as the substituted isotope is not directly involved in the bond that is breaking or forming and thus this is classified as a secondary effect. This is not the case when sodium borodeuteride is used during β-elimination as the deuterium is directly involved in the reaction, and thus, the relative change in rate associated with the incorporation of these isotopes is a function of the inverse square root of the ratios of the reduced masses of the atoms involved in the bonds based on the calculation of the vibrational energy required to break or form a bond. All of these labeling processes are intended to proceed to completion, and thus a kinetic isotope effect may decrease the rate of reaction, but if care is taken to drive the reaction to completion this kinetic effect is

not expected to be problematic. One area where the isotope effect introduced by hydrogen/deuterium substitution is often noticed involves LC separations. This phenomenon is widely reported in the proteomics literature, where the peptide labeled with the deuterated tag elutes at a later time than the same peptide with the non-deuterated tag and have been reported to cause errors in excess of 25% (5). Glycans labeled with deuterated forms of standard reductive amination reagents have also been reported to be partially resolved from their hydrogen-containing counterparts (26). As discussed above, the rational for introducing an isotopically labeled version of the analyte is to ensure that the analyte and standard are analyzed at the exact same time to compensate for differential instrument response and matrix effects; this criterion is not met when the analyte does not co-elute with standard, and thus care should be taken when using hydrogen/deuterium isotopic tags.

The introduction of an isotopic label with a reductive amination tag is an attractive procedure because this procedure does not alter existing workflows, i.e., the researcher simply derivatizes with the deuterated form of the tag. This labeling strategy also introduces a fixed shift in mass between the light and heavy pair, thus simplifying the identification of matched glycans. When this type of approach has been utilized for relative quantitation, coefficients of variation have been reported as 20% (29), which makes this approach significantly more precise than the label-free procedures discussed above. One limitation for these procedures is that they cannot be used to quantify structurally distinct isomeric glycans when these are present in isomeric mixtures as this approach provides a measure of the total abundance of the collection of isomers at a particular mass, rather than the abundances of individual species in that collection. One possible solution in these instances is to chromatographically separate these isomers; however, this may be difficult due to the chromatographic shifts associated with the use of deuterium as the heavy isotope.

2.2.2. Isotopic Incorporation During Permethylation

Because of the advantages associated with permethylation (discussed above) many glycans are subjected to this procedure prior to MS analysis. This process provides the opportunity to isotopically label oligosaccharides with the use of heavy methyl iodide ($^{13}CH_3$ or $^{12}CD_3$) and light methyl iodide ($^{12}CH_3$) in the standard permethylation procedures (21, 31, 32), as shown in **Fig. 3.3** where a milk oligosaccharide was. An example of this approach is shown in **Fig. 3.3**. The use of methyl iodide with varying degrees of deuterium content (i.e., $^{12}CH_3$, $^{12}CH_2D$, $^{12}CHD_2$, and $^{12}CD_3$) allows for the simultaneous analysis of four different samples (32). The kinetic isotope effect resulting from the use of CH_2DI has been estimated to be less than 2%, and thus this is not expected to significantly effect the permethylation

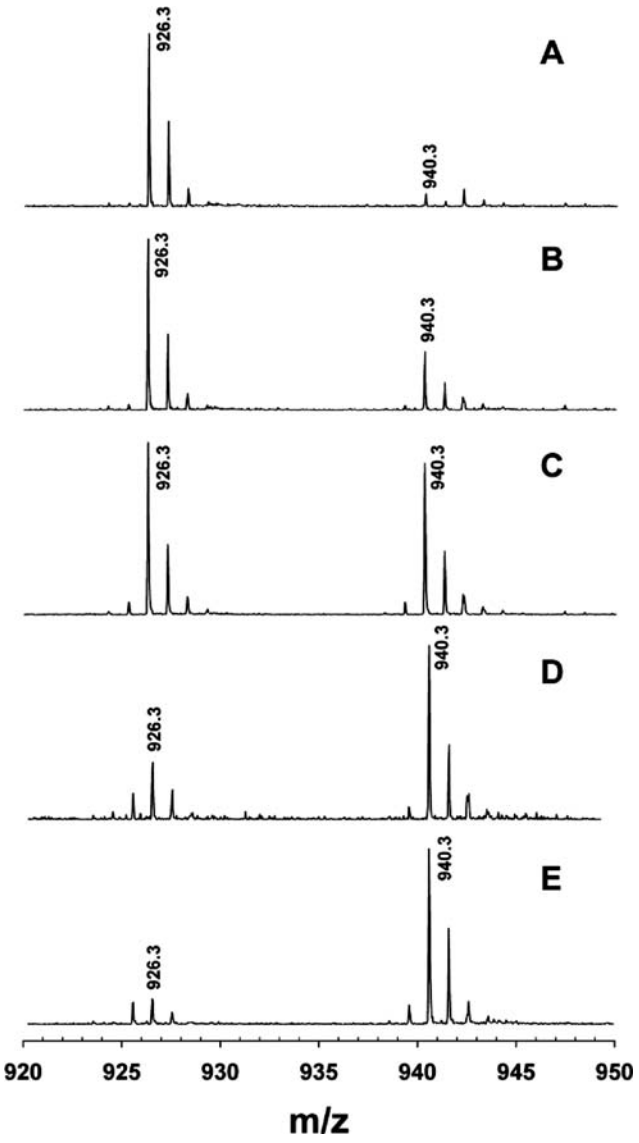


Fig. 3.3. MALDI-TOF MS spectra showing the isotopic pairs for the tetrasaccharide mixture (Gal2-GlcNAc-Glc) labeled with different proportions of ^{12}C and ^{13}C methyl iodide. For this experiment, two aliquots with similar amounts of oligosaccharides from human milk were separately labeled with either ^{12}C or ^{13}C methyl iodide. Before MS, these aliquots were mixed in the following ^{12}C:^{13}C proportions: 9:1 (A), 7.5:2.5 (B), 1:1 (C), 2.5:7.5 (D), and 1:9 (E). The monoisotopic *m/z* for the ^{12}C- and ^{13}C-labeled glycans are 926.3 and 940.3, respectively.

process (33). However, the use of deuterated methyl iodide is expected to lead to a chromatographic shift due to isotope effects, which because of the large number of methyl groups introduced is expected to be more pronounced than that observed with the reductive amination tags where only a few deuteriums are added

to the glycan (26). This chromatographic shift is not expected when $^{13}CH_3$ and $^{12}CH_3$ are used, and was one of the reasons this pair was selected (31).

A potential issue unique to the permethylation labeling strategy is that large errors can be introduced by small variation in labeling efficiency since the number of modified sites is so large. For instance, changing the labeling efficiency from 99.1 to 99.0% leads to a 0.1% error when the isotope is introduced via reductive amination since this label is only added to a single site on the glycan. A similar 0.1% decrease during permethylation has a much more profound effect on the error level since glycans have many sites that become methylated. A fully sialylated triantennary complex glycan, for example, has 50 sites of permethylation, and thus the 0.1% change in labeling efficiency results in a 5.0% error. Consequently, with the permethylation approach reproducible quantitation is much more dependent on reproducible labeling than with the other methods of introducing an isotopic label.

An important aspect of incorporating the isotopic label during permethylation is that it does not alter the typical glycan workflow and can be performed on any glycan that is amenable to permethylation. One limitation of these isotopic labeling approaches, however, is that the mass difference (Δm) between the heavy and light forms of each glycan is variable and can be very large, as Δm is proportional to the number of methylation sites on the glycan (31). This variability can confound the analysis of complex mixtures, as it can be difficult to match the differentially labeled forms of the same chemical species. In addition, this approach cannot be used to quantify the structurally distinct isomeric glycans that are often encountered in glycomic analyses, as was the case with the labeled reductive amination tags. Despite these limitations, permethylation tagging procedures have been reported to provide linear quantification over two orders of magnitude and yielded coefficients of variation in the range of 15–30% (31, 32).

An isobaric labeling approach based upon the incorporation of an isobaric label during permethylation, called quantitation by isobaric labeling (QUIBL), has recently been introduced (33, 34). In this approach, glycans are permethylated with either $^{13}CH_3I$ or $^{12}CH_2DI$. This pair of reagents has the same nominal mass, but differ in their exact mass by 0.002922 Da per label. This mass difference is difficult to resolve with current mass spectrometers in cases where only a single label is attached to the analyte. However, glycans typically contain multiple hydroxyl groups, which increase the delta mass between the two samples. Since the number of hydroxyl groups increases with the mass of the glycan, the difference between these isobaric species also increases and thus the resolution needed is approximately 25,000 $\Delta M/M$ and is independent of the glycan's size for typical N-linked and

O-linked species. This level of mass resolution is easily attained with FT-MS and Orbitrap MS systems.

The advantages of QUIBL are numerous and result primarily from the isobaric ions appearing at the same nominal mass to charge ratio. This characteristic leads to increased ion intensity, as ions from both samples are not distributed between isotopic species having different *m/z* values. The small mass difference between these isobars allows the two species to be simultaneously selected for MS^n analysis, permitting the relative quantitation of isomeric glycans, as shown in **Fig. 3.4**. This later characteristic was used to demonstrate a decrease in the level of N-linked glycans containing the Lewis X structure glycans when mouse embryonic stem cells were allowed to spontaneously differentiate. In addition, the light and heavy analyte ions resulting from reagents not having 100% isotope incorporation can still be resolved, unlike the typical isotope labeling strategies where the ion produced by under incorporation of the heavy isotope results in an ion that is indistinguishable from an ion produced by the light species. The characteristic of QUIBL results from the presence of multiple labeling sites on the glycan, and thus replacing one of the many ^{13}C atoms with a ^{12}C atom or replacing one of the many D (or ^{2}H) atoms with an ^{1}H atom decreases the analyte's mass by approximately 1 Da, however, the resulting ion is detected in the appropriate ($^{13}CH_3$-labeled or $^{12}CH_2D$-labeled) ion series because it still contains a large number of

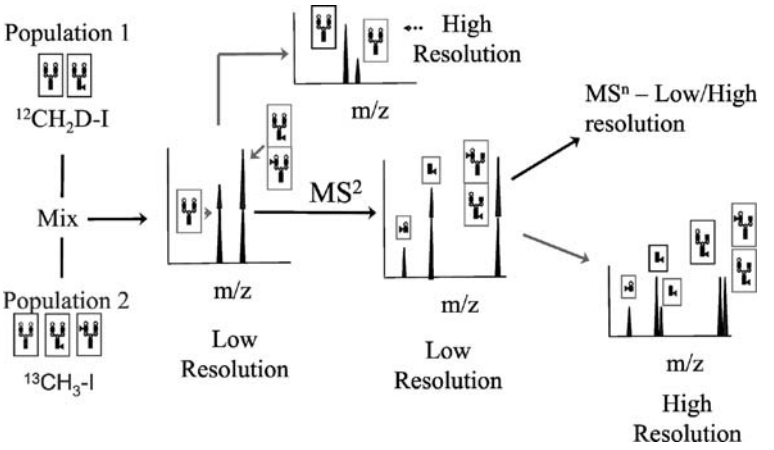


Fig. 3.4. Flow chart for quantitative glycan analysis using isobaric labeling. Glycans from two biological samples are permethylated in either $^{13}CH_3I$ or $^{12}CH_2DI$ and mixed together prior to analysis. At low mass resolution, the two labeled species appear at the same *m/z* value thereby increasing their abundance and decreasing sample complexity. Analysis of the glycans by high-resolution MS separates the differentially labeled glycan precursor ions permitting their relative quantitation by comparing the peak intensities from the $^{13}CH_3$ to the $^{12}CH_2D$ labeled glycans. Structural information on the glycan is provided by low-resolution MS^n, which does not alter the ratio of isobaric labels. High-resolution analysis of the MS^n fragment ions permits the isomeric glycans to be quantified.

isotopic labels. This greatly simplifies quantitation, which is accomplished by summing the ion abundances for the $^{13}CH_3$-labeled and $^{12}CH_2D$-labeled series and comparing these two values.

QUIBL offers a relatively straightforward approach to compensate for quantitative errors resulting from instrument performance and matrix effects, provided that the researcher has access to a high-resolution MS system. This procedure does not alter the workflow for labs that routinely permethylate glycans prior to analysis. As with the other glycan-tagging procedures discussed here, QUIBL does not compensate for differential losses resulting from multiple samples processed in parallel. A chromatographic shift is also expected between the light and heavy labeled species as a result of the CH_2D label. Despite these limitations, QUIBL has been reported to provide linear quantification over at least two orders of magnitude and yielded coefficients of variation in the range of 10–26% (33, 34), as demonstrated in **Fig. 3.5**. This level of accuracy and dynamic range is approximately the same as that seen using the other glycan labeling strategies. The one advantage of QUIBL over these other methods is that it is capable of quantitating individual glycan present in isomeric mixtures, **Fig. 3.4**.

2.2.3. In Vivo Labeling

An in vivo labeling strategy has recently been described for glycomic studies (35). A flow chart for quantitative glycan analysis using in vivo isotopic labeling is illustrated in **Fig. 3.6**. This methodology termed IDAWG, *I*sotopic *D*etection of *A*mino sugars *W*ith *G*lutamine, relies on the side-chain of glutamine being the sole donor source of nitrogen for amino sugars in the production of sugar nucleotides according to the hexosamine biosynthetic pathway, as shown in **Fig. 3.7**. Thus, introduction of glutamine with an ^{15}N-labeled side-chain (amide-^{15}N-Gln) into Gln-free media allows for the incorporation of one ^{15}N into all amino sugars, including GlcNAc, GalNAc, and sialic acids. Consequently, the mass of all N-linked and O-linked glycans, glycolipids, and extracellular matrix polysaccharides, is increased by +1 Da per amino sugar. This approach was demonstrated by the analysis of both N-linked and O-linked glycans released from proteins of murine embryonic stem cells grown in both the light and amide-^{15}N-Gln. By incorporating the isotopic label into glycans as they are being synthesized, IDAWG is similar to the SILAC procedure used in proteomics. Both of these approaches share the advantage that the differentially labeled cells can be mixed together at the beginning of the analytic procedure, minimizing the contribution of handling and work-up to overall variability. IDAWG is expected to compare with SILAC and thus have a linear response for at least two orders of magnitude and yielded coefficients of variation in the range of 10–20%.

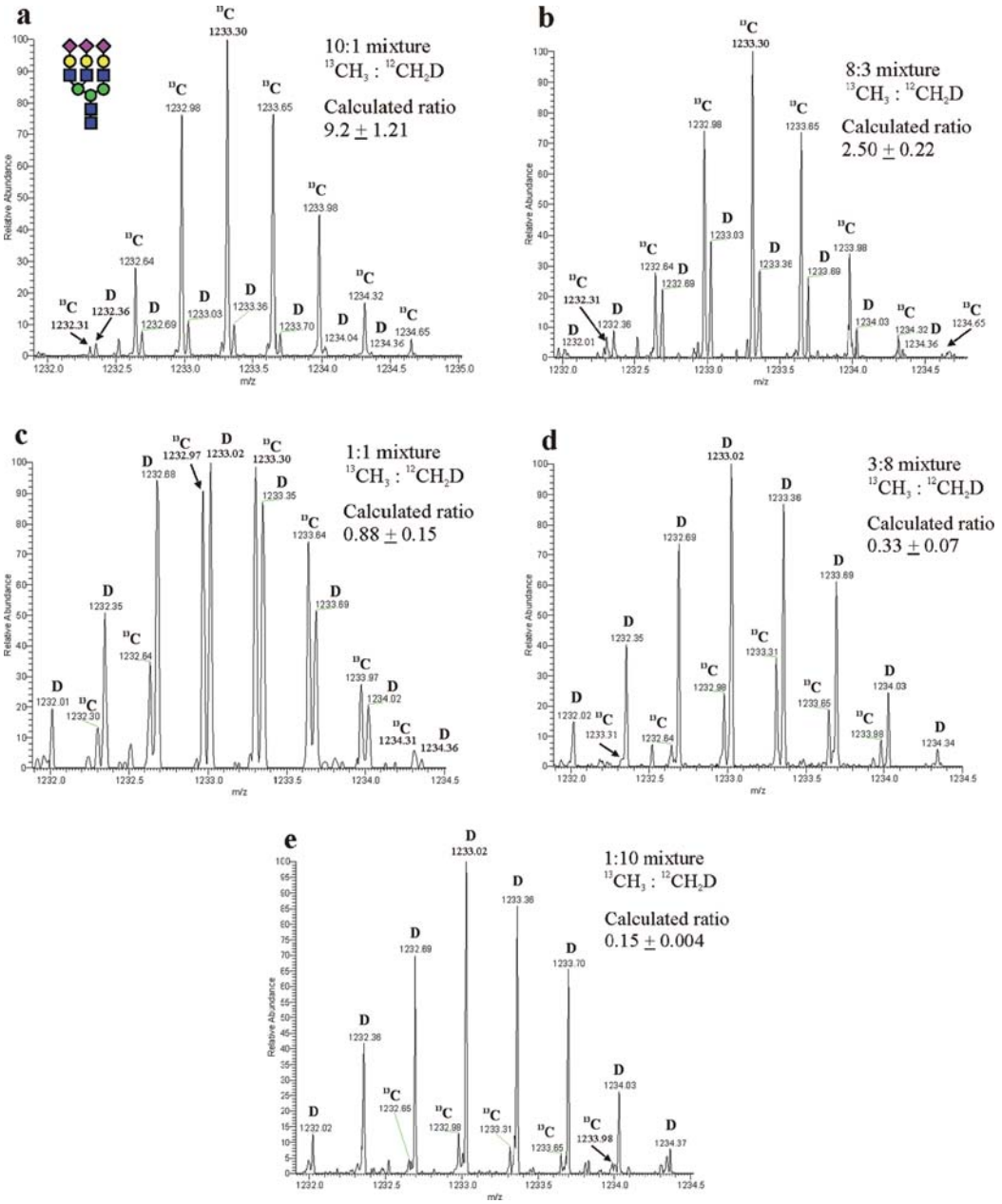


Fig. 3.5. QUIBL analysis of a differently labeled fetuin glycan mixed at five different ratios. Two fetuin glycan mixtures were permethylated in either $^{13}CH_3I$ or $^{12}CH_2DI$. The two differentially labeled glycan mixtures were then mixed together at the ratios 10:1, 8:3, 1:1, 3:8, and 1:10 ($^{13}CH_3$:$^{12}CH_2D$) and analyzed by FT MS (a, b, c, d, e). Accurate quantitation was achieved at all ratios over two orders of magnitude.

IDAWG is a relatively straightforward procedure. Cell culture media are typically supplied glutamine free since this amino acid rapidly decomposes in aqueous solutions, which leads to glutamine being one of the most common supplements to cell

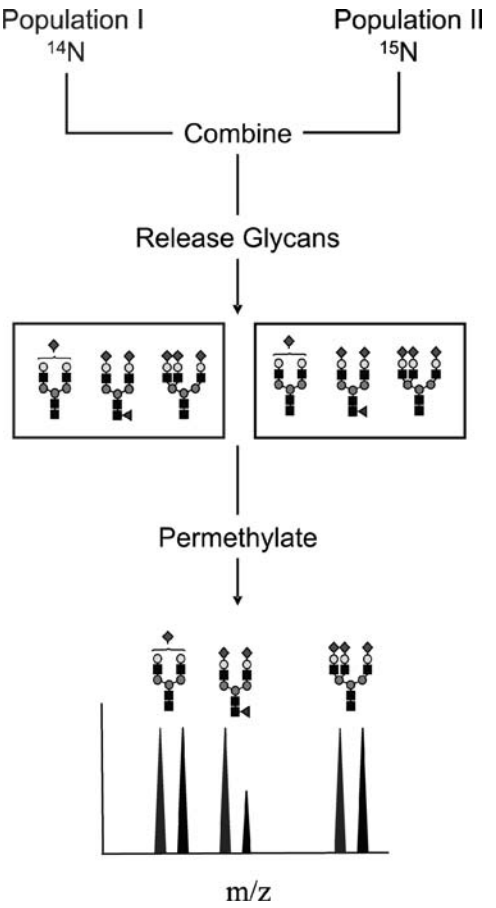


Fig. 3.6. Flow chart for quantitative glycan analysis using in vivo isotopic labeling. Here, glycans in one of the populations are labeled with ^{14}N while glycans in the other population are labeled with ^{15}N. The two populations are mixed, followed by glycan release, permethylation and MS analysis. Changes in glycan abundance are determined by comparing the peak heights of the light and heavy labeled glycans.

culture media. This phenomenon greatly simplifies the IDAWG approach, since the only change in standard operating procedures it to use amide-^{15}N-Gln when preparing fresh media. The concentration of amide-^{15}N-Gln used for IDAWG is the same as that used for normal cell growth with ^{14}N-Gln. Using these conditions, 96% incorporation of ^{15}N into N-linked and O-linked glycans has been reported after labeling mouse embryonic stem cells for 3 days (35). Consequently, this in vivo labeling strategy provides a strategy to isotopically label a glycan population without significantly altering the experimental procedures.

Metabolic labeling of glycans provides new opportunities for assessing the dynamics of glycan turnover during the course of any cellular behavior that can be induced or sustained in culture. By completely labeling cells with heavy Gln and then replacing

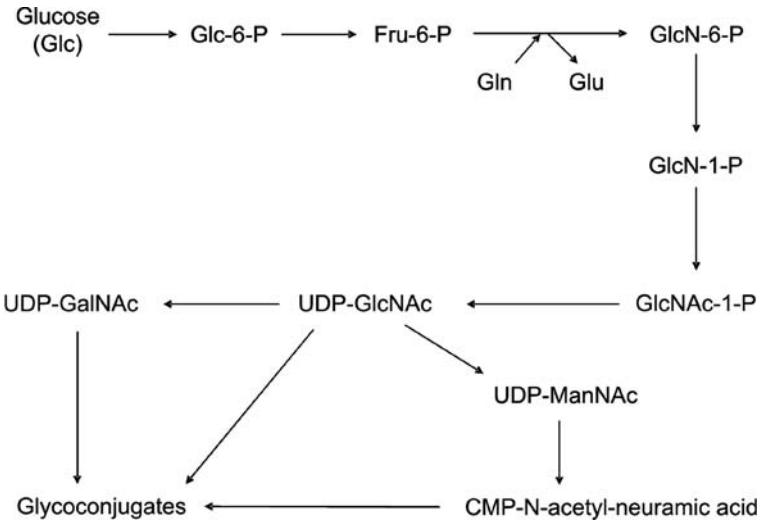


Fig. 3.7. The hexosamine biosynthetic pathway demonstrating that the side-chain of glutamine is the sole donor source of nitrogen for amino sugars in the production of sugar nucleotides, which allows the introduction of an ^{15}N isotopic tag into all amino sugars, including GlcNAc, GalNAc, and sialic acids. Species containing ^{15}N are indicated by blue type.

the media supplement with light Gln, the half-life of any amino sugar-containing glycans can be determined. Previously, glycan turnover studies required incorporation of radioactive monosaccharide and extensive subsequent fractionation to identify specific changes in glycan expression. Generally, these radiotracer techniques allowed for very sensitive detection of glycan classes, but lacked the resolution to follow individual glycan structures or subsets of biosynthetically related species. The stable isotope incorporation method reported here merges the analytic advantages of high-resolution mass spectrometry to the biological necessity of understanding the dynamics of glycan turnover. Thus, IDAWG appears to be a powerful quantitative tool for exploring the biological role of glycans, glycoproteins, and glycolipids in cell culture systems.

3. Summary

A range of procedures have been developed for quantitative glycomics, each of which has its own strengths and weaknesses. In general, the label-free approaches are the easiest to implement; however, these also show the lowest accuracy and reproducibility, particularly in the analysis of glycan with low abundance. The label-free approaches, thus, provide a method for quickly sur-

veying samples for glycan changes that can be targeted by the more accurate, but also more expensive, isotopic labeling strategies. The in vivo labeling approach is expected to offer the highest level of accuracy, but this appears to be limited to the study of cell culture systems. The presence of isomeric glycans causes problems with quantitation, and only the QUIBL approach has demonstrated an ability to quantitate individual glycans present in isomeric mixtures. Despite these limitations, all of these quantitative approaches offer powerful new tools in the glycomics toolbox.

References

1. Liu, H., Sadygov, R. G., and Yates 3rd, J. R. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201.
2. Radulovic, D., Jelveh, S., Ryu, S., Hamilton, T. G., Foss, E., Mao, Y., and Emili, A. (2004) Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. *Mol. Cell Proteomics* **3**, 984–997.
3. Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G. Z., McKenna, T., Nold, M. J., Richardson, K., Young, P., and Geromanos, S. (2005) Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* **77**, 2187–2200.
4. Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., Anderle, M., and Becker, C. H. (2003) Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* **75**, 4818–4826.
5. Goshe, M. B. and R.D. Smith (2003) Stable isotope-coded proteomic mass spectrometry. *Curr. Opin. Biotechnol.* **14**, 101–109.
6. Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K., Florens, L., and Washburn, M. P. (2006) Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **5**, 2339–2347.
7. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A., Ahn, N. G. (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **4**, 1487–1502.
8. Hendrickson, E. L., Xia, Q., Wang, T., Leigh, J. A., and Hackett, M. (2006) Comparison of spectral counting and metabolic stable isotope labeling for use with quantitative microbial proteomics. *The Analyst* **131**, 1335–1341.
9. Goshe, M. B. and Smith, R. D. (2003) Stable isotope-coded proteomic mass spectrometry. *Curr. Opin. Biotechnol.* **14**, 101–109.
10. Tao, W. A. and Aebersold, R. (2003) Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. *Curr. Opin. Biotechnol.* **14**, 110–118.
11. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**, 994–999.
12. Liu, P. and Regnier, F. E. (2002) An isotope coding strategy for proteomics involving both amine and carboxyl group labeling. *J. Proteome Res.* **1**, 443–450.
13. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* **1**, 376–386.
14. Rao, K. C., Carruth, R. T., and Miyagi, M. (2005) Proteolytic ^{18}O labeling by peptidyl-Lys metalloendopeptidase for comparative proteomics. *J. Proteome Res.* **4**, 507–514.
15. Schnolzer, M., Jedrzejewski, P., and Lehmann, W. D. (1996) Protease-catalyzed incorporation of ^{18}O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* **17**, 945–953.
16. Vosseller, K., Hansen, K. C., Chalkley, R. J., Trinidad, J. C., Wells, L., Hart, G. W., and Burlingame, A. L. (2005) Quantitative analysis of both protein expression and serine/threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics* **5**, 388–398.

17. Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J., and Hart, G. W. (2002) Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell Proteomics* **1**, 791–804.
18. Yao, X., Freas, A., Ramirez, J., Demirev, P. A., and Fenselau, C. (2001) Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* **73**, 2836–2842.
19. Ong, S. E., Foster, L. J., and Mann, M. (2003) Mass spectrometric-based approaches in quantitative proteomics. *Methods* **29**, 124–130.
20. Wada, Y., Azadi, P., Costello, C. E., Dell, A., Dwek, R. A., et al. (2007) Comparison of the methods for profiling glycoprotein glycans – HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **17**, 411–422.
21. Aoki, K., Perlman, M., Lim, J. M., Cantu, R., Wells, L., and Tiemeyer, M. (2007) Dynamic developmental elaboration of N-linked glycan complexity in the *Drosophila melanogaster* embryo. *J. Biol. Chem.* **282**, 9127–9142.
22. Sutton, C. W., O'Neill, J. A., and Cottrell, J. S. (1994) Site-specific characterization of glycoprotein carbohydrates by exoglycosidase digestion and laser-desorption mass spectrometry. *Anal. Biochem.* **218**, 34–46.
23. Kang, P., Mechref, Y., Klouckova, I., and Novotny, M. V. (2005) Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun. Mass Spectrom.* **19**, 3421–3428.
24. Kang, P., Mechre, Y. and Novotny, M. V. (2008) High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Commun. Mass Spectrom.* **22**, 721–734.
25. Karlsson, N. G., Wilson, N. L., Wirth, H. J., Dawes, P., Joshi, H., and Packer, N. H. (2004) Negative ion graphitised carbon nano-liquid chromatography–mass spectrometry increases sensitivity for glycoprotein oligosaccharide analysis. *Rapid Commun. Mass Spectrom.* **18**, 2282–2292.
26. Yuan, J., Hashii, N., Kawasaki, N., Itoh, S., Kawanishi, T., and Hayakawa, T. (2005) Isotope tag method for quantitative analysis of carbohydrates by liquid chromatography–mass spectrometry. *J. Chromatogr. A* **1067**, 145–152.
27. Hitchcock, A. M., Costello, C. E., and Zaia, J. (2006) Glycoform quantification of chondroitin/dermatan sulfate using an LC–MS/MS platform. *Biochemistry* **45**, 2350–2361.
28. Hsu, J., Chang, S. J., and Franz, A. H. (2006) MALDI-TOF and ESI-MS analysis of oligosaccharides labeled with a new multifunctional oligosaccharide tag. *J. Am. Soc. Mass Spectrom.* **17**, 194–204.
29. Bowman, M. J. and Zaia, J. (2007) Tags for the stable isotopic labeling of carbohydrates and quantitative analysis by mass spectrometry. *Anal. Chem.* **79**, 5777–5784.
30. Xie, Y., Liu, J., Zhang, J., Hedrick, J. L., and Lebrilla, C. B. (2004) Method for the comparative glycomic analyses of O-linked, mucin-type oligosaccharides. *Anal. Chem.* **76**, 5186–5197.
31. Alvarez-Manilla, G., Warren, N. L., Abney, T., Atwood 3rd, J., Azadi, P., York, W. S., Pierce, M., and Orlando R. (2007) Tools for glycomics: relative quantitation of glycans by isotopic permethylation using $^{13}CH_3I$. *Glycobiology* **17**, 677–687.
32. Kang, P., Mechref, Y., Kyselova, Z., Goetz, J. A., Novotny, M. V. (2007) Comparative glycomic mapping through quantitative permethylation and stable-isotope labeling. *Anal. Chem.* **79**, 6064–6073.
33. Botelho, J. C., Atwood 3rd, J. A., Cheng, L., Alvarez-Manilla, G., York, W. S., and Orlando, R. (2008) Quantification by isobaric labeling (QUIBL) for the comparative glycomic study of O-linked glycans. *Intl. J. Mass Spectrom.* **278**(2–3): 137–142.
34. Atwood 3rd, J. A., Cheng, L., Alvarez-Manilla, G., Warren, N. L., York, W. S., and Orlando, R. (2008) Quantitation by isobaric labeling: applications to glycomics. *J. Proteome Res.* **7**, 367–374.
35. Orlando, R., Lim, J., Atwood 3rd, J. A., Angel, P. M., Fang, M., Alvarez-Manilla, G., Moremen, K. W., York, W. S., Tiemeyer, M., Pierce, M., Dalton, S., Wells, L. (2009) IDAWG: metabolic incorporation of stable isotope labels for quantitative glycomics of cultured cells. *J. Proteome Res.* **8**, 3816–3823.

Chapter 4

Human Gangliosides and Bacterial Lipo-oligosaccharides in the Development of Autoimmune Neuropathies

Nobuhiro Yuki

Abstract

Guillain–Barré syndrome (GBS), the most frequent cause of acute flaccid paralysis, can develop after infection by *Campylobacter jejuni*. The condition is often associated with serum anti-GM1 or anti-GD1a IgG antibodies. Gangliosides contribute to stability of paranodal junctions and ion channel clusters in myelinated nerve fibers. Autoantibodies to GM1 and GD1a disrupt lipid rafts, paranodal or nodal structures, and ion channel clusters in peripheral motor nerves. Molecular mimicry exists between GM1 and GD1a gangliosides and lipo-oligosaccharides of *C. jejuni* isolates from GBS patients. Sensitization of rabbits with GM1 or *C. jejuni* lipo-oligosaccharide produces replica of GBS. These findings provide strong evidence for carbohydrate mimicry being a cause of GBS and show the role of gangliosides in peripheral nerves.

Key words: Autoantibody, autoimmune disease, *Campylobacter jejuni*, carbohydrate mimicry, ganglioside, Guillain–Barré syndrome, Fisher syndrome, lipo-oligosaccharide, molecular mimicry, sialyltransferase.

1. Introduction

Guillain–Barré syndrome (GBS), which is characterized by acute onset of limb weakness and absence of muscle stretch reflexes, is a prototype of autoimmune-mediated peripheral neuropathies. Peripheral nerves are composed of axons and myelin. Primary target is axons in some GBS patients and myelin in others. For this reason, GBS is classified into two major subtypes, acute motor axonal neuropathy (AMAN) and acute inflammatory demyelinating neuropathy (AIDP) (1, 2). Fisher syndrome (FS), which is characterized by ophthalmoplegia, ataxia, and areflexia, is

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_4,

considered to be a variant of GBS because these illnesses sometime overlap. Although the pathogenic mechanism remains unknown in most autoimmune diseases, our understanding of the pathogenesis of GBS and FS has greatly advanced in the last two decades. This also has let us recognize physiological and pathological significance of carbohydrates such as ganglioside and bacterial lipo-oligosaccharide (LOS). Here I review our research that has elucidated ganglioside function in the peripheral nerves and pathological role of LOS in the development of GBS and FS.

2. Carbohydrate Mimicry as a Cause of Autoimmune Disease

Molecular mimicry between human tissue and microorganism has been proposed to be a pathogenic mechanism of autoimmune diseases. Four criteria should be satisfied to conclude that a disease is triggered by molecular mimicry (3): (i) establishment of an epidemiological association between the infectious agent and immune-mediated disease, (ii) identification of T cells or antibodies directed against the patient's target antigens, (iii) identification of microbial mimics of the target antigen, and (iv) reproduction of the disease in an animal model. Among a number of autoimmune diseases, GBS has first fulfilled all the four criteria and verified that molecular mimicry is a true cause of an autoimmune disease.

2.1. Campylobacter jejuni *Infection*

GBS is a typical post-infectious autoimmune disorder, which develops in one to four weeks subsequent to microbial infection. The Gram-negative bacterium *Campylobacter jejuni*, a leading cause of gastroenteritis in developed countries, is the most common antecedent agent (4). A prospective case–control study established an epidemiological association between *C. jejuni* infection and GBS (5). *C. jejuni* infection is associated with AMAN, but not with AIDP (6).

2.2. Autoantibodies Against Gangliosides

Gangliosides constitute a large family of glycosphingolipids predominantly distributed on the cell-surface membrane and anchored in the external leaflet of the lipid bilayer by a ceramide moiety. Sialylated oligosaccharides are exposed extracellularly. In 1990, we reported two patients with AMAN subsequent to *C. jejuni* enteritis who had serum anti-GM1 IgG antibodies, whereas these were absent in 10 patients with uncomplicated *C. jejuni* enteritis (7). This report set a new stage for research on the pathogenesis of GBS. Anti-GD1a IgG antibodies are also associated with AMAN subsequent to *C. jejuni* infection (8).

2.3. Ganglioside Mimicry in C. jejuni LOS

Gangliosides extracted from bovine brain have been used to treat various neurological disorders. Since the report of a case of amyotrophic lateral sclerosis-like disorder that started shortly after ganglioside therapy (9), there have been other reports of patients who developed GBS after ganglioside administration. After parenteral ganglioside treatment, several patients developed AMAN in combination with serum anti-GM1 or anti-GD1a IgG antibodies (10). These observations suggested that *C. jejuni* carries ganglioside-like structure. LOSs constitute a family of phosphorylated glycolipids and are anchored in the outer membrane of *C. jejuni*. To test the molecular mimicry hypothesis, LOS was extracted from *C. jejuni* isolated from an AMAN patient who had anti-GM1 IgG antibodies (11). That LOS reacted with the cholera toxin B-subunit, a specific ligand for GM1-oligosaccharide. Gas-liquid chromatography-mass spectrometric analysis and ^{1}H nuclear magnetic resonance showed that the terminal tetrasaccharides of the purified LOS were identical to that of GM1 ganglioside (**Fig. 4.1**) (12). This study proved the existence of molecular mimicry between human peripheral nerve and an antecedent infectious agent of GBS. The discovery promoted research on the pathogenesis of *C. jejuni*-related GBS. This bacterial strain also carried a GD1a-like LOS (13).

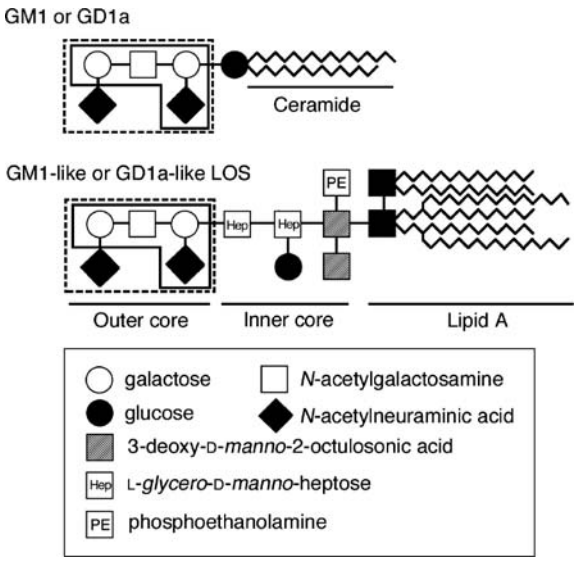


Fig. 4.1. Carbohydrate mimicry between gangliosides and *Campylobacter jejuni* lipo-oligosaccharides (LOSs). The terminal tetrasaccharide of GM1-like LOS is identical to that of GM1 (shown by *solid lines*). The terminal pentasaccharide structure of GD1a-like LOS is identical to that of GD1a (shown by *dashed lines*). (Reproduced from (12) with permission from the American Association of Immunologists.)

2.4. Animal Models

We sensitized Japanese white rabbits as well as New Zealand white rabbits with a bovine brain ganglioside mixture that included GM1 or with purified GM1 fraction (14). The rabbits developed high anti-GM1 IgG antibody titers, followed by flaccid limb weakness of acute onset with a monophasic course. The antibody titers did not differ before and after the disease onset, but high-affinity antibodies were detected only at the disease onset (15). This suggests that high affinity of anti-GM1 IgG antibodies is essential for the disease development. Pathological findings in the rabbit peripheral nerves were predominantly Wallerian-like degeneration with neither lymphocytic infiltration nor demyelination. Anterior spinal nerves showed macrophage infiltration in the periaxonal space, but surrounding myelin sheaths remained almost intact (16). These pathological findings correspond well with those in human AMAN. GM1 is expressed in both central and peripheral nervous systems, but the pathological changes are seen only in the peripheral nerves. Probably because blood–brain barrier is tighter than blood–nerve barrier, autoantibodies cannot go into the brain.

The most straightforward way to verify whether molecular mimicry between microbes and autoantigens causes GBS is to establish a disease model by immunization of animals with components of an antecedent infectious agent. We injected Japanese white rabbits repeatedly with GM1-like and GD1a-like LOSs of *C. jejuni* isolated from a patient with AMAN (11, 13, 17). Some rabbits developed limb weakness associated with anti-GM1 IgG antibodies. The pathological findings, compatible with the features of human AMAN, are evidence that rabbits inoculated with *C. jejuni* LOS constitute a valid AMAN model.

Research on molecular mimicry and autoimmunity has focused principally on T cell-mediated, anti-peptide responses, rather than on antibody responses to carbohydrate structures. As mentioned, however, carbohydrate mimicry between GM1 and *C. jejuni* LOS triggers production of pathogenic autoantibodies and development of AMAN. This new concept that carbohydrate mimicry can cause an autoimmune disease provides a clue to the resolution of the pathogenesis of other immune-mediated diseases.

2.5. CD1-Independent Antibody Response

Whereas IgG antibodies against bacterial polysaccharides belong to IgG2 subclass, the cross-reactive IgG antibodies to LOS and gangliosides belong to IgG1 and IgG3 (18). This subclass pattern suggests that T cells help in the production of anti-ganglioside IgG antibodies. Human group 1 CD1 molecules, CD1a, CD1b, and CD1c, bind microbial glycolipids as well as self-glycolipids including GM1 and present the antigens to T cells and NKT cells. Group 2 CD1 molecule, CD1d, also binds microbial and

self-glycolipids and presents the antigens to NKT cells. Because glycolipid-specific, CD1-restricted T cells are present, we assumed that a certain group 1 CD1 molecule functions in the production of IgG antibodies against *C. jejuni* LOS, that cross-react with self-gangliosides. We investigated in vitro in which CD1 molecule binds to *C. jejuni* LOS (12. Neither human CD1a, CD1b, nor CD1c bound GM1-like LOS, whereas both human and mouse CD1d bound GM1-like LOS. Sensitization of GM1/GD1a-deficient mice (GalNAcT$^{-/-}$ mice) lacking the functional gene for (*N*-acetylneuraminyl)-galactosylglucosylceramide *N*-acetylgalactosaminyltransferase (EC 2.4.1.92) with *C. jejuni* develops high titers of IgG antibodies against GM1 and GD1a, because they have no natural tolerance to these gangliosides. CD1d knockout mice were intercrossed to clarify whether CD1d is essential for the production of anti-ganglioside IgG antibodies, but sensitization of the mice with *C. jejuni* developed high antibody responses against GM1 and GD1a. These results indicate that CD1d does not function in the production of anti-ganglioside IgG antibodies at least in mice.

3. Ganglioside Function in Nerves

GM1 and GD1a are autoantigens for AMAN, but ganglioside function of myelinated nerves has yet to be clarified. Myelinated axons are divided into four functional regions: nodes of Ranvier, paranodes, juxtaparanodes, and internodes. In myelinated nerve fibers, paranodal axo-glial junctions are important for ion channel clustering and rapid action potential propagation. To elucidate ganglioside function at and near nodes of Ranvier, we examined those in GalNAcT$^{-/-}$ mice (19). In both peripheral and central nervous systems, some paranodal loops failed to attach to the axolemma. Voltage-gated Na^+ (Nav) channels are highly concentrated at the nodes of Ranvier and voltage-gated K^+ channels are localized at the juxtaparanodes. In the GM1/GD1a-deficent mice, however, K^+ channels are aberrantly present at the paranodes (**Fig.4.2**). Abnormal protrusion of paranodal and nodal axolemmas is seen. These findings increase with age. The defects were more prevalent in ventral than dorsal roots and less frequent in mutant mice lacking the b-series gangliosides but with excess GM1 and GD1a. Electrophysiological studies revealed nerve conduction slowing and reduced nodal Na^+ current in mutant peripheral motor nerves. These results indicate that gangliosides are components that contribute to stability and maintenance of neuron–glia interactions at paranodes. The amounts of essential components of paranodal junctions in

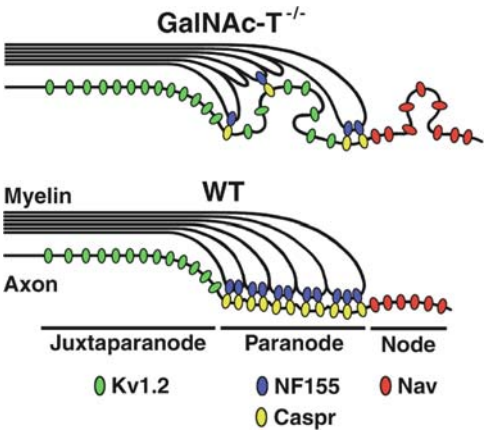


Fig. 4.2. Schematic representation of the paranodal and nodal disruption in peripheral nerves of GalNAc-T$^{-/-}$ mice which lack complex gangliosides such as GM1 and GD1a. In wild-type mice, voltage-gated Na^+ (Nav) and K^+ (Kav) clusters are located at nodes and juxtaparanodes, respectively. The paranodal junction tightly attaches the myelin sheath to the axon. Neurofascin 155 (NF155) and contactin-associated protein (Caspr) are located at myelin membrane and axolemma at paranodes, respectively. In GalNAc-T$^{-/-}$ mice, axo-glial junctions are not formed in some paranodal myelin loops. Kv channels aberrantly localized at the paranodes. Abnormal protrusions of paranodal and nodal axolemma are frequently seen. Nodal Nav channel clusters are lengthened.

low-density, detergent-insoluble membrane fractions were reduced in mutant brains. The results indicate that gangliosides are lipid raft components that contribute to stability maintenance of neuron–glia junctions at paranodes.

3.1. Nav Channel Dysfunction by Anti-ganglioside Antibodies

Whether anti-GM1 antibodies can block Nav channels at the nodes, which are responsible for nerve conduction, has been controversial (20, 21). In the spinal anterior roots in AMAN rabbits, IgG antibodies bind to nodes of Ranvier, where GM1 is highly expressed, and activate complements (**Fig. 4.3**) (22). C3 fragments are deposited, and membrane attack complex, MAC, is formed at the nodal axolemma. Nav channel clusters disappear at lengthened nodes with complement deposition. This pathological change is able to produce muscle weakness. In the advanced stage of paranodal disruption, potassium channel clusters at the juxtaparanodes are disrupted and disappear. Complement deposition is prominent in the acute progressive phase, but decreases with the clinical course. Modulation of Nav channel property by autoantibodies has been proposed as a novel mechanism in some neuroimmunological diseases (23). The results suggest that Nav channel alterations occur as a consequence of complement-mediated disruption of interactions between axons and Schwann cells and provide new insights into the pathogenesis of autoimmune neuropathies.

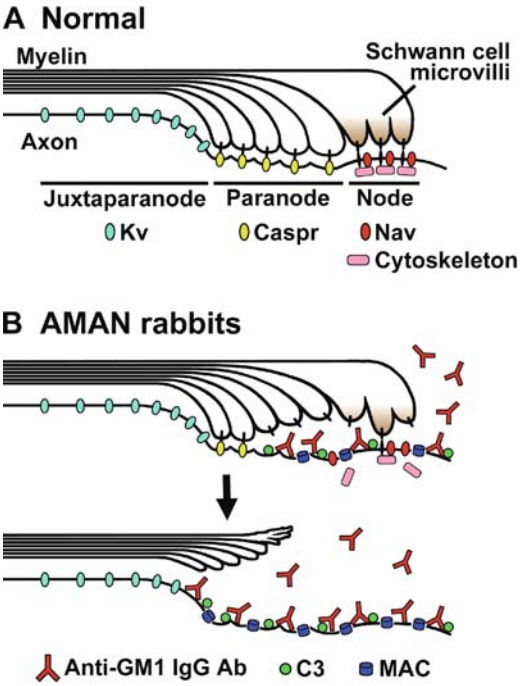


Fig. 4.3. Autoimmune-mediated disruption of nodes in acute motor axonal neuropathy model. Anti-GM1 IgG antibodies cause complement-mediated attack with membrane attack complex (MAC) formation at the nodal and paranodal axolemmas. Nav channel clusters are altered by the destruction of structures that mediate their stabilization, including the axonal cytoskeleton at nodes, Schwann cell microvilli, and paranodal junctions. As autoimmune-mediated destruction spreads, voltage-gated Na^+ (Nav) channel clusters and other components at and near nodes disappear. Voltage-gated K^+ (Kav) channel clusters are preserved unless immune attack extends to the juxtaparanodes. (Reproduced from (22) with permission from the Society for Neuroscience.)

The possible pathogenesis of AMAN subsequent to *C. jejuni* enteritis is as follows: (i) Infection by *C. jejuni* bearing GM1-like and GD1a-like LOS induces the production of anti-GM1 or anti-GD1a IgG antibodies. (ii) These autoantibodies bind to GM1 or GD1a at the nodes of Ranvier in peripheral motor nerves. (iii) Bound anti-GM1 or anti-GD1a IgG antibodies induce local complement activation resulting in the formation of MAC. (iv) The autoimmune attack disrupts Nav channels clusters, producing muscle weakness in the early phase of illness. In severe cases Wallerian-like degeneration subsequently occurs.

3.2. Complement Inhibitor Therapy

Because of the demonstrated complement dependency of the neurotoxic effects of anti-ganglioside antibodies, complement inhibitors are a rational new form of treatment. Murine FS model has been treated successfully with the complement inhibitors Mirococept and eculizumab (24, 25): the former is derived from a region of type 1 complement receptor expressed in bacteria and

is given by local administration to rheumatoid arthritis patients, whereas the latter is a humanized monoclonal antibody that binds to and blocks cleavage of C5 and is approved to treat paroxysmal nocturnal hemoglobinuria which is caused by a genetic defect leading to deficiency of natural membrane-bound complement inhibitors. Nafamostat mesilate (6-amidino-2-naphthyl-*p*-guanidinobenzoate dimethanesulfonate), a synthetic serine protease inhibitor, has been clinically used in Japan over 20 years with no serious adverse effects in patients with disseminated intravascular coagulation, acute pancreatitis, and plasmapheresis. Complement system contains several serine proteases such as C1r, C1s, C3 convertases, and C5 convertases. Nafamostat mesilate so efficiently inhibits these serine proteases that it blocks the formation of the MAC (26–28). At disease onset, rabbits were randomly divided into two groups, and treated using osmotic pumps with or without nafamostat mesilate (29). Total C3 depositions and large or long forms were significantly less in the treated group than in the non-treated group. Disrupted Nav channel clusters were significantly decreased in the group treated by nafamostat mesilate. The experimental results confirm that nafamostat mesilate is a suitable candidate for treatment of AMAN.

As mentioned, GBS is divided into two subtypes, AMAN and AIDP. In AMAN patients as well as in the AMAN rabbits, nodal and internodal axolemmas of motor fibers are intensely immunostained for activated C3 fragments and MAC (1). In patients with AIDP, many fibers have a rim of C3 and MAC along the outer surface of Schwann cells (2). The autopsy study findings indicate that complement activation followed by MAC formation plays an important role in the development of both subtypes. In other words, complement inhibitors such as nafamostat mesilate are likely to be effective for AIDP as well as for AMAN.

4. *C. jejuni* Genes Responsible for Neuropathy Development

As stated earlier, ganglioside mimicry of *C. jejuni* LOS is one cause of GBS. Ganglioside-like LOSs are synthesized by sialyltransferase Cst-II, *N*-acetylgalactosaminyltransferase CgtA and galactosyltransferase CgtB. The *cst-II*, *cgtA*, and *cgtB* genes that encode these enzymes have been cloned and sequenced (30). The *cst-II* gene frequency between GBS/FS and uncomplicated enteritis strains studied did not differ (31). Compared with enteritis-related isolates, GBS-related *C. jejuni* isolates are strongly associated with the expression of GD1a-like LOS rather than with GM1-like LOS (32). The presence of all three

cst-II, *cgtA*, and *cgtB* genes is strongly associated with GBS-related strains rather than with enteritis isolates.

LOS biosynthesis enzymes are encoded by a large gene cluster (33), and many glycosyltransferase genes necessary for synthesizing ganglioside-like structures in the LOS outer core have been cloned and sequenced (34). Gene contents of LOS biosynthesis loci differ between strains creating variations in the LOS structure. These LOS biosynthesis loci have been divided into several classes based on gene organization (34, 35). Classes A–C carry *cgtA*, *cgtB*, and *cst-II* or *cst-III*, and the strains belonging to these classes express ganglioside-like LOSs. Most isolates from GBS and FS patients belong to classes A, B, or C, and the frequency is significantly higher than isolates from uncomplicated enteritis patients (13, 36). *C. jejuni* genes associated with the development of GBS and FS have been found, and they are glycosyltransferase genes essential for synthesizing ganglioside-like LOSs.

The *cst-II* gene encodes an enzyme that transfers sialic acid to LOS, and *neuA1* encodes an enzyme that synthesizes the donor (CMP-sialic acid) used by Cst-II sialyltransferase (30). Because both genes function in LOS sialylation, they are also essential for ganglioside-like LOS synthesis. Mutants of *C. jejuni* that lack these genes have been produced and analyzed (36). Whereas GM1-like and GD1a-like LOSs have been identified in wild-type *C. jejuni* strains isolated from GBS patients, neither has been found in the corresponding *cst-II* and *neuA1* knockout mutants. Unlike the wild types, *cst-II* and *neuA1* knockout mutants have decreased reactivity to GBS patients' sera. Mice lacking GM1 and GD1a are immune-naïve hosts and can be used to obtain high-titer anti-ganglioside antibody responses. Immunization with the wild-type strain induced an anti-GD1a IgG antibody response in these mice, whereas mutant strain immunization did not. This means that the genes involved in ganglioside-like LOS biosynthesis are essential for the induction of cross-reactive anti-ganglioside antibodies during *C. jejuni* infection.

4.1. Sialyltransferase Gene Polymorphism as a Determinant of Clinical Feature

Anti-GQ1b IgG antibodies, which are identified in FS, cross-react with GT1a, indicating that the terminal disialosyl structure is the epitope (37, 38). Our prospective case–control serologic study has shown that FS is related to infections by *C. jejuni* and *Haemophilus influenzae* (39). Mass spectrometry analysis identified a GT1a-like LOS structure that partially mimics GQ1b in a *C. jejuni* strain from a patient with FS (**Fig. 4.4**). A GD1c-like LOS structure, which also partially mimics GQ1b, has also been identified in *C. jejuni* isolates from FS patients (40, 41). Mice lacking GQ1b and GT1a are immune-naïve hosts and can be used to obtain high anti-ganglioside antibody responses. Immunization of these mice with the *C. jejuni* LOS generates monoclonal IgG antibodies that react with GQ1b and GT1a (39).

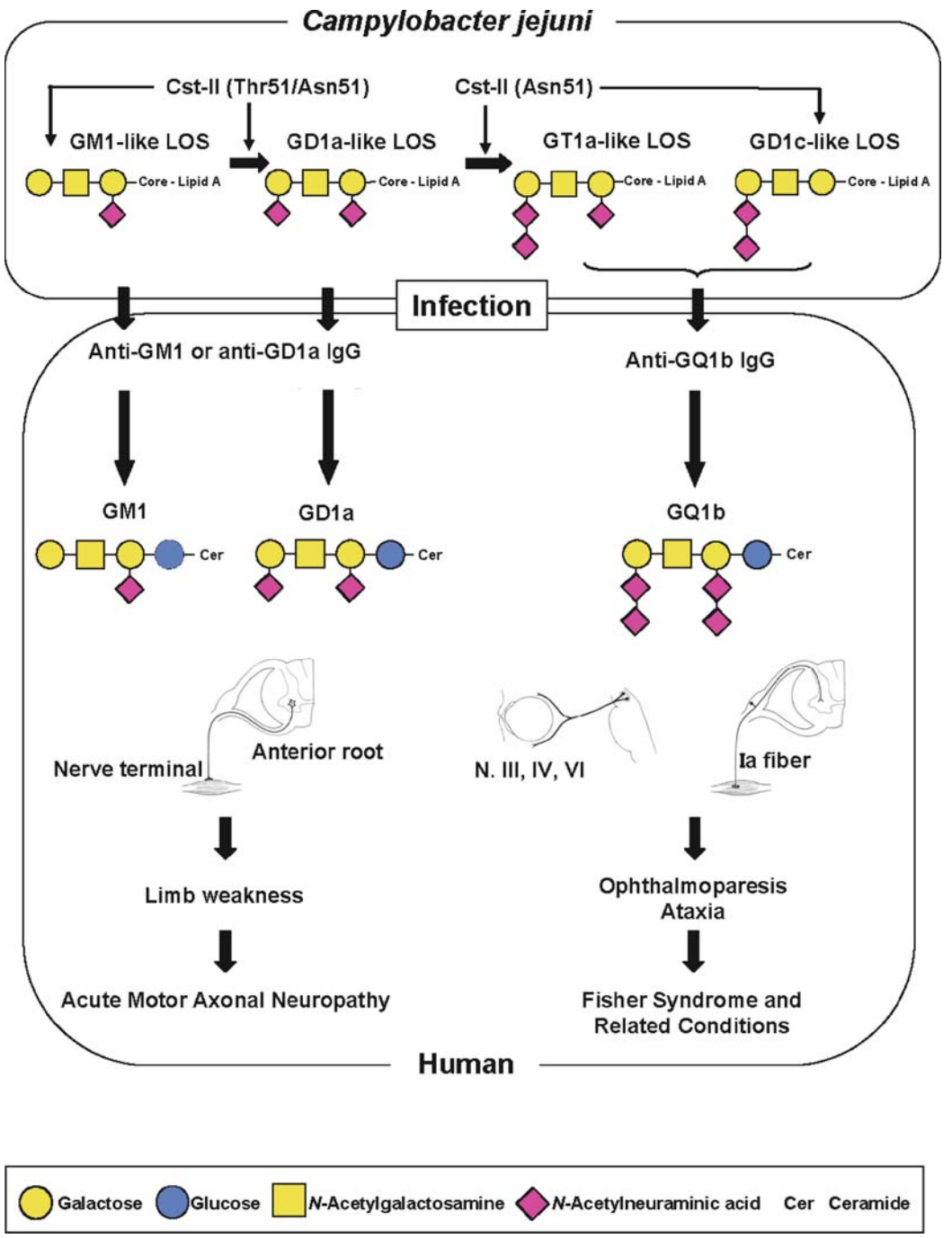


Fig. 4.4. *Campylobacter jejuni* gene polymorphism as a determinant of clinical neuropathies after infection by that bacterium. *C. jejuni* that carries *cst-II* (Thr51) can express GM1-like or GD1a-like LOS on its cell surface. Infection by such a *C. jejuni* strain may induce anti-GM1 or anti-GD1a IgG antibodies in some patients. The autoantibodies bind to the GM1 or GD1a expressed on motor nerves of the four limbs, inducing acute motor axonal neuropathy. By contrast, *C. jejuni* that carries *cst-II* (Asn51) expresses GT1a-like or GD1c-like LOS on the cell surface. Infection by such *C. jejuni* strains may induce anti-GQ1b IgG in some patients. The autoantibodies bind to the GQ1b that is expressed on oculomotor nerves and primary sensory neurons, inducing Fisher syndrome or related conditions. (Reproduced from (45) with permission from John Wiley & Sons, Inc.)

Anti-GQ1b antibodies bind to FS-related *C. jejuni* more commonly than to the GBS-related or enteritis-related strains. An *H. influenzae* strain isolated from a patient with FS carries a GD3-like LOS structure which partially mimics GQ1b (42). Infections by both these bacteria bearing the GQ1b epitope may induce production of anti-GQ1b IgG antibodies.

The reason why a certain microbial infection could induce the development of several autoimmune diseases has yet to be clarified. For example, the mechanism through which group A streptococcal infection induces the development of acute rheumatic fever in some patients and acute glomerulonephritis in others is unknown. The mechanism through which *C. jejuni* induces the development of GBS in some patients and FS in others, however, has been clarified (43). The Cst-II sialyltransferase consists of 291 amino acids, and residue 51 determines its enzymatic specificity (44). Cst-II (Thr51) has only α-2,3-sialyltransferase activity (monofunctional) and produces GM1-like and GD1a-like LOSs (**Fig. 4.4**) (45). In contrast, Cst-II (Asn51) has both α-2,3-sialyltransferase and α-2,8-sialyltransferase activities (bifunctional) and synthesizes GT1a-like and GD1c-like LOSs that mimic GQ1b. Neuropathic strains more frequently had *cst-II*, in particular *cst-II* (Thr51), than did enteritic ones. Strains with *cst-II* (Asn51) regularly expressed the GQ1b epitope, whereas those with *cst-II* (Thr51) had GM1 and GD1a epitopes. The presence of these bacterial epitopes in neuropathy patients corresponded to autoantibody reactivity. Patients infected with *C. jejuni* (*cstII* Asn51) are more often positive for anti-GQ1b IgG and to have ophthalmoparesis and ataxia, whereas patients infected with *C jejuni* (*cstII* Thr51) are more frequently positive for anti-GM1 and anti-GD1a IgG and to have limb weakness.

GM1 and GD1a are expressed on motor nerve axons (46) and GQ1b is expressed in the oculomotor nerves and some primary sensory neurons (37, 47). Briefly, the molecular pathogenesis of GBS or FS subsequent to *C. jejuni* enteritis (**Fig. 4.4**) is as follows: *C. jejuni* (*cstII* Thr51) strains produce GM1-like and GD1a-like LOSs that induce anti-GM1 or anti-GD1a IgG antibodies, causing patients to develop limb weakness. *C. jejuni* (*cstII* Asn51) strains synthesize GT1a-like and GD1c-like LOSs which induce anti-GQ1b IgG antibodies, causing patients to develop ophthalmoplegia and ataxia. In other words, *cst-II* polymorphism (Thr/Asn51) determines whether GBS or FS develops after *C. jejuni* enteritis.

5. Conclusions

Gangliosides contribute to stability of paranodal junctions and ion channel clusters in myelinated nerve fibers. Autoantibodies to GM1 and GD1a and subsequent complement activation disrupt lipid rafts, paranodal or nodal structures, and ion channel clusters in peripheral motor nerves. Convincing evidence has shown that carbohydrate mimicry between ganglioside and LOS is a cause of GBS and FS subsequent to *C. jejuni* enteritis, and that *cst-II* polymorphism determines this carbohydrate mimicry and the clinical presentations, GBS or FS.

Acknowledgments

The dedicated efforts of the scientific staff in my laboratory and our many external collaborators are greatly appreciated. I thank Dr. Keiichiro Susuki (Department of Neuroscience, Baylor College of Medicine) for his help with the figure preparations; and Drs. Bart C Jacobs (Departments of Neurology and Immunology, Erasmus MC), and Michel Gilbert (Institute for Biological Sciences, National Research Council Canada) for their critical reading and English editing of this review.

References

1. Hafer-Macko, C., Hsieh, S.-T., Li, C. Y., Ho, T. W., Sheikh, K., Cornblath, D. R., McKhann, G. M., Asbury, A. K., and Griffin, J. W. (1996) Acute motor axonal neuropathy: an antibody-mediated attack on axolemma. *Ann. Neurol.* **40**, 635–644.
2. Hafer-Macko, C. E., Sheikh, K. A., Li, C. Y., Ho, T. W., Cornblath, D. R., McKhann, G. M., Asbury, A. K., and Griffin, J. W. (1996) Immune attack on the Schwann cell surface in acute inflammatory demyelinating polyneuropathy. *Ann. Neurol.* **39**, 625–635.
3. Ang, C. W., Jacobs, B. C., and Laman, J. D. (2004) The Guillain-Barré syndrome: a true case of molecular mimicry. *Trends Immunol.* **25**, 61–66.
4. Jacobs, B. C., Rothbarth, P. H., van der Meché, F. G. A., Herbrink, P., Schmitz, P. I. M., de Klerk, M. A., and van Doorn, P. A. (1998) The spectrum of antecedent infections in Guillain-Barré syndrome: a case-control study. *Neurology* **51**, 1110–1115.
5. Rees, J. H., Soudain, S. E., Gregson, N. A., and Hughes, R. A. C. (1995) *Campylobacter jejuni* infection and Guillain-Barré syndrome. *N Engl. J. Med.* **333**, 1374–1379.
6. Kuwabara, S., Ogawara, K., Misawa, S., Koga, M., Mori, M., Hiraga, A., Kanesaka, T., Hattori, T., and Yuki, N. (2004) Does *Campylobacter jejuni* infection elicit "demyelinating" Guillain-Barré syndrome? *Neurology* **63**, 529–533.
7. Yuki, N., Yoshino, H., Sato, S., and Miyatake, T. (1990) Acute axonal polyneuropathy associated with anti-GM1 antibodies following *Campylobacter* enteritis. *Neurology* **40**, 1900–1902.
8. Ho, T. W., Willison, H. J., Nachamkin, I., Li, C. Y., Veitch, J., Ung, H., Wang, G. R., Liu, R. C., Cornblath, D. R., Asbury, A. K., Griffin, J. W., and McKhann, G. M. (1999) Anti-GD1a antibody is associated with axonal but not demyelinating forms of Guillain-Barré syndrome. *Ann. Neurol.* **45**, 168–173.

9. Yuki, N., Sato, S., Miyatake, T., Sugiyama, K., Katagiri, T., and Sasaki, H. (1991) Motoneuron-disease-like disorder after ganglioside therapy. *Lancet* **337**, 1109–1110.
10. Illa, I., Ortiz, N., Gallard, E., Juarez, C., Grau, J. M., and Dalakas, M. C. (1995) Acute axonal Guillain-Barré syndrome with IgG antibodies against motor axons following parenteral gangliosides. *Ann. Neurol.* **38**, 218–224.
11. Yuki, N., Taki, T., Inagaki, F., Kasama, T., Takahashi, M., Saito, K., Handa, S., and Miyatake, T. (1993) A bacterium lipopolysaccharide that elicits Guillain-Barré syndrome has a GM1 ganglioside-like structure. *J. Exp. Med.* **178**, 1771–1775.
12. Matsumoto, Y., Yuki, N., van Kaer, L., Furukawa, K., Hirata, K., and Sugita, M. (2008) Guillain-Barré syndrome-associated IgG responses to gangliosides are generated independently of CD1 function in mice. *J. Immunol.* **180**, 39–43.
13. Koga, M., Gilbert, M., Takahashi, M., Li, J., Koike, S., Hirata, K., and Yuki, N. (2006) Comprehensive analysis of bacterial risk factors for the development of Guillain-Barré syndrome after *Campylobacter jejuni* enteritis. *J. Infect. Dis.* **193**, 547–555.
14. Yuki, N., Yamada, M., Koga, M., Odaka, M., Susuki, K., Tagawa, Y., Ueda, S., Kasama, T., Ohnishi, A., Hayashi, S., Takahashi, H., Kamijo, M., and Hirata, K. (2001) Animal model of axonal Guillain-Barré syndrome induced by sensitization with GM1 ganglioside. *Ann. Neurol.* **49**, 712–720.
15. Comín, R., Yuki, N., Lopez, P. H., and Nores, G. A. (2006) High affinity of anti-GM1 antibodies is associated with disease onset in experimental neuropathy. *J. Neurosci. Res.* **84**, 1085–1090.
16. Susuki, K., Nishimoto, Y., Yamada, M., Baba, M., Ueda, S., Hirata, K., and Yuki, N. (2003) Acute motor axonal neuropathy rabbit model: immune attack on nerve root axons. *Ann. Neurol.* **54**, 383–388.
17. Yuki, N., Susuki, K., Koga, M., Nishimoto, Y., Odaka, M., Hirata, K., Taguchi, K., Miyatake, T., Furukawa, K., Kobata, T., and Yamada, M. (2004) Carbohydrate mimicry between human ganglioside GM1 and *Campylobacter jejuni* lipooligosaccharide causes Guillain-Barré syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11404–11409.
18. Yuki, N., Ichihashi, Y., and Taki, T. (1995) Subclass of IgG antibody to GM1 epitope-bearing lipopolysaccharide of *Campylobacter jejuni* in patients with Guillain-Barré syndrome. *J. Neuroimmunol.* **60**, 161–164.
19. Susuki, K., Baba, H., Tohyama, K., Kanai, K., Kuwabara, S., Hirata, K., Furukawa, K., Furukawa, K., Rasband, M. N., and Yuki, N. (2007) Gangliosides contribute to stability of paranodal junctions and ion channel clusters in myelinated nerve fibers. *Glia* **55**, 746–757.
20. Hirota, N., Kaji, R., Bostock, H., Shindo, K., Kawasaki, T., Mizutani, K., Oka, N., Kohara, N., Saida, T., and Kimura, J. (1997) The physiological effect of anti-GM1 antibodies on saltatory conduction and transmembrane currents in single motor axons. *Brain* **120**, 2159–2169.
21. Takigawa, T., Yasuda, H., Kikkawa, R., Shigeta, Y., Saida, T., and Kitasato, H. (1995) Antibodies against GM1 ganglioside affect K^+ and Na^+ currents in isolated rat myelinated nerve fibers. *Ann. Neurol.* **37**, 436–442.
22. Susuki, K., Rasband, M. N., Tohyama, K., Koibuchi, K., Okamoto, S., Funakoshi, K., Hirata, K., Baba, H., and Yuki, N. (2007) Anti-GM1 antibodies cause complement-mediated disruption of sodium channel clusters in peripheral motor nerve fibers. *J. Neurosci.* **27**, 3956–3967.
23. Waxman, S. G. (1995) Sodium channel blockade by antibodies: a new mechanism of neurological disease? *Ann. Neurol.* **37**, 421–423.
24. Halstead, S. K., Humphreys, P. D., Goodfellow, J. A., Wagner, E. R., Smith, R. A., and Willison, H. J. (2005) Complement inhibition abrogates nerve terminal injury in Miller Fisher syndrome. *Ann. Neurol.* **58**, 203–210.
25. Halstead, S. K., Zitman, F. M., Humphreys, P. D., Greenshields, K., Verschuuren, J. J., Jacobs, B. C., Rother, R. P., Plomp, J. J., and Willison, H. J. (2008) Eculizumab prevents anti-ganglioside antibody-mediated neuropathy in a murine model. *Brain* **131**, 1197–1208.
26. Fujii, S., and Hitomi, Y. (1981) New synthetic inhibitors of C1r, C1 esterase, thrombin, plasmin, kallikrein and trypsin. *Biochim. Biophys. Acta* **661**, 342–345.
27. Fujita, Y., Inoue, I., Inagi, R., Miyata, T., Shinzato, T., Sugiyama, S., Miyama, A., and Maeda, K. (1993) Inhibitory effect of FUT-175 on complement activation and its application for glomerulonephritis with hypocomplementemia. *Nippon Jinzo Gakkai Shi* **35**, 393–397.
28. Inagi, R., Miyata, T., Maeda, K., Sugiyama, S., Miyama, A., and Nakashima, I. (1991) FUT-175 as a potent inhibitor of C5/C3 convertase activity for production of C5a and C3a. *Immunol. Lett.* **27**, 49–52.

29. Phongsisay, V., Susuki, K., Matsuno, K., Yamahashi, T., Okamoto, S., Funakoshi, K., Hirata, K., Shinoda, M., and Yuki, N. (2008) Complement inhibitor prevents disruption of sodium channel clusters in a rabbit model of Guillain-Barré syndrome. *J. Neuroimmunol.* **205**, 101–104.
30. Gilbert, M., Brisson, J. R., Karwaski, M. F., Michniewicz, J., Cunningham, A. M., Wu, Y., Young, N. M., and Wakarchuk, W. W. (2000) Biosynthesis of ganglioside mimics in *Campylobacter jejuni* OH4384: identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-MHz ^{1}H and ^{13}C NMR analysis. *J. Biol. Chem.* **275**, 3896–3906.
31. van Belkum, A., van den Braak, N., Godschalk, P. C. R., Ang, W., Jacobs, B. C., Gilbert, M., Wakarchuk, W. W., Verbrugh, H., and Endtz, H. P. (2001) A *Campylobacter jejuni* gene associated with immune-mediated neuropathy. *Nat. Med.* **7**, 752–753.
32. Nachamkin, I., Liu, J., Li, M., Ung, H., Moran, A. P., Prendergast, M. M., and Sheikh, K. (2002) *Campylobacter jejuni* from patients with Guillain-Barré syndrome preferentially expresses a GD1a-like epitope. *Infect. Immun.* **70**, 5299–5303.
33. Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A. V., Moule, S., Pallen, M. J., Penn, C. W., Quail, M. A., Rajandream, M. A., Rutherford, K. M., van Vliet, A. H., Whitehead, S., and Barrell, B. G. (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**, 665–668.
34. Gilbert, M., Godschalk, P. C. R., Parker, C. T., Endtz, H. P., and Wakarchuk, W. W. (2005) Genetic bases for the variation in the lipooligosaccharide outer core of *Campylobacter jejuni* and possible association of glycosyltransferase genes with postinfectious neuropathies. In: Ketley J, Konkel M, editors. *Campylobacter*: Molecular and Cellular Biology. Norwich, UK: Horizon Scientific Press; pp. 219–248.
35. Parker, C. T., Horn, S. T., Gilbert, M., Miller, W. G., Woodward, D. L., and Mandrell, R. E. (2005) Comparison of *Campylobacter jejuni* lipooligosaccharide biosynthesis loci from a variety of sources. *J. Clin. Microbiol.* **43**, 2771–2781.
36. Godschalk, P. C. R., Heikema, A. P., Gilbert, M., Komagamine, T., Ang, C. W., Glerum, J., Brochu, D., Li, J., Yuki, N., Jacobs, B. C., van Belkum, A., and Endtz, H. P. (2004) The crucial role of *Campylobacter jejuni* genes in anti-ganglioside antibody induction in Guillain-Barré syndrome. *J. Clin. Invest.* **114**, 1659–1665.
37. Chiba, A., Kusunoki, S., Obata, H., Machinami, R., and Kanazawa, I. (1993) Serum anti-GQ1b IgG antibody is associated with ophthalmoplegia in Miller Fisher syndrome and Guillain-Barré syndrome: clinical and immunohistochemical studies. *Neurology* **43**, 1911–1917.
38. Chiba, A., Kusunoki, S., Shimizu, T., and Kanazawa, I. (1992) Serum IgG antibody to ganglioside GQ1b is a possible marker of Miller Fisher syndrome. *Ann. Neurol.* **31**, 677–679.
39. Koga, M., Gilbert, M., Li, J., Koike, S., Takahashi, M., Furukawa, K., Hirata, K., and Yuki, N. (2005) Antecedent infections in Fisher syndrome: a common pathogenesis of molecular mimicry. *Neurology* **64**, 1605–1611.
40. Kimoto, K., Koga, M., Odaka, M., Hirata, K., Takahashi, M., Li, J., Gilbert, M., and Yuki, N. (2006) Relationship of bacterial strains to clinical syndromes of *Campylobacter*-associated neuropathies. *Neurology* **67**, 1837–1843.
41. Nam Shin, J. E., Ackloo, S., Mainkar, A. S., Monteiro, M. A., Pang, H., Penner, J. L., and Aspinall, G. O. (1997) Lipo-oligosaccharides of *Campylobacter jejuni* serotype O:10: structures of core oligosaccharide regions from a bacterial isolate from a patient with the Miller-Fisher [sic] syndrome and from the serotype reference strain. *Carbohydr. Res.* **305**, 223–232.
42. Houliston, R. S., Koga, M., Li, J., Jarrell, H. C., Richards, J. C., Vitiazeva, V., Schweda, E. K., Yuki, N., and Gilbert, M. (2007) A *Haemophilus influenzae* strain associated with Fisher syndrome expresses a novel disialylated ganglioside mimic. *Biochemistry* **46**, 8164–8171.
43. Koga, M., Takahashi, M., Masuda, M., Hirata, K., and Yuki, N. (2005) *Campylobacter* gene polymorphism as a determinant of clinical features of Guillain-Barré syndrome. *Neurology* **65**, 1376–1381.
44. Gilbert, M., Karwaski, M. F., Bernatchez, S., Young, N. M., Taboada, E., Michniewicz, J., Cunningham, A. M., and Wakarchuk, W. W. (2002) The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni*: biosynthesis of sialylated ganglioside mimics in the core oligosaccharide. *J. Biol. Chem.* **277**, 327–337.

45. Yuki, N. (2007) Ganglioside mimicry and peripheral nerve disease. *Muscle Nerve.* **35**, 691–711.
46. Gong, Y., Tagawa, Y., Lunn, M. P., Laroy, W., Heffer-Lauc, M., Li, C. Y., Griffin, J. W., Schnaar, R. L., and Sheikh, K. A. (2002) Localization of major gangliosides in the PNS: implications for immune neuropathies. *Brain* **125**, 2491–2506.
47. Kusunoki, S., Chiba, A., and Kanazawa, I. (1999) Anti-GQ1b IgG antibody is associated with ataxia as well as ophthalmoplegia. *Muscle Nerve.* **22**, 1071–1074.

Chapter 5

Biotinylated Multivalent Glycoconjugates for Surface Coating

Alexander A. Chinarev, Oxana E. Galanina, and Nicolai V. Bovin

Abstract

Systematic studying of biological processes driven by multipoint high-cooperative carbohydrate recognition requires application of multivalent carbohydrates as tools. In this regard polyacrylamides with various pendant carbohydrate residues and labels are probably the most well advanced class of carbohydrate multimerics. Here we describe a synthetic approach to polyacrylamide-based glycoconjugates with biotin tag, with special emphasis to development of carbohydrate biosensors and arrays.

Key words: Glycoarrays, glycopolymers, radical polymerization, biotin.

Abbreviations

AC	6-aminocaproyl
AIBN	α,α'-azoisobutyronitrile
biot	biotin
B_{tri}	Galα1-3(Fucα1-2)Gal-
DP	average degree of polymerization
ELISA	enzyme-linked immunosorbent assay
Glyc	glycoside residue
GPC	gel-permeation chromatography
M_n	number average molecular mass
M_w	weight average molecular mass
NPA	4-nitrophenylacrylate
pHEAA	poly(*N*-(2-hydroxyethyl)acrylamide)
pNPA	poly(4-nitrophenylacrylate)
Str	streptavidin.

1. Introduction

Multivalent glycoconjugates based on linear poly(acrylamides) with attached side carbohydrates groups (Glyc) have turned to routine instruments for glycobiology research (1, 2). The latter often demand modification of surfaces with complex

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_5,

carbohydrates in multivalent form (3). A convenient approach of glycopolymer deposition on surface is based on application of biotin–streptavidin (Str) system, when glycopolymers bearing biotin tags are anchored to Str-coated surface (**Fig. 5.1**). This "click"-like procedure is known to give quantitative yield.

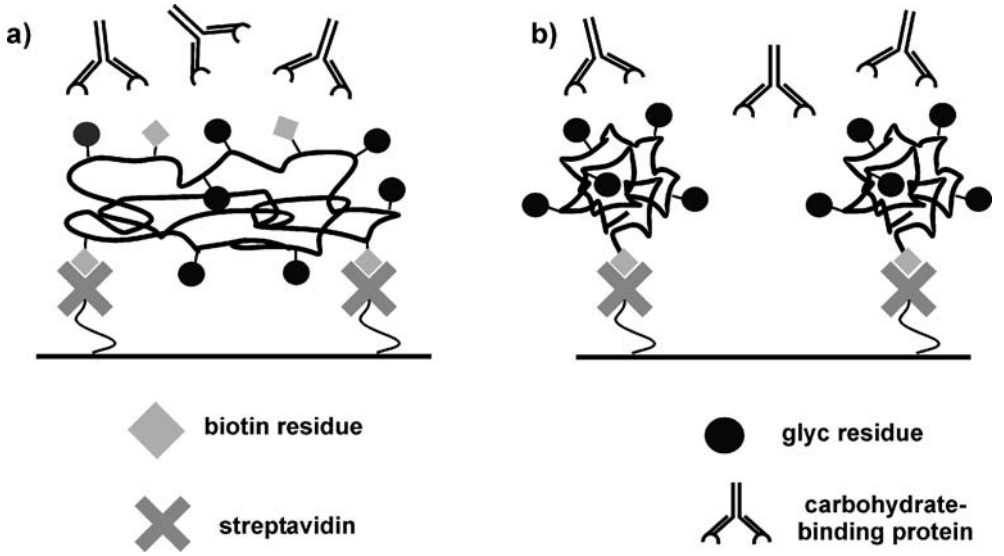


Fig. 5.1. Construction of glycoarrays on Str-coated surface using glycopolymers carrying several side biotin groups (a) or one end biotin group (b) to study carbohydrate specificity of antibodies, lectins, enzymes, cells, pathogens.

We have suggested two approaches to biotinylated glycopolymers. In accordance with the first one, activated polyacrylic acid is initially prepared by radical polymerization of 4-nitrophenylacrylate (NPA). Then, a biotin derivative containing amino group in linker, biot-$NH(CH_2)_6NH_2$, and an ω-aminoalkyl glycoside, Glyc-$O(CH_2)_3NH_2$, are consequently coupled to poly(4-nitrophenylacrylate), pNPA. Finally, the remaining active ester groups in the polymer are quenched by treatment with ethanolamine. Resultant substituted poly (*N*-(2-hydroxyethyl)acrylamide), pHEAA-Glyc$_x$-biot$_y$, contains several Glyc and biotin residues (**Fig. 5.2a**). In the second method, biotin residue is introduced into a polymer scaffold as end group with a fragment of biotinylated initiator. After that, like in the first case, biotinylated poly(4-nitrophenylacrylate), biot$_1$-pNPA, is treated with Glyc-$O(CH_2)_3NH_2$ followed by ethanolamine (**Fig. 5.2b**). The obtained glycopolymer, biot$_1$-pHEAA–Glyc$_x$, with end biotin group possesses a pronounced advantage: in immobilized macromolecules the single end biotin is crypted inside streptavidin matrix, thus does not impact on potential unspecific interactions. Both synthetic routes are considered in detail below. Also we describe a deposition procedure of the obtained biotinylated glycopolymers on the surface

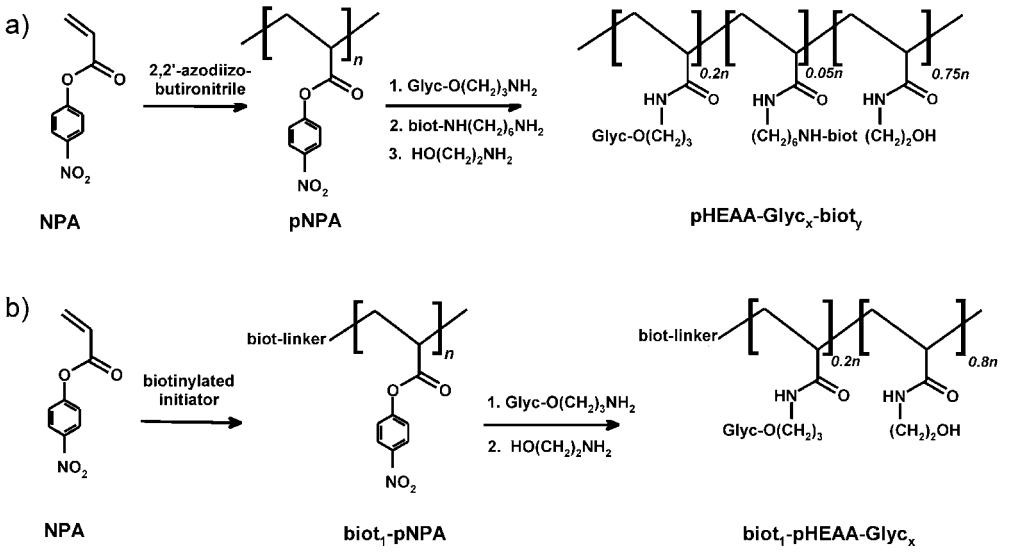


Fig. 5.2. Preparation of biotinylated polyacrylamide glycoconjugates. Synthetic approach leading to glycopolymer with several pendant biotins (2, 4) (a); an approach suggests incorporation of biotin group into pNPA as the fragment of biotinylated initiator followed by ligation with Glyc-$O(CH_2)_3NH_2$ (5) (b); *x*, *y* – average numbers of Glyc and biotin groups per polymer molecule (*see* **Notes 1** and **2**).

of Str-covered plates and the use of these plates in enzyme-linked immunosorbent assay (ELISA).

2. Materials

2.1. Chemicals

1. Acrylic acid 4-nitrophenyl (NPA) ester was synthesized from acryloyl chloride and 4-nitropenyl as described in (7) (*see* **Notes 3, 4** and **5**).
2. α,α'-Azoisobutyronitrile (AIBN) (Fluka).
3. [biot-AC-$NH(CH_2)_5$Co(7-Me-salen)(en)]Br] was synthesized as described in (5) (*see* **Note 6**).
4. B_{tri}-$O(CH_2)_3NH_2$ and B_{tri}-$O(CH_2)_3$NH-biotin (Lectinity Inc., Russia).
5. Biot-$NH(CH_2)_6NH_2$ (Lectinity Holding, Inc., Russia).
6. Kiesselgel 60 (Merck).
7. NEt_3 (Fluka) was kept for 5 days over BaO (Aldrich), and then distilled under fresh BaO.
8. DMF (Fluka) was distilled in vacuum under BaO.
9. Et_2O (Fluka) was distilled with CaH_2 (Sigma).

2.2. Other Commercial Reagents and Solvents

1. Globular protein calibration kit, M_w = 12.4–450 kDa (Serva, USA).
2. Sephadex LH-20 (Pharmacia BioTech, Austria).
3. TSK-2000SW column, 7.5 × 300 mm (Ultrapack, Sweden).
4. Reacti-Bind Streptavidin Coated High Binding Capacity Black 96-Well Plates (Pierce, USA).

2.3. Buffers and Coating Solutions

1. Coating solution: 200 mg/mL pHEAA-B_{tri}(20)-biot (5) or $biot_1$-pHEAA-B_{tri}(20) in PBS (pH 7.4, Sigma).
2. Washing buffer: 0.1% Tween-20 (Sigma) in PBS.
3. Blocking buffer: 3% BSA (Sigma) in PBS.
4. Carbonate buffer, Na_2CO_3–$NaHCO_3$, pH 9.6, was prepared as described in (6).

2.4. Antibodies

1. Mouse monoclonal antibodies B8 against B_{tri} were obtained from All-Russian Hematology Research Center (Moscow, Russia).
2. Anti-mouse IgG+IgM (H+L)-alkaline phosphatase conjugate (Ig-AP) was the product of AP Biotech Inc. (UK).

3. Methods

Two methods of NPA polymerization induced by a typical radical azo-initiator, AIBN, are described. In the first method adopted from literature (7), the polymerization is performed in benzene. The growing macromolecules become insoluble when reached certain molecular weight; thus, the forming polymer is steadily evolved from the reaction mixture during the course of polymerization. Further growth of pNPA molecules presumably takes place in the polymer gel phase, until diffusion of the monomer to the active centers is allowed. Effective control of such occlusion polymerization process is rather difficult, the resultant pNPA possess relatively low meanings of number average molecular weight (M_n = 15–45 kDa) and wide molecular mass distribution (M_w/M_n = 2–7). Moreover, the molecular weight characteristics of the polymer are poorly reproducible in this method, and different batch of pNPA may noticeably differ in this respect.

The second method of NPA polymerization is original; the polymerization is carried on in solution, i.e., in homogenous conditions; DMSO is used as the solvent (8). In this case, molecular mass characteristic of the obtained polymer can be easily controlled. Thus, in consistency with the theory of

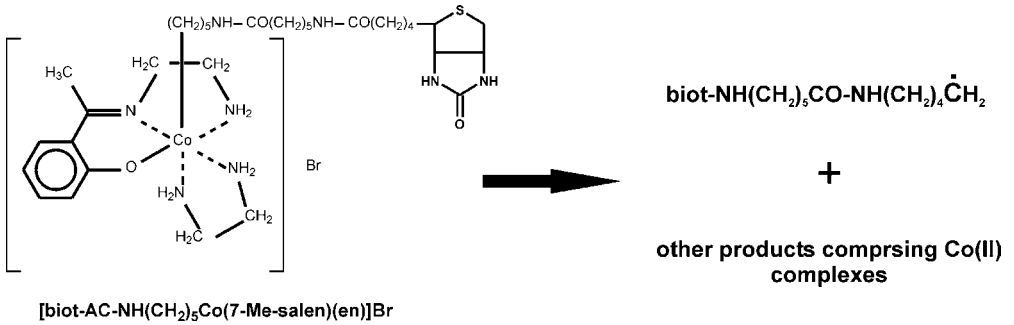
radical polymerization, a decrease in AIBN concentration ($[I]_0$) caused an increase in pNPA number average molecular weight (M_n = 110–145 kDa). The growth of M_n values is also observed upon increase in the monomer concentration ($[M]_0$), although this dependence is less pronounced. Noteworthy, the polymer obtained by this method has narrower molecular mass distribution than pNPA prepared in accordance with the first approach (*see* **Table 5.1**).

Table 5.1
Molecular weight characteristics of activated polyacrylic acids depending on the amount of initiator taken for polymerization of 4-nitrophenylacrylate ($[I]_0$). Starting concentration of monomer in polymerizing mixture ($[M]_0$) was 1 mol/L for all experiments

	$[I]_0$, mass % on monomer weight	M_n, kDa	M_w/M_n
pNPA, Method 1	5	~15–45	2–7
pNPA, Method 2	0.5	110	1.85
	0.25	135	2.01
	0.1	145	2.03
$Biot_1$-pNPA	3	35	1.88
	1.5	40	1.93
	0.5	47	2.15

Finally, we describe a method leading to an active ester homopolymer with end biotin group, $biot_1$-pNPA. To accomplish this biotinylated organocobalt(III) chelate, [biot-NH$(CH_2)_5$Co(7-Me-salen)(en)]Br], is used as initiator of NPA polymerization in DMSO. Homolytic splitting of the chelate Co–C bonds taking place with heating of the reaction mixture leads to carbon-centered radicals, biotin-AC-NH$(CH_2)_5$ (*see* **Fig. 5.3**). The latter triggers polymer formation and provides incorporation of biotin end group into pNPA molecule. Starting concentration of the initiator in reaction mixture may be varied to get the polymers of different molecular weights (*see* **Table 5.1**).

Small part of the activated polymers, pNPA and $biot_1$-pNPA, was converted into pHEAA and $biot_1$-PHEAA, correspondingly, and number average molecular mass (M_n) was defined by analytical gel-permeation chromatography (GPC). This is an important



[biot-AC-NH$(CH_2)_5$Co(7-Me-salen)(en)]Br

Fig. 5.3. Structure of the biotinylated organocobalt initiator and the scheme of its decomposition due to splitting of the alkyl–Co bond. Noteworthy, only one radical is released under the decomposition, which diminishes the "cage" effect and enhances the initiator efficiency.

characteristic of the polymer scaffold as it defines number of pendant side groups in glycoconjugate and hence its capacity to bind carbohydrate recognizing proteins. Data on molecular mass properties of pNPA and biot$_1$-pNPA obtained under different conditions are given in **Table 5.1**.

A universal approach used earlier to get a wide variety of label-free neoglycoconjugates as well as probes bearing different label types, neoglycolipids, adsorbents, etc. (1, 2), is adopted here for biotinylated polyacrylamide glycoconjugate synthesis (**Fig. 5.2**). Consequent condensation of the aminoligands, biot-$NH(CH_2)_6NH_2$ and B_{tri}-$O(CH_2)_3NH_2$, with the activated polyacrylic acids, pNPA (M_n~23 kDa, Method 1) and biot$_1$-pNPA (M_n~47 kDa), is carried on in DMSO at 40°C with quantitative yield (*see* **Notes 1** and **2**). Then, the non-reacted activated acryloyl groups of the polymer are quenched by treatment with an excess of ethanolamine giving rise to inert *N*-(2-hydroxyethyl)acrylamide form. The resultant glycopolymers bear Glyc residues statistically distributed along the backbone; biotins are attached to the polymer scaffold as side (for pHEAA-Btri(20)-biot (5)) or end groups (for biot$_1$-pHEAA-B_{tri}(20)).

The prepared biotinylated polyacrylamide glycoconjugates are used to coat the surface of commercial Str-covered plates for ELISA. The coating procedure is simple and consists just in incubation of the plates with the biotinylated glycopolymers in buffer solution and washing. In the similar way we coated commercial fluorescent beads for suspension assay, modified surface of red blood cells aimed at designing serological reagents with controllable hapten density and printed the polymers onto microchip (5).

Binding of monoclonal antibodies (mAbs) specific to B_{tri} to the glycopolymers, pHEAA-Btri(20)-biot (5) and biot$_1$-pHEAA-B_{tri}(20), coated onto surface of polystyrene plates, was tested using ELISA. For comparison, mAbs binding with the plates coated with the monomeric glycoside, B_{tri}-*O*$(CH_2)_3$NH-biotin,

was determined. The intensity of the signal detected in ELISA was plotted against glycoconjugate amount (**Fig. 5.4**).

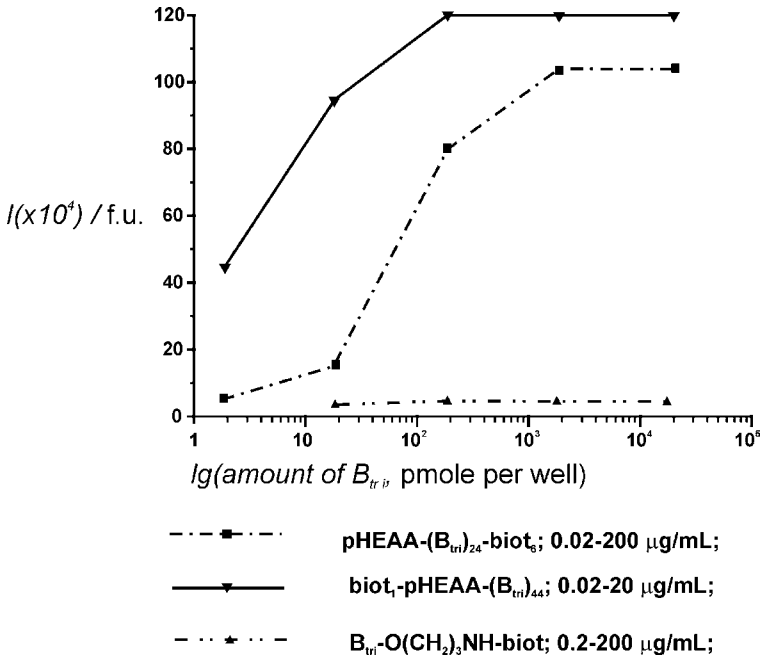


Fig. 5.4. Dependence of fluorescent signal in ELISA on the amount of B_{tri} added per well as the biotinylated glycoconjugates. The plot displays higher binding of mAbs with the glycopolymers, whereas the lack of binding is observed with the monomeric conjugate. Our earlier data showed that the difference in binding affinity between the glycopolymers cannot be attributed to the observed minor disparity in molecular weight and valence (1, 2). It is rather connected with their different anchoring on surface and, correspondingly, favored presentation of glycoligands on surface (5).

3.1. Chromatography

3.1.1. Thin Layer Chromatography (TLC)

1. TLC was performed on silica gel covered plates Kiesselgel 60.
2. Spots were visualized by charring a plate with 7% H_3PO_4, by a plate exposure to NH_3 vapor (for NPA) or by ninhydrin treatment (for aminocompounds).

3.1.2. Preparative GPC

1. GPC was performed on Sephadex LH-20 in MeCN/H_2O, 1:1, using 75 × 1.5 cm column.

3.1.3. Analytical GPC

1. Analytical gel-permeation chromatography of pHEAA and $biot_1$-pHEAA was carried out using HPLC TSK-2000SW column, 7.5 × 300 mm; mobile phase 0.2 M NaCl; flow rate –1 mL/min; and UV detection at 210 nm.
2. The column was calibrated with a globular protein calibration kit, M_w = 12.4–450 kDa.

3. Based on the polymer elution profiles, molecular weight distribution differential curves were plotted by the Schultz method and analyzed using standard approaches (9) to enable the calculation of the polymer molecular weight characteristics M_n, and M_w/M_n (*see* **Table 5.1**).

3.2. Preparation of Activated Poly(acrylic) Acids

3.2.1. Poly (4-Nitrophenylacrylate), pNPA. Method 1

1. Apparatus consisting of a two-neck round-bottomed 50 mL flask equipped with magnetic stirrer, inlet tube (not too narrow in diameter) provided with current of nitrogen, and a water condenser was assembled. The apparatus was closed by a calcium chloride tube; care should be taken before the experiment to ensure that there is a free gas passage through the tube.
2. 1 g (5.18 mmol) NPA, 50 mg (0.3 mmol) AIBN, and 12 mL benzene were placed in the flask.
3. The mixture was stirred under slow nitrogen flow (extra pure) for 1 h and then heated at 70°C for 24 h.
4. Precipitate evolved from the reaction mixture was collected on a glass filter, dissolved in 10 mL DMF and re-precipitated by adding 30 mL MeOH.
5. The product was washed on a glass filter with 3×30 mL MeOH and dried in vacuum.
6. Yield of the product ~60%; fine yellowish powder (*see* **Note 7**).

3.2.2. Poly (4-Nitrophenylacrylate), pNPA. Method 2

1. A solution containing 1 g (5.18 mmol) NPA and 1–5 mg AIBN in 5.2 mL DMSO was placed in a long-stem glass 10 mL ampoule.
2. The mixture was degassed by 3–5 times repeating of freeze–pump–thaw cycles; the ampoule was flame-sealed under vacuum.
3. The mixture was heated at 70°C for 24 h.
4. The ampoule was opened, and the reaction mixture was poured into flask with 30 mL intensively stirred anhydrous Et_2O to wash from DMSO. At times Et_2O was replaced with fresh portions, until the polymer did not turn into glassy state.
5. Pieces of the glassy polymer were pounded and then stirred with Et_2O to form fine powder.
6. The product was collected on a glass filter and dried in vacuum.

7. Yield of the product 50–70%; fine yellowish powder (*see* **Note 7**).

3.2.3. Poly (4-Nitrophenylacrylate) with End Biotin Group, $biot_1$-pNPA

1. A solution containing 1 g (5.18 mmol) NPA and 5–30 mg [biot-NH$(CH_2)_5$Co(7-Me-salen)(en)]Br] in 5.2 mL DMSO was placed in a long-stem glass 10 mL ampoule.
2. The mixture was degassed by 3–5 times repeating of freeze–pump–thaw cycles; the ampoule was flame-sealed under vacuum.
3. The mixture was heated at 70°C for 24 h.
4. The further treatment was the same as for pNPA (*see* **Sect. 3.2.2**).
5. Yield of the product –50 to 70%; fine olive colored powder (*see* **Notes 7** and **8**).

3.3. Preparation of Poly (N-(2-Hydroxyethyl)acrylamide), pHEAA and $biot_1$-pHEAA

1. One hundred microliters of ethanolamine (#411000, Aldrich) was added to a solution of 30 mg pNPA or $biot_1$-pNPA in 1 mL DMSO and the mixture was kept for 48 h under 70°C.
2. Product was purified by GPC, fractions with product were evaporated, and the remains were dried in vacuum.
3. Yield 93–95%; white solids.

3.4. Preparation of Polyacrylamide-Based Glycoconjugates

3.4.1. Preparation of Glycopolymers with Side Biotin Groups, pHEAA-B_{tri}(20)-biot (5)

1. To a solution of 4.7 mg (27.5 μmol) pNPA in 400 mL DMSO (*see* **Note 7**) were added 0.5 mg (1.3 μmol) biot-NH$(CH_2)_6NH_2$ and 0.4 mL (2.6 μmol) NEt_3 (*see* **Notes 9** and **10**).
2. A solution containing 3 mg (5.5 μmol) B_{tri}-*O*$(CH_2)_3NH_2$ and 1.5 μL (11 μmol) NEt_3 in 100 mL DMSO was mixed with the solution of pNPA.
3. The mixture was kept for 12 h at 40°C, absence of the free glycoside was confirmed by TLC (*see* **Note 11**).
4. An excess of ethanolamine (5 equiv.) was added to the reaction and it was kept for 24 h at 40°C.
5. The product was purified by GPC, fractions containing the product were evaporated and the residue was dried in vacuum.
6. Yield 85–95%; white solid.

3.4.2. Preparation of Glycopolymers with End Biotin Group, $biot_1$-pHEAA-B_{tri}(20)

1. To a solution of 4.7 mg (27.5 μmol) $biot_1$-pNPA in 500 mL DMSO were added 3 mg (5.5 μmol) B_{tri}-O$(CH_2)_3NH_2$ and 1.5 μL (11 μmol) NEt_3.

2. The further treatment was the same as for pHEAA-B_{tri}(20)-biot (5) (*see* **Sect. 3.4.1**)
3. Yield 85–95%; white solids.

3.5. Coating of Str-Plates with Biotinylated Glycopolymers

1. Reacti-Bind Streptavidin Coated High Binding Capacity Black 96-Well Plates were rinsed twice with PBS.
2. Serial 10-fold dilution of coating solution in PBS (containing 0.02–200 mg/mL biotinylated glycoconjugate) was added to the plates (100 mL per well); the plates were kept for 1 h at 37°C.
3. The plates were washed (hereinafter, the washing buffer, 200 mL per well).
4. The plates were blocked (the blocking buffer, 200 mL per well) and washed.

3.6. ELISA

1. B8 mAbs (1:100 in PBS containing 0.3% BSA) were added to the plates (100 mL per well).
2. The plates were incubated for 1 h at 37°C and washed.
3. Ig-AP (1:5000 in PBS containing 0.3% BSA) was added to the plates (100 mL per well); the plates were incubated for 1 h at 37°C and washed.
4. A solution of 10^{-4} M 4-methylumbelliferyl phosphate disodium salt in the carbonate buffer was added to the plates (100 mL per well); the plates were incubated for 30 min at room temperature.
5. Fluorescence intensity (355 nm/460 nm) was measured by Victor^2 multilabel counter. Each assay was done in duplicate, and blank reaction was performed by omitting mAb. The blank reading was subtracted from the final fluorescence to provide the corrected fluorescence intensity values.

4. Notes

1. Routinely, molar fractions of Glyc and biotin in pHEAA-$Glyc_x$-$biot_y$ are 20 and 5%, respectively; this optimal loading be changed if necessary. In glycopolymer notations, meanings of the molar fractions are given in brackets for the corresponding ligands; for example, pHEAA-Glyc(20)-biot (5) or $biot_1$-pHEAA-Glyc(20).
2. Obviously, the average number of pendant Glyc and biotin groups (x, y) depends not only on the substitution rate of the acrylate units with corresponding residues, but also on

the molecular weight of the polymer. Polyacrylamide with higher molecular weight can bear larger number of the pendant residues.

3. Some characteristics of NPA are given bellow. NPA: yellow needle crystals; m.p. 44–47°C; TLC: toluene/Me_2CO, 4:1, R_f 0.68; ^1H NMR (δ, $CDCl_3$, 303 K): 6.11 (dd, 1H, J_{cys} 1 Hz, J_{trans} 10.6 Hz, $CH_2 = CHCO$), 6.35 (dd, 1H, J_{hem} 17.4 Hz, J_{trans} 10.6 Hz, $CH_2 = CH$), 6.67 (dd, 1H, J_{cys} 1 Hz, J_{hem} 17.4 Hz, $CH_2 = CH$), 7.35 and 8.30 (d, J 7.6 Hz, 4H, Ar).
4. The monomer taken for polymerization could contain up to 20% of oligomeric admixtures, this does not affect the polymerization course.
5. The monomer should be kept at 2–8°C avoiding contact with moisture and amines.
6. The biotinylated organocobalt chelate should be kept in darkness at –20°C avoiding contact with acids.
7. pNPA and $biot_1$-pNPA should be kept at 2–8°C avoiding contact with moisture and amines.
8. $Biot_1$-pNPA is slightly impure with cobalt-containing products of [biot-$NH(CH_2)_5$Co(7-Me-salen)(en)]Br] decomposition. Nevertheless, it proved fit for further modifications without additional purification.
9. Concentration of pNPA or $biot_1$-pNPA solutions in DMSO taken for glycoconjugates preparation should be in the range 10–20 mg/mL, otherwise gelation of the polymers can occur.
10. Stock solution of pNPA-$biot_x$ can be prepared, stored at 5°C for a long time without decomposition of the active ester groups, and used to obtain a glycoconjugate if necessary.
11. To control coupling of the aminoligands with pNPA and $biot_1$-pNPA, TLC may be used. Recommended eluents: MeOH/1 M Py·HOAc 3:1, R_f 0.54, for biot-$NH(CH_2)_6NH_2$; EtOH/BuOH/Py/H_2O/AcOH 10:1:1:1:0.3, R_f 0.3, for B_{tri}-$O(CH_2)_3NH_2$.

Acknowledgments

This work was supported by the RAS Presidium Program “Molecular and cell biology” and RFBR grant # 07-04-00630.

References

1. Bovin, N.V. (2003) Neoglycoconjugates as probes in glycobiology. In Chemical Probes in Biology (M. P. Schneider Ed.), Kluwer Academic Publishers, Dordrecht, pp. 207–225.
2. Bovin, N.V. (1998) Polyacrylamide-based glycoconjugates as tools in glycobiology. *Glycoconj. J.* **15**: 431–446.
3. Consortium for Functional Glycomics Home Page. http://www.functionalglycomics.org.
4. Bovin, N.V., Korchagina, E.Yu., Zemlyanukhina, T.V., et al. (1993) Synthesis of polymeric neoglycoconjugates based on N-substituted polyacrylamide. *Glycoconj. J.* **10**: 142–151.
5. Bovin, N., Chinarev, A. (2008) Bioanalytic systems and methods. US Patent Appl 0076137.
6. Dawson, R.M.C., Elliot, D.C., Jones, K.M. Data for Biochemical Research. Clarendon, Oxford, 1986.
7. Mammen, M., Dahmann, G., Whitesides, G.M. (1995) Effective inhibitors of hemagglutinin by influenza virus synthesized from polymers having active ester groups. Insight into mechanism of inhibition. *J. Med. Chem.* **38**:4179–4190.
8. Dikusar, M.A., Kubrakova, I.V., Chinarev, A.A., Bovin, N.V. (2001) Polymerization of 4-nitrophenyl acrylate under microwave heating conditions. *Russ. J. Bioorg. Chem. (Engl. Transl.)* **27**: 408–412.
9. Rabek, J.F. (1980) Experimental Methods in Polymer Chemistry. Wiley, New York.

Chapter 6

Profiling LPS Glycoforms of Non-typeable *Haemophilus influenzae* by Multiple-Stage Tandem Mass Spectrometry

Elke K.H. Schweda and James C. Richards

Abstract

Non-typeable (acapsular) *Haemophilus influenzae* (NTHi) is a major cause of otitis media accounting for 25–30% of all cases of the disease. Lipopolysaccharide (LPS) is an essential and exposed component of the *H. influenzae* cell wall. A characteristic feature of *H. influenzae* LPS is the extensive inter-strain and intra-strain heterogeneity of glycoform structure which is key to the role of the molecule in both commensal and disease-causing behavior of the bacterium. However, to characterize LPS structure unambiguously is a major challenge due to the extreme heterogeneity of glycoforms that certain strains express. A powerful tool for obtaining sequence and branching information is multiple-stage tandem ESI-MS (ESI-MS^n) performed on dephosphorylated and permethylated oligosaccharide material using an ESI-quadrupole ion trap mass spectrometer. In general, permethylation increases the MS response by several orders of magnitude and sequence information is readily obtained since methyl tagging allows the distinction between fragment ions generated by cleavage of a single glycosidic bond and inner fragments resulting from the rupture of two glycosidic linkages. Using this approach we are now able to identify all isomeric glycoforms in very heterogeneous LPS preparations.

Key words: Multiple-stage tandem mass spectrometry (MS^n), lipopolysaccharide, *Haemophilus influenzae*, isomeric glycoforms.

1. Introduction

Haemophilus influenzae is an important cause of human disease worldwide and exists in encapsulated (types a–f) and unencapsulated (non-typeable) forms. Type b capsular strains are associated with invasive bacteremic diseases, including meningitis, epiglottitis, cellulitis, and pneumonia, while acapsular or non-typeable strains of *H. influenzae* (NTHi) are a significant cause of otitis

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_6,

media and both acute and chronic lower respiratory tract infections (1). The potential of *H. influenzae* to cause disease depends on its surface expressed carbohydrate antigens, which include capsular polysaccharide in capsulated strains (2) and lipopolysaccharide (LPS) (3).

LPS is an essential and characteristic surface component of *H. influenzae*. This bacterium has been found to express short-chain LPS, lacking O-specific polysaccharide chains and is often referred to as lipooligosaccharide (LOS). Structure, genetics, and expression of variant LPS glycoforms and their implications on virulence have recently been reviewed (4).

The structure of LPS from *H. influenzae* has been extensively investigated and a structural model has been advanced consisting of a conserved tri-heptosyl inner-core moiety attached to a lipid A component via a phosphorylated 3-deoxy-D-*manno*-octulosonic acid (Kdo) linker (*see* **Fig. 6.1**).

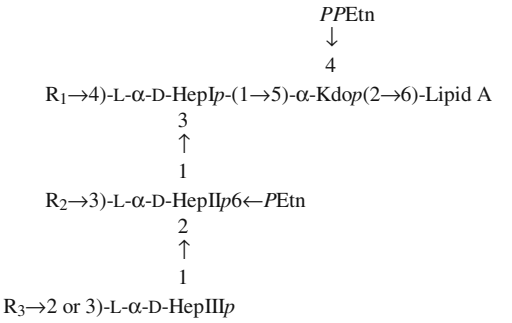


Fig. 6.1. Schematic representation of the inner core structure of *Haemophilus influenzae* LPS where R_1, R_2, R_3=H, sugar residue or OS unit.

H. influenzae lipid A has a relatively conserved structure consisting of a partially acylated β-1,6-linked D-glucosamine disaccharide which is phosphorylated at the 1 and 4´positions (5, 6). The structures of the oligosaccharide (OS) regions of *H. influenzae* LPS have been extensively investigated (7–24). Each of the inner-core L-*glycero*-D-*manno*-heptose units (HepI, HepII, and HepIII) can provide a point for attachment of sugar units and further oligosaccharide extension (**Fig. 6.1**) resulting in considerable intra-strain and inter-strain heterogeneity (4). Moreover, the presence of phosphate substituents, including free phosphate (*P*), phosphoethanolamine (*P*Etn), phosphocholine (*P*Cho), and O-acyl substituents (acetyl and glycine), contributes to the structural variability. Intra-strain variability in structure can result from (i) incomplete biosynthesis during stepwise addition of sugar residues, (ii) competition during biosynthesis when a particular sugar residue is the acceptor for two or more transferases, (iii) variable addition to more than one acceptor by a single transferase, and (iv) phase variable expression of genes encoding

specific transferases. This can lead to a manifold of isomeric OS structures (isomeric glycoforms) from a single strain greatly increasing the analytical challenge for structural analysis. In addition, strain-to-strain variation resulting from genomic differences adds to the overall structural diversity of short-chain LPS (4).

Structural analysis of short-chain LPS has traditionally involved initial delipidation to generate water-soluble oligosaccharides followed by characterization by mass spectrometry (MS) and nuclear magnetic resonance (NMR) methods. MS techniques have played an increasingly important role in structural profiling and oligosaccharide characterization over the last 10–15 years. Ionization methods involving electrospray (ESI) and matrix-assisted laser desorption (MALDI) in conjunction with quadrupole, ion trap, time-of-flight and Fourier transform mass spectrometers have provided detailed information on both core OS and lipid A structures (25, 26). When combined with multiple-stage tandem MS techniques (MS^n) and on-line high performance liquid chromatography (HPLC), information on isomeric glycoforms in heterogeneous mixtures can readily be obtained. This technique has been extensively applied to permethylated oligosaccharide samples following acid-catalyzed delipidation and dephosphorylation of *H. influenzae* LPS (17–19, 21–24, 27–30).

2. Materials

2.1. Growth of Bacterial Strains

1. Chocolate agar plates.
2. Liquid growth medium: brain–heart infusion broth (BHI) supplemented with hemin (10 μg/mL) and nicotinamide adenine dinucleotide (NAD) (2 μg/mL).
3. Ethanol, 95%, acetone and light petroleum, 35–60°C.

2.2. Extraction of LPS

1. Phenol:chloroform:light petroleum (PCP)-extraction: phenol, 90%, phenol, chloroform, light petroleum, 35–60°C.
2. Acetone, ethyl ether.

2.3. Preparation and Derivatization of OS Samples

2.3.1. Delipidation and Dephosporylation

1. Glacial acetic acid, borane-4-methylmorpholine complex.
2. 48% Aqueous hydrofluoric acid (HF).

2.3.2. Permethylation

1. Dimethylsulfoxide (DMSO), butyllithium 15% in *n*-hexane, methyl iodide, ethanol, 95%, acetonitrile, acetonitrile 10% in water, Sep-Pak C18 cartridge.

2.4. Profiling of Carbohydrate Backbone Oligosaccharides by ESI-MS^n

1. Gradient solvents: (A) methanol; (B) ultrapure water, 18.2 MΩ, each containing 0.1 mmol anhydrous sodium acetate, p.a., and 1% glacial acetic acid.
2. HPLC: 2690 Waters HPLC system (Waters, Milford) with autosampler and a solvent sparging system using helium. Column: microbore Luna 5μ C18 (2); 100°A, 150×0.50 mm; (Phenomenex). Store the column at 21°C in methanol/water (1:1, v/v).
3. Electrospray ionization quadrupole ion trap mass spectrometer (ESI-MS): LCQ (ThermoFinnigan) used in positive mode.

3. Methods

3.1. Growth of Bacterial Strains

Expression levels of certain oligosaccharide epitopes of *H. influenzae* LPS are dependent on growth conditions (*see* **Notes 1** and **2**). The following describes our general protocol for culturing in liquid media (31).

1. Resuscitate bacterial strains from frozen stocks on chocolate agar plates and incubate overnight at 37°C. Select colonies from plates and cultivate in 10-L batches of BHI broth supplemented with hemin and NAD at 37°C for 20 h.
2. Harvest cells by centrifugation and discard centrifugate. Cells are then dehydrated by successive suspension and pelleting in ethanol, then acetone, and finally light petroleum.
3. Cells are air-dried to remove organic solvents prior to lyophilization.

3.2. Extraction of LPS

1. Prepare mixture of 90% phenol:chloroform:light petroleum (PCP) in the ratio 2:5:8 (v/v/v). If it is cloudy add crystalline phenol in small portions under stirring until it becomes a clear solution.
2. Stir bacterial cells (approx. 2.4×*g*) in 75 mL PCP overnight. After centrifugation, combine the PCP phases and remove any precipitation by filtering the solvent through filter paper.
3. Chloroform and light petroleum are then removed by evaporation and the LPS is precipitated from the aqueous phenol by adding 120 mL acetone/ether (5:1, v/v). After centrifugation (7500 rpm) the LPS is washed with acetone and ether. Finally the LPS can be further purified by ultracentrifugation (27,000 rpm).

3.3. Preparation and Derivatization of OS

Short-chain *H. influenzae* LPS comprises heterogeneous populations of low-molecular-mass hydrophilic oligosaccharide components that are covalently linked to a hydrophobic lipid A moiety (*see* **Fig. 6.1**). LPS molecules are poorly soluble in aqueous media as well as common organic solvents. Thus, structural profiling of *H. influenzae* LPS involves in most cases initial delipidation to obtain water-soluble oligosaccharides that are suitable for subsequent analyses (*see* **Note 3**). For isomeric glycoform profiling we have employed mild acid hydrolysis of intact or O-deacylated LPS with dilute aqueous acetic acid which affords insoluble lipid A and a soluble OS fraction.

3.3.1. Delipidation and Dephosphorylation

Mild acid hydrolysis of LPS with dilute aqueous acetic acid takes advantage of the acid lability of Kdo ketosodic linkage that joins the LPS oligosaccharide components to the lipid A moiety. The general protocol involving a gel filtration step was provided previously (32). Here we have adapted the protocol for microanalysis without gel filtration to include a dephosphorylation step (*see* **Notes 4** and **5**).

1. Prepare aqueous acetic acid solution by adding glacial acetic acid to water until pH 3.1 is reached. Add borane-N-methylmorpholine complex to a concentration of 1 mg/mL.
2. Transfer LPS or O-deacylated LPS (1 mg) to Eppendorf tube (1.5 mL) and add 1 mL of the aqueous acetic acid solution. Place it at 90°C for 2 h.
3. Cool the sample on ice and centrifuge it for 25 min (14,000 rpm). Lyophilize the water-soluble part containing OS in another Eppendorf vial.
4. OS is treated with 0.1 mL chilled 48% HF for 48 h at 4°C. Then unreacted HF is evaporated with a stream of nitrogen at 0°C following several lyophilizations with water (2×1 mL).

3.3.2. Permethylation

Permethylation of OS samples has proven to be effective in increasing the MS response by several orders of magnitude compared to underivatized samples. This provides a more complete picture of intra-strain and inter-strain variation as well as environmental effects (e.g., during the course of disease pathogenesis) on the distribution of isomeric glycoforms (33). The method involves the conversion of all free hydroxyl groups on the oligosaccharide to the corresponding methyl ethers using the strong base, butyllithium in DMSO to facilitate reaction with methyl iodide (*see* **Note 6**).

1. Transfer dry dephosphorylated OS to a serum flask (5 mL) containing a stirrer and seal the flask with a rubber septum. All subsequent additions of reagents are made using a 1-mL

glass syringe. Add 0.5 mL of dry DMSO, flush gently with N_2 and stir for 2 h at RT.

2. Cool serum flask on ice and slowly add 0.4 mL of butyllithium. Then warm reaction mixture in an oil bath at 40°C for 1 h.
3. Cool vial on ice and add 0.4 mL methyl iodide and stir when melted for at least 2 h at 20°C. Remove excess methyl iodide with vacuum and add 0.5 mL H_2O.
4. Permethylated dephosphorylated OS is recovered using a Sep-Pak C18 cartridge. Precondition the cartridge by subsequent rinsing with 10 mL ethanol, 4 mL acetonitrile, and 4 mL water. Apply sample and wash the Sep-Pak C18 with 10 mL H_2O followed by 10 mL 10% CH_3CN in H_2O.
5. Permethylated OS is then eluted with acetonitrile (2×2 mL), collected into a 13×100 mm screw cap tube and concentrated to dryness.

3.4. Profiling by HPLC-ESI-MS^n

ESI-MS is a sensitive analytical procedure that is ideally suited for profiling the structural diversity of short-chain LPS (32). Direct and on-line capillary electrophoretic analysis of O-deacylated LPS in the negative ion mode or of core OS in the positive ion mode has been described (34, 35). In the present application, permethylated OS samples derived from LPS are analyzed by HPLC coupled to an ion trap spectrometer for characterization by multiple-stage MS (MS^n) in the positive ion mode (*see* **Notes 7 and 8**).

1. Dissolve the dephosphorylated and permethylated sample in 150 μL methanol/water (7:3, v/v).
2. Using the above described instrumentation for HPLC–ESI-MS, select a solvent flow rate of 0.020 mL/min. Select a sample volume of 10 μL to be injected automatically onto the column. Select a column temperature of 21°C.
3. The gradient starting condition is solvent A/B (60/40) which is increased to 100% A within 50 min and then kept at 100% for 10 min. After completed run, allow the column to equilibrate for 20 min at starting conditions.
4. In the electrospray ionization mode, apply a capillary voltage of 3 kV. Singly charged sodium adduct ions ($[M+Na]^+$) are identified in either full scan MS or multiple-stage tandem MS/MS using an instrument tune file for optimized detection at a range of *m/z* 150–3000. For full scan MS/MS analysis select parent ions with an isolation width of *m/z* 2 and fragment in the ion trap by collision-induced dissociation using helium. Apply a maximum injection time of 200 ms using a number of 3 microscans. For data acquisition apply centroid scan. Please note that instrument settings for detection might vary depending on the instrument in use.

4. Notes

1. It is important to carry out growth and manipulations of *H. influenzae* bacteria under containment level II conditions to ensure proper biosafety. Precautions normally followed in the analytical laboratory may be followed once bacteria are dried by successive treatment with ethanol, acetone, and petroleum ether.
2. *H. influenzae* strains 1003 and R2486*lpsA* referred to in this chapter are from the culture collection of Professor E. R. Moxon (Oxford University, UK). NTHi 1003 was a middle ear isolate obtained as part of the Finnish Otitis Media Cohort Study (36). Wild-type R2486 was a middle ear isolate from an otitis media patient (37) from which the *lpsA* mutant was constructed (24).
3. Both mild acid hydrolysis and anhydrous hydrazine provide effective methods for O-deacylation of LPS under mild conditions and analytical applications have been described (32). Treatment of LPS samples with anhydrous hydrazine at ambient temperature leads to the removal of O-acyl groups from the lipid A and core regions of the molecule. O-Deacylated LPS (LPS-OH) samples have been used for rapid glycoform profiling by ESI-MS and CE-ESI-MS in the negative ion mode and is well established for probing the heterogeneity of LPS from *H. influenzae* (34, 38). We have used this approach for initial characterization of LPS glycoforms from ex vivo chinchilla middle ear samples (*see* **Note 9**) (33, 39). While analysis of LPS-OH samples provides valuable information on acid-labile substituents such as sialic acid (N-acetylneuraminic acid) (*see* **Note 4**) only limited information is available on the patterns of isomeric glycoforms due to limited fragmentation in the negative ion mode (40).
4. Partial acid hydrolysis of short-chain LPS in dilute acetic acid affords a soluble OS fraction and the insoluble lipid A which is removed by centrifugation. This takes advantage of the fact that the 3-deoxy-2-keto-sugars are more acid labile than aldoses. Borane-N-methylmorpholine complex is added to the hydrolysis mixture to reduce the Kdo carbonyl group that is liberated during hydrolysis (41). This procedure also results in cleavage of acid-labile sialic acid residues. Structures containing terminal sialic acid residues have been identified in many *H. influenzae* strains (4) and are typically characterized by CE-ESI-MS analysis of O-deacylated LPS samples (21, 23) or by NMR when sufficient quantities of sample are available (14).

5. Phosphate substituents (*P*, *P*Etn, and *P*Cho) are removed by treatment with cold aqueous HF, relatively mild conditions which do not normally lead to hydrolysis of glycosidic linkages. Hydrofluoric acid can cause severe burns and is reactive with silicon-based glass. Latex gloves should be worn and all manipulations should be carried out in a fume hood. Standard polypropylene Eppendorf vials should be used for the HF treatment step.
6. It is important that all procedures are conducted under anhydrous conditions by carrying out each of the steps under a flow of dry nitrogen.
7. HPLC–ESI-MS is conducted using the standard interface to the LCQ quadrupole ion trap mass spectrometer. Multiple-stage tandem MS can be carried out on this spectrometer effectively to MS^4 making it possible to unambiguously identify isomeric distributions of LPS glycoforms (18) (*see* **Note 8**). It is important to degas eluant solvents A and B during preparation and to sparge with helium gas during runs to prevent air bubbles. HPLC provides on-line sample clean-up and partial separation of glycoforms for subsequent MS analysis. The addition of NaOAc to the HPLC eluants promotes formation of Na^+ adducts which increases MS-sensitivity (30).
8. Sequence information is readily obtained from permethylated LPS OS samples since methyl tagging permits the distinction between fragment ions generated by cleavage of a single glycosidic bond and inner fragments resulting from the rupture of two glycosidic linkages. By contrast in underivatized samples, the fragment ions arising from these cleavage pathways are isobaric rendering them indistinguishable from each other. **Figure 6.2** shows the application to permethylated OS derived from non-typeable *H. influenzae* strain 1003 (16). In this spectrum the most prominent signal at *m/z* 1468.1 represents a glycoform having a single hexose attached to both the proximal and the terminal heptose (**Fig. 6.1**; R_1=R_3=Hex; R_2=H) (33). A low-abundance ion corresponding to Hex4 glycoforms is evident in the full scan spectrum at *m/z* 1875.3 from which it is possible to identify three isomeric forms in the MS^2 spectrum (**Fig. 6.2B**). Thus, fragment ions at *m/z* 1002 and 754 were indicative of consecutive losses from a t-Hex–Hex–Hex–Hep–Hep unit in isomeric glycoform **1** while the isomer **2** was defined by ions at *m/z* 1206 and 958 due to consecutive losses from a t-Hex-Hex-Hep-Hep unit. The third Hex4 isomeric glycoform (3) was characterized by ions at *m/z* 1410 and 1162 arising from consecutive loss from t-Hex-Hep-Hep (33). In the MS spectrum of

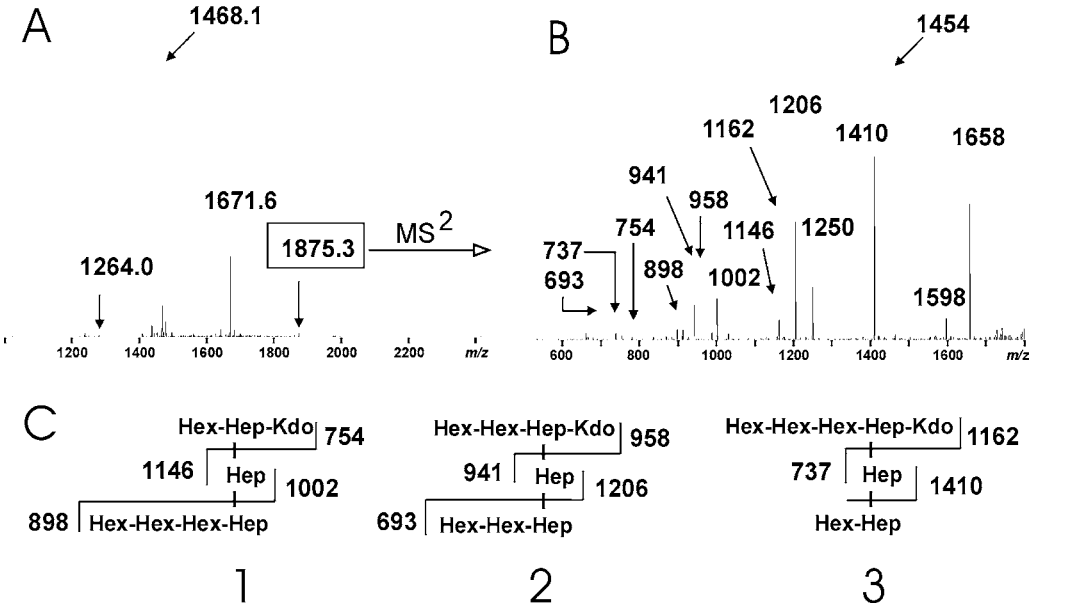


Fig. 6.2. HPLC–ESI-MS of permethylated OS from NTHi 1003 LPS (A) and the product ion (MS^2) spectrum of the ion at *m/z* 1875.3 (B). The structure of the isomeric Hex4 glycoforms and fragmentation patterns identified in B are shown (C).

O-deacylated LPS (underivatized) the Hex4 glycoforms are not detectable (16).

In another example, identification of a novel branching OS extension from the proximal heptose of the LPS inner-core unit (*see* **Fig. 6.1**) was identified in NTHi R2846 (24) by subsequent multiple-stage MS experiments. This strain of NTHi was of particular interest since it was obtained as a clinical isolate from the middle ear of an otitis media patient (37) for which the complete genome sequence was recently made available. The *lpsA* mutant (i.e., referred to as R2846*lpsA*) was constructed (24) to facilitate structural characterization of complex OS extensions from the proximal heptose (i.e., at R_1 in **Fig. 6.1**) since this gene is required for Hex addition from the terminal Hep (i.e., at R_3 in **Fig. 6.1**). In the major glycoform, depicted as Hex_2Hep in **Fig. 6.3A** (*m/z* 1716), the proximal heptose is substituted by the Hex-Hep-Hex trisaccharide extension while the central and distal heptose units are unsubstituted (**Fig. 6.1**; $R_2{=}R_3{=}H$). The minor ion at *m/z* 2164.6 in the full scan spectrum corresponds to an OS extension by $HexNAc_1Hex_3Hep_1$. Even though the abundance of this glycoform was only just above background it was possible to unambiguously determine its sequence by multiple-stage tandem MS experiments. Thus, MS^2 on *m/z* 2164.6 gave a fragment ion at *m/z* 1656 corresponding to loss of the HepII–HepIII unit of the inner-core region (*see* **Fig. 6.1**;

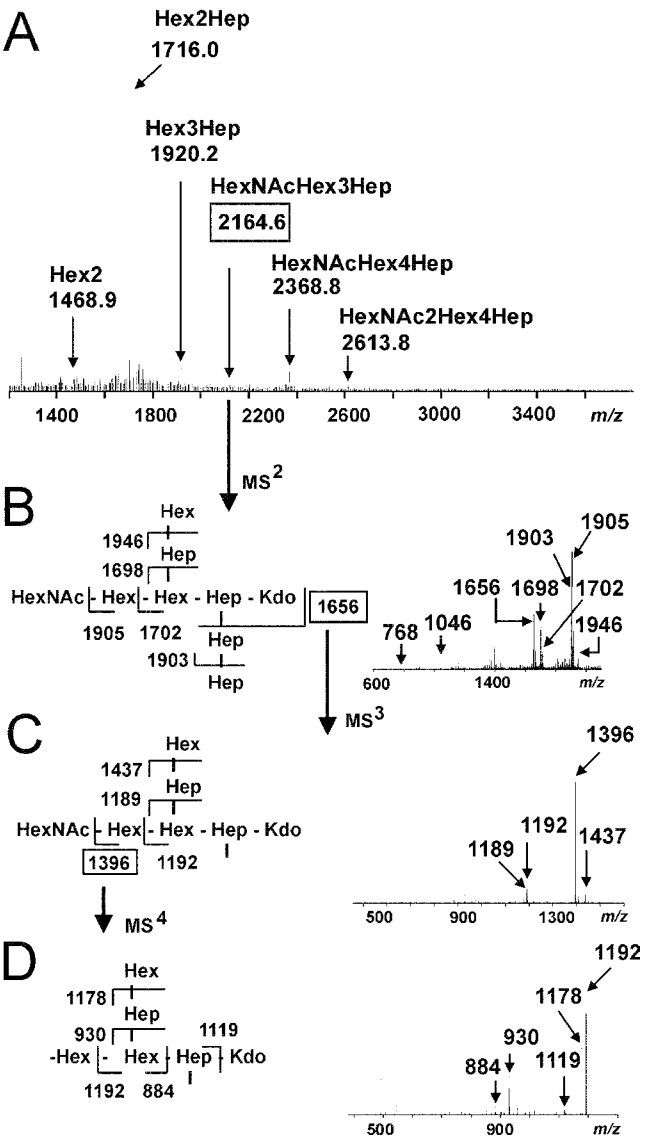


Fig. 6.3. HPLC-ESI-MS^n analysis on permethylated OS glycoforms derived from LPS of NTHi R2846*lpsA*. Full scan spectrum in which proposed glycoforms (all contain additional Hep_3 *An*Kdo-ol) are indicated (A). Product ion spectrum (MS^2) of $[M+Na]^+$ ion at *m/z* 2164.6 showing structure and key fragmentations (B). MS^3 of ion at *m/z* 1656 showing structure and key fragmentations (C). MS^4 of ion at *m/z* 1396 showing structure and key fragmentations (D).

Fig. 6.3B). MS^3 on this ion resulted in the loss t-HexNAc giving rise to the ion at *m/z* 1396 (**Fig. 6.3C**) which in the MS^4 spectrum gave fragments at *m/z* 1178, 930, and 1192 corresponding to loss of t-Hex, t-Hex-Hep, and of the substituted Hex, respectively, confirming the sequence of the branching OS substituent (**Fig. 6.3D**).

9. We have employed CE-ESI-MS to determine glycoform profiles on LPS-OH for bacteria from direct middle ear fluid (MEF) samples during the course of experimental otitis media in the chinchilla model of infection (33, 39). With this information in hand, permethylated OS samples prepared from LPS-OH (*see* **Section 3.3.1**) of NTHi 1003 were subjected to HPLC–ESI-MS in the positive ion mode with selective ion monitoring/selective reaction monitoring (SIM/SRM) to increase sensitivity (33). The enhanced sensitivity afforded by analysis of permethylated samples under these conditions permitted the detection

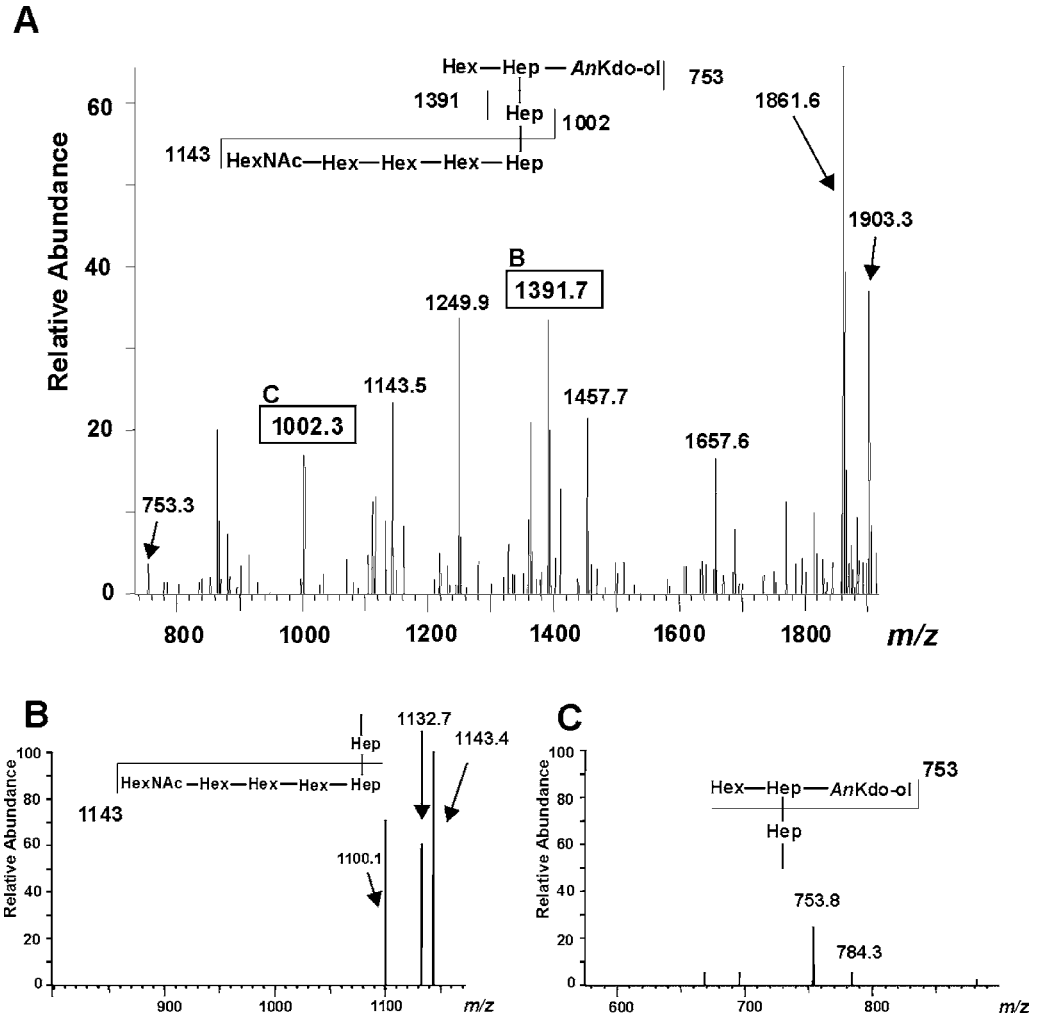


Fig. 6.4. HPLC–ESI-MS^2 spectrum of the Hex4HexNAc glycoform (*m/z* 2121.0) identified in the permethylated OS from NTHi 1003 LPS obtained from the middle ear fluid of a chinchilla 7 days after infection and singly passaged on solid media (33) (A). MS^3 spectrum of the fragment ions at *m/z* 1391.7 (B) and 1002.3 (C). Fragmentation patterns are shown in structural insets.

of minor glycoforms not observed by analysis of the underivatized LPS-OH samples. Moreover, when the MEF was passaged on solid media, ions from less prominent glycoforms were of sufficient abundance to permit further tandem MS analysis beyond MS^2. Thus, a Hex4HexNAc glycoform was detected in the MEF taken from three of ten animals 7 days after infection as evidenced from an ion at m/z 2121. This glycoform has not been observed in laboratory bacterial growth samples. As illustrated in the product ion spectrum (MS^2) of this ion (**Fig. 6.4A**), a fragment ion at m/z 1391.7 was indicative of the loss of a $HexNAcHex_3Hep_2$ unit. The MS^3 spectrum of this ion provided confirmation that the terminal heptose (i.e., HepIII in **Fig. 6.1**) is the point of substitution (**Fig. 6.4B**). MS^3 of the fragment at m/z 1002 (**Fig. 6.4C**) indicated that the proximal heptose is substituted by a single hexose in this glycoform completing the characterization of the glycoform.

Acknowledgments

We thank Dr. Derek Hood and Professor E. Richard Moxon for the contributions they have made leading to the development of this strategy. Brigitte Twelkmeyer, Varvara Vitiazeva, and Mikael Engskog are acknowledged for assistance in preparing this manuscript.

References

1. Campagnari, A.A., Gupta, M.R., Dudas, K.C., Murphy, T.F., and Apicella, M.A. (1987) Antigenic diversity of lipooligosaccharides of nontypeable *Haemophilus influenzae*. *Infect. Immun.* **55**, 882–887.
2. Anderson, P., Johnston, R.B., Jr., and Smith, D.H. (1972) Human serum activities against *Hemophilus influenzae*, type b. *J. Clin. Invest.* **51**, 31–38.
3. Zwahlen, A., Rubin, L.G., and Moxon, E.R. (1986) Contribution of lipopolysaccharide to pathogenicity of *Haemophilus influenzae*: comparative virulence of genetically-related strains in rats. *Microb. Pathog.* **1**, 465–473.
4. Schweda, E.K., Richards, J.C., Hood, D.W., and Moxon, E.R. (2007) Expression and structural diversity of the lipopolysaccharide of *Haemophilus influenzae*: implication in virulence. *Int. J. Med. Microbiol.* **297**, 297–306.
5. Helander, I.M., Lindner, B., Brade, H., Altmann, K., Lindberg, A.A., Rietschel, E.T., and Zähringer, U. (1988) Chemical structure of the lipopolysaccharide of *Haemophilus influenzae* strain I-69 Rd-/b+. Description of a novel deep-rough chemotype. *Eur. J. Biochem.* **177**, 483–492.
6. Mikhail, I., Yildirim, H.H., Lindahl, E.C., and Schweda, E.K. (2005) Structural characterization of lipid A from nontypeable and type f *Haemophilus influenzae*: variability of fatty acid substitution. *Anal. Biochem.* **340**, 303–316.
7. Phillips, N.J., Apicella, M.A., Griffiss, J.M., and Gibson, B.W. (1992) Structural characterization of the cell surface lipooligosaccharides from a nontypeable strain of *Haemophilus influenzae*. *Biochemistry* **31**, 4515–4526.
8. Phillips, N.J., Apicella, M.A., Griffiss, J.M., and Gibson, B.W. (1993) Structural studies of the lipooligosaccharides from *Haemophilus influenzae* type b strain A2. *Biochemistry* **32**, 2003–2012.

9. Schweda, E.K., Hegedus, O.E., Borrelli, S., Lindberg, A.A., Weiser, J.N., Maskell, D.J., and Moxon, E.R. (1993) Structural studies of the saccharide part of the cell envelope lipopolysaccharide from *Haemophilus influenzae* strain AH1-3 (lic3+). *Carbohydr. Res.* **246**, 319–330.
10. Schweda, E.K., Jansson, P.E., Moxon, E.R., and Lindberg, A.A. (1995) Structural studies of the saccharide part of the cell envelope lipooligosaccharide from *Haemophilus influenzae* strain galEgalK. *Carbohydr. Res.* **272**, 213–224.
11. Masoud, H., Moxon, E.R., Martin, A., Krajcarski, D., and Richards, J.C. (1997) Structure of the variable and conserved lipopolysaccharide oligosaccharide epitopes expressed by *Haemophilus influenzae* serotype b strain Eagan. *Biochemistry* **36**, 2091–2103.
12. Risberg, A., Masoud, H., Martin, A., Richards, J.C., Moxon, E.R., and Schweda, E.K. (1999) Structural analysis of the lipopolysaccharide oligosaccharide epitopes expressed by a capsule-deficient strain of *Haemophilus influenzae* Rd. *Eur. J. Biochem.* **261**, 171–180.
13. Schweda, E.K., Brisson, J.-R., Alvelius, G., Martin, A., Weiser, J.N., Hood, D.W., Moxon, E.R., and Richards, J.C. (2000) Characterization of the phosphocholine-substituted oligosaccharide in lipopolysaccharides of type b *Haemophilus influenzae*. *Eur. J. Biochem.* **267**, 3902–3913.
14. Månsson, M., Bauer, S.H., Hood, D.W., Richards, J.C., Moxon, E.R., and Schweda, E.K. (2001) A new structural type for *Haemophilus influenzae* lipopolysaccharide. Structural analysis of the lipopolysaccharide from nontypeable *Haemophilus influenzae* strain 486. *Eur. J. Biochem.* **268**, 2148–2159.
15. Cox, A.D., Masoud, H., Thibault, P., Brisson, J.-R., van der Zwan, M., Perry, M.B., and Richards, J.C. (2001) Structural analysis of the lipopolysaccharide from the nontypeable *Haemophilus influenzae* strain SB 33. *Eur. J. Biochem.* **268**, 5278–5286.
16. Månsson, M., Hood, D.W., Li, J., Richards, J.C., Moxon, E.R., and Schweda, E.K. (2002) Structural analysis of the lipopolysaccharide from nontypeable *Haemophilus influenzae* strain 1003. *Eur. J. Biochem.* **269**, 808–818.
17. Månsson, M., Hood, D.W., Moxon, E.R., and Schweda, E.K. (2003) Structural characterization of a novel branching pattern in the lipopolysaccharide from nontypeable *Haemophilus influenzae*. *Eur. J. Biochem.* **270**, 2979–2991.
18. Månsson, M., Hood, D.W., Moxon, E.R., and Schweda, E.K. (2003) Structural diversity in lipopolysaccharide expression in nontypeable *Haemophilus influenzae*. Identification of L-*glycero*-D-*manno*-heptose in the outer-core region in three clinical isolates. *Eur. J. Biochem.* **270**, 610–624.
19. Yildirim, H.H., Hood, D.W., Moxon, E.R., and Schweda, E.K. (2003) Structural analysis of lipopolysaccharides from *Haemophilus influenzae* serotype f. Structural diversity observed in three strains. *Eur. J. Biochem.* **270**, 3153–3167.
20. Masoud, H., Martin, A., Thibault, P., Moxon, E.R., and Richards, J.C. (2003) Structure of extended lipopolysaccharide glycoforms containing two globotriose units in *Haemophilus influenzae* serotype b strain RM7004. *Biochemistry* **42**, 4463–4475.
21. Yildirim, H.H., Li, J., Richards, J.C., Hood, D.W., Moxon, E.R., and Schweda, E.K. (2005) An alternate pattern for globoside oligosaccharide expression in *Haemophilus influenzae* lipopolysaccharide: structural diversity in nontypeable strain 1124. *Biochemistry* **44**, 5207–5224.
22. Yildirim, H.H., Li, J., Richards, J.C., Hood, D.W., Moxon, E.R., and Schweda, E.K. (2005) Complex O-acetylation in non-typeable *Haemophilus influenzae* lipopolysaccharide: evidence for a novel site of O-acetylation. *Carbohydr. Res.* **340**, 2598–2611.
23. Lundström, S.L., Twelkmeyer, B., Sagemark, M.K., Li, J., Richards, J.C., Hood, D.W., Moxon, E.R., and Schweda, E.K. (2007) Novel globoside-like oligosaccharide expression patterns in nontypeable *Haemophilus influenzae* lipopolysaccharide. *FEBS J.* **274**, 4886–4903.
24. Lundström, S.L., Li, J., Deadman, M.E., Hood, D.W., Moxon, E.R., and Schweda, E.K. (2008) Structural analysis of the lipopolysaccharide from nontypeable *Haemophilus influenzae* strain R2846. *Biochemistry* **47**, 6025–6038.
25. Schilling, B., McLendon, M.K., Phillips, N.J., Apicella, M.A., and Gibson, B.W. (2007) Characterization of lipid A acylation patterns in *Francisella tularensis*, *Francisella novicida*, and *Francisella philomiragia* using multiple-stage mass spectrometry and matrix-assisted laser desorption/ionization on an intermediate vacuum source linear ion trap. *Anal. Chem.* **79**, 1034–1042.
26. Vinogradov, E., Lindner, B., Seltmann, G., Radziejewska-Lebrecht, J., and Holst, O. (2006) Lipopolysaccharides from *Serratia marcescens* possess one or two

4-amino-4-deoxy-L-arabinopyranose 1-phosphate residues in the lipid A and D-*glycero*-D-*talo*-oct-2-ulopyranosonic acid in the inner core region. *Chemistry* **12**, 6692–6700.
27. Landerholm, M.K., Li, J., Richards, J.C., Hood, D.W., Moxon, E.R., and Schweda, E.K. (2004) Characterization of novel structural features in the lipopolysaccharide of nondisease associated nontypeable *Haemophilus influenzae*. *Eur. J. Biochem.* **271**, 941–953.
28. Schweda, E.K., Landerholm, M.K., Li, J., Moxon, E.R., and Richards, J.C. (2003) Structural profiling of lipopolysaccharide glycoforms expressed by non-typeable *Haemophilus influenzae*: phenotypic similarities between NTHi strain 162 and the genome strain Rd. *Carbohydr. Res.* **338**, 2731–2744.
29. Gulin, S., Pupo, E., Schweda, E.K., and Hardy, E. (2003) Linking mass spectrometry and slab-polyacrylamide gel electrophoresis by passive elution of lipopolysaccharides from reverse-stained gels: analysis of gel-purified lipopolysaccharides from *Haemophilus influenzae* strain Rd. *Anal. Chem.* **75**, 4918–4924.
30. Reinhold, V.N., Reinhold, B.B., and Costello, C.E. (1995) Carbohydrate molecular weight profiling, sequence, linkage, and branching data: ES-MS and CID. *Anal. Chem.* **67**, 1772–1784.
31. Hood, D.W., Cox, A.D., Wakarchuk, W.W., Schur, M., Schweda, E.K., Walsh, S.L., Deadman, M.E., Martin, A., Moxon, E.R., and Richards, J.C. (2001) Genetic basis for expression of the major globotetrose-containing lipopolysaccharide from *H. influenzae* strain Rd (RM118). *Glycobiology* **11**, 957–967.
32. Schweda, E.K. and Richards, J.C. (2003) Structural profiling of short-chain lipopolysaccharides from *Haemophilus influenzae*. *Methods Mol. Med.* **71**, 161–183.
33. Lundström, S.L., Li, J., Månsson, M., Figueira, M., Leroy, M., Goldstein, R., Hood, D.W., Moxon, E.R., Richards, J.C., and Schweda, E.K. (2008) Application of capillary electrophoresis mass spectrometry and liquid chromatography multiple-step tandem electrospray mass spectrometry to profile glycoform expression during *Haemophilus influenzae* pathogenesis in the chinchilla model of experimental otitis media. *Infect. Immun.* **76**, 3255–3267.
34. Li, J. and Richards, J.C. (2007) Application of capillary electrophoresis mass spectrometry to the characterization of bacterial lipopolysaccharides. *Mass Spectrom Rev.* **26**, 35–50.
35. Li, J., Bauer, S.H., Månsson, M., Moxon, E.R., Richards, J.C., and Schweda, E.K. (2001) Glycine is a common substituent of the inner core in *Haemophilus influenzae* lipopolysaccharide. *Glycobiology* **11**, 1009–1015.
36. Cody, A.J., Field, D., Feil, E.J., Stringer, S., Deadman, M.E., Tsolaki, A.G., Gratz, B., Bouchet, V., Goldstein, R., Hood, D.W., and Moxon, E.R. (2003) High rates of recombination in otitis media isolates of non-typeable *Haemophilus influenzae*. *Infect. Genet. Evol.* **3**, 57–66.
37. Barenkamp, S.J. and Leininger, E. (1992) Cloning, expression, and DNA sequence analysis of genes encoding nontypeable *Haemophilus influenzae* high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of *Bordetella pertussis*. *Infect. Immun.* **60**, 1302–1313.
38. Melaugh, W., Engstrom, J.J., Auriola, S., Phillips, N.J., and Gibson, B.W. (1996) *Mass Spectrometry in the Biological Sciences*. Humana Press, Totowa, N.S.
39. Bouchet, V., Hood, D.W., Li, J., Brisson, J.-R., Randle, G.A., Martin, A., Li, Z., Goldstein, R., Schweda, E.K., Pelton, S.I., Richards, J.C., and Moxon, E.R. (2003) Host-derived sialic acid is incorporated into *Haemophilus influenzae* lipopolysaccharide and is a major virulence factor in experimental otitis media. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8898–8903.
40. Schweda, E.K., Twelkmeyer, B., and Li, J. (2008) Profiling structural elements of short-chain lipopolysaccharide of non-typeable *Haemophilus influenzae*. *Innate Immun.* **14**, 199–211.
41. Risberg, A., Schweda, E.K., and Jansson, P.E. (1997) Structural studies of the cell-envelope oligosaccharide from the lipopolysaccharide of *Haemophilus influenzae* strain RM.118-28. *Eur. J. Biochem.* **243**, 701–707.

Chapter 7

In Vitro Reconstitution of *Escherichia coli* 086 O Antigen Repeating Unit

Weiqing Han, Lei Li, Nicholas Pettit, Wen Yi, Robert Woodward, Xianwei Liu, Wanyi Guan, Veer Bhatt, Jing Katherine Song, and Peng George Wang

Abstract

Polysaccharides constitute a major component of the bacterial cell surface. They play critical roles in the interactions between bacteria and the host environments, and consequently contribute to the virulence of pathogens. The lipopolysaccharide (LPS) found on the surface of gram-negative bacteria consists of three parts: lipid A, a core oligosaccharide, and the O antigen. The O antigen is the outermost part of LPS and contains multiple oligosaccharide repeating units. Biosynthesis of the O-repeating unit is the first committed step in LPS biosynthesis. We sequenced and characterized the O-antigen biosynthetic gene cluster of *Escherichia coli serotype* O86. Four glycosyltransferases encoded by the genes within the cluster were cloned and overexpressed. In vitro reconstitution of the O-repeating unit of *E. coli* 086 was achieved via using these enzymes.

Key words: *Escherichia coli*, lipopolysaccharide, O antigen, O-repeating unit, glycosyltransferase.

1. Introduction

There are generally large amounts of polysaccharides occurring on the surface of bacteria which serve a variety of purposes. In gram-negative bacteria, LPS constitutes the major component of the outer membrane and contributes significantly to the cell structural integrity and pathogenicity. LPS consists of three structural parts: (i) the lipid A, glucosamine-based phospholipids; (ii) the core oligosaccharide; and (iii) the O antigen that is composed of multiple copies of oligosaccharide repeating units (O-RU). The

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_7,

lipid A anchors LPS in the outer membrane and in conjunction with the core oligosaccharide both highly conserved structurally and genetically (1). However, the O antigen can observe considerable variation due to the possible variation of sugar composition, the arrangement of the sugars, the linkages between the sugars within the O-RU, and as well as the linkages between different O-RUs.

In the past decade much effort has been put in to elucidate the biosynthetic pathway of O antigens. Three major biosynthetic mechanisms, i.e., *wzy*-dependent, ABC-transporter-dependent, and synthase-dependent pathways, have been proposed for the assembly of O antigens (2). Among these, the *wzy*-dependent pathway is the most widely used in nature. In the *wzy*-dependent pathway, the assembly of the O-unit starts with the addition of a sugar phosphate onto undecaprenyl phosphate (Und-P) to form an Und-PP-linked sugar. This initiation reaction is catalyzed by an integral membrane protein WecA or WbaP, depending on the identity of the first sugar of the O-unit structure (either N-acetylhexosamine or hexose) (3, 4). Following the initiation reaction, subsequent steps involve the sequential addition of different sugar residues onto an Und-PP-sugar intermediate, catalyzed by specific glycosyltransferases. These glycosyltransferases are predicted to be soluble or peripheral membrane proteins. Upon completion of the RU-PP-Und, they are translocated to the periplasmic side of the inner membrane by the flippase Wzx (5, 6). In the periplasm, the RU-PP-Und are polymerized into a sugar polymer by the integral membrane protein Wzy. The polymerization reaction involves the transfer of the nascent polymer from its Und-PP carrier to the non-reducing end of the new RU-PP-Und (7). The final component of the *wzy*-dependent pathway is the Wzz protein, which works as a chain length regulator to generate strain-specific polysaccharide chain lengths (8). The PS-PP-Und is then ligated to lipid A by WaaL, an integral membrane protein, and then the complete LPS is finally translocated to the outer membrane via a protein complex Imp/RlpB (9).

Although increasing efforts involving gene cluster sequencing, putative enzyme annotation, and gene knockout/complementation approaches have provided remarkable insight into the mechanistic details of this pathway, the current model of *wzy*-dependent polysaccharide biosynthesis has never been biochemically reconstituted in a cell-free in vitro system. One of the major difficulties that hampered this study is that the complete RU-PP-Und is needed as substrate to assay Wzy protein in vitro and is as such very difficult to obtain.

E. coli O86 is interesting since it exhibits strong human blood group B activity. Based on the chemical and serological investigations, Springer et al. (10) and Kochibe et al. (11) showed that the strong blood group activity of *E. coli* O86 came from

LPS that contained the blood group B trisaccharide partial structure. The histo-blood group ABO system is defined by ABO blood group carbohydrate antigens, which play important roles in blood transfusion, cell differentiation, and oncogenesis (12). The blood group A and B antigens are formed by the transfer of two different sugar residues (GalNAc and Gal) to the common H-disaccharide epitope by human blood group transferases GTA and GTB, respectively. Chemical synthesis methodology developed by Lemieux set a milestone in the in vitro synthesis of blood group antigens (13, 14). On the other hand, the enzymatic synthesis using recombinant GTA and GTB has proven less laborious and more regioselective and stereoselective (15, 16). Microbial glycosyltransferases have recently become an attractive alternative to the mammalian counterparts, since they are easy to express in active form without complicated gene manipulation and can catalyze similar glycosylation reactions. The close association between *E. coli* O86 O antigen and human blood group B antigen prompts us to investigate the glycosyltransferases involving the assembly of O86 O antigens.

E. coli O86 O antigen is derived from the O-repeating unit structure consisting of -4-Fucα (1–2)[Galα (1–3)]Galβ (1–3)GalNAcα (1–3)GalNAcα1- (**Fig. 7.1**). The O-antigen biosynthetic gene cluster was sequenced and the sequence analysis indicates that the O-antigen biosynthesis of *E. coli* O86 is *wzy*-dependent since the potential *wzx* and *wzy* genes were found within the gene cluster. Among the 14 *ORF*s deduced from the sequence, four (*wbnH*, *wbnI*, *wbnJ*, and *wbnK*) encode glycosyltransferases involved in the O-RU assembly (**Fig. 7.2**). In present study, specific functions of WbnI, WbnJ, WbnK, and WbnH were biochemically characterized and they were employed sequentially to in vitro synthesis of the RU-PP-lipid. In this study, an undecaprenyl mimic, *cis*-pentaprenyl was used instead of undecaprenyl.

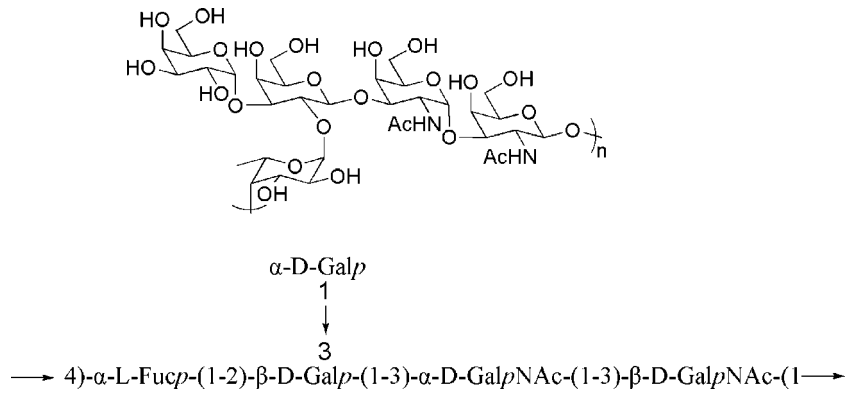


Fig. 7.1. O-repeating unit structure of *E. coli* O86:H2.

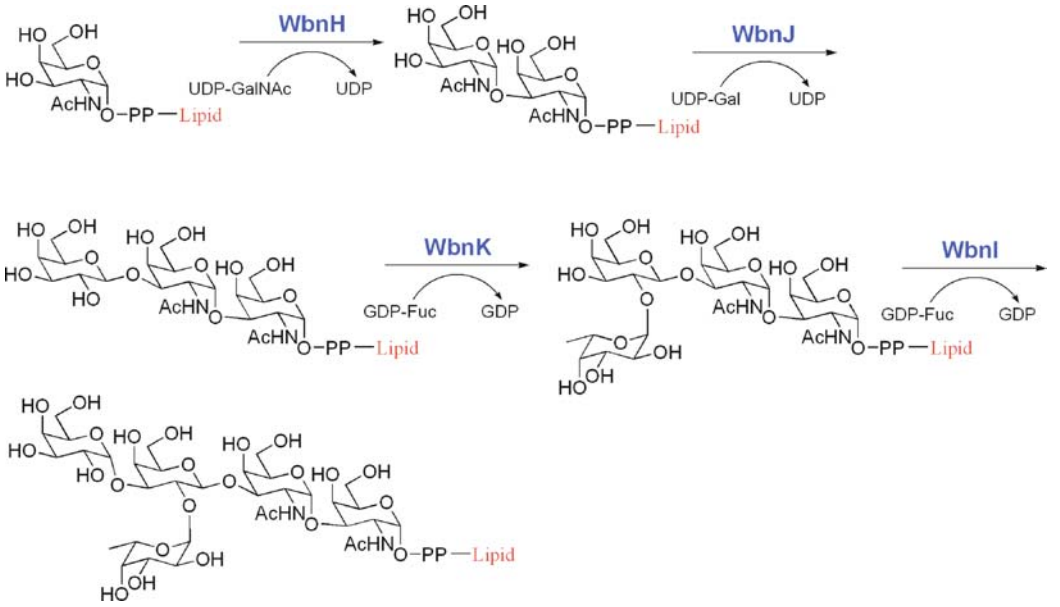


Fig. 7.2. Proposed model of RU-PP-Und biosynthesis in *E. coli* O86.

2. Materials

All Chemicals and solvents were from Sigma-Aldrich unless otherwise indicated.

2.1. Sequencing and Analysis of E. coli *O86:H2 O Antigen Gene Cluster*

1. *E. coli* O86:H2 was obtained from Department of Health and Human Service, Food and Drug Administration.
2. *E. coli* DH5α competent cell [*lacZDM15 hsdR recA*] (Gibco-BRL LifeTechnology).
3. Luria–Bertani medium (LB medium) for the cultivation of all *E. coli* strains used in this study: 1% (w/v) tryptone, 0.5% (w/v) yeast extract, and 1% NaCl. Adjust pH to 7.0. After autoclaving, kept at room temperature. LB agar: LB medium supplemented with 1.5% agar.
4. ELU-QUIK DNA Purification kit (Schleicher & Schuell Bioscience).
5. LA *Taq* DNA polymerase (Takara Bio, Inc.) for amplification of *E. coli* O86:H2 O antigen gene cluster.
6. Wizard PCR Preps DNA Purification System (Promega) for the purification of long-range PCR products.
7. Nebulizers (Invitrogen) for fragmentation of long-range PCR products.

8. T4 DNA polymerase (Novagen).
9. Single deoxyribosyladenine tailing kit (Novagen).
10. Chloroform–isoamyl alcohol (24:1), store at room temperature.
11. pGEM-T-easy cloning vector (Novagen).
12. T4 DNA ligase (Promega).
13. 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) stock solution: dissolve the X-gal in N, N'-dimethylformamide to make a 20 mg/mL solution. Use a glass or polypropylene tube. Wrap the tube containing the solution in aluminum foil to prevent damage by light and store at –20°C.
14. Isopropyl-β-D-1-thiogalactopyranoside (IPTG) stock solution: dissolve the IPTG in H_2O to make a 200-mg/mL solution. Sterilize by passing it through a 0.22-μm disposable filter. Store at –20°C.
15. Ampicillin stock solution: prepare a stock of 100 mg/mL in water. Sterilize by using a 0.22-μm filter. Store at –20°C.
16. *Taq* DNA polymerase (New England Biolabs).
17. QIAprep 96 Turbo Miniprep kit (Qiagen) for plasmid preparation.

2.2. Cloning, Expression, Purification, and Activity Assay of Glycosyltransferases

2.2.1. Cloning, Expression, and Purification of Glycosyltransferase

1. The Herculase DNA polymerase for protein cloning was purchased from Stratagene, USA.
2. All restriction enzymes were purchased from New England Biolabs, Inc.
3. The expression vectors pET15b and pGEX-4T-1 were purchased from Novagen and GE Healthcare, respectively.
4. *E. coli* competent cell BL21 (DE3) [*F⁻ ompT hsdSB* er^-_B m^-_B *T gal dcm* (DE3)] (Novagen).
5. Binding buffer for His-tag fusion proteins: 50 mM Tris–HCl, pH 8.0, 0.2 M NaCl, 5 mM imidazole, and 1 mM PMSF for purification of WbnH, WbnJ, and WbnI. Store at 4°C.
6. Binding buffer for GST-tag fusion proteins: 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.3 for purification of WbnK. Store at 4°C.
7. Sonicator Branson Sonifier 450 (VWR Scientific).
8. Five milliliters of nickel-nitrilotriacetic acid (Ni^{2+}-NTA) chelating column (Amersham); 1 mL GST-Bind column (Amersham).

9. Wash buffer for His-tag fusion proteins: 50 mM Tris–HCl, pH 8.0, 0.2 M NaCl, and 80 mM imidazole for purification of WbnH, WbnJ, and WbnI. Store at 4°C.
10. Wash buffer for GST-tag fusion proteins: 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.3 for purification of WbnK. Store at 4°C.
11. Elution buffer for His-tag fusion proteins: 50 mM Tris–HCl, pH 8.0, 0.2 M NaCl, and 250 mM imidazole for purification of WbnH, WbnJ, and WbnI. Store at 4°C.
12. Elution buffer for GST-tag fusion proteins: 50 mM Tris–HCl, 10 mM reduced Glutathione, pH 8.0 for purification of WbnK. Store at 4°C.
13. Bradford Protein Assay kit (Bio-rad, USA).

2.2.2. Enzymatic Assay for Glycosyltransferase Activity

1. Donor substrates: UDP-GalNAc for WbnH; radioactively labeled UDP-D-[6-^{3}H]Galactose (10,000 cpm) for WbnJ and WbnI; and radioactively labeled GDP-L-[U-^{14}C]fucose (7000 cpm) for WbnK.
2. Acceptor substrates: GalNAcα-PP-O-$(CH_2)_{11}$-Oph and GalNAcα-OMe that are chemically synthesized as acceptors for WbnH and WbnJ, respectively; Galβ1,3-GalNAcα-OMe and Fucα1,2-Galβ1,3-GalNAcα-OMe that are enzymatically synthesized as below described as acceptors for WbnK and WbnI, respectively.
3. Dowex 1×8-200 chloride anion exchange resin.
4. ScintiVerse BD liquid.
5. LS-3801 scintillation counter (Beckmann).
6. 0.5 M MES (pH 7.0).
7. 100 mM $MnCl_2$.
8. 200 mM Tris–HCl (pH7.0).
9. 10 mM ATP.
10. 0.1 M EDTA.
11. 10 mM dithiothreitol (DTT).
12. Thin-layer chromatography solution: i-PrOH/H_2O/NH_4OH=8:2:2 (v/v/v).
13. Baker Si250F silica gel TLC plates.
14. Staining solution: anisaldehyde/MeOH/H_2SO_4=1:15:2 (v/v/v).
15. Gel filtration chromatography Bio-Gel P2 (Bio-Rad, USA).

2.3. In Vitro Reconstitution of Und-pp-Linked O-Repeating Unit of E. coli O86

1. 200 mM Tris–HCl (pH 7.0).
2. 50 mM $MnCl_2$.
3. Triton X-100.
4. 50 mM ammonium formate (pH 4.5).
5. Acetonitrile containing 5 mM ammonium formate.
6. Initial sugar acceptor: 5 mg of GalNAc-PP-Und (solubilized in DMSO) that is chemically synthesized as below described in Methods section.
7. Sugar donors: UDP-GalNAc, UDP-Gal, and GDP-Fuc.

3. Methods

3.1. Sequencing and Analysis of E. coli O86:H2 O Antigen Gene Cluster

1. *E. coli* O86:H2 was cultured overnight in 5 mL Luria Bertani medium. Chromosomal DNA was isolated via using ELU-QUIK DNA Purification kit from Schleicher & Schuell Bioscience, Inc.
2. The *E. coli* O-antigen biosynthetic gene cluster is normally located between *galF* and *gnd* that are two highly conserved house-keeping genes. A pair of primers: F (5′-CTCTCTGAATACTCCGTCATC) and R (5′-ACCTGCTTTCACCATTAACAGGAT) were designed based on the two genes. Long-range PCR was carried out by using LA *Taq* DNA polymerase from Takara Bio, Inc. The PCR amplification mix is as following: 5 μL of PCR buffer (10×), 2 μL of dNTP (each 10 mM), 1 μL of primer F (5 μM), 1 μL of primer R (5 μM), 0.5 μL of chromosomal DNA and 0.5 μL of DNA polymerase, and distilled H_2O added to a final volume of 50 μL. The PCR cycles were as follows: 94°C for 2 min; 30 cycles of 94°C for 10 s, 60°C for 30 s, and 68°C for 15 min; and finally 68°C for 7 min.
3. A total of six long-range PCR products were pooled to limit the PCR errors. All the PCR products were collected into an Eppendorf tube and purified using the Wizard PCR Preps DNA Purification kit from Promega. The DNA was resuspended in 50 μL of distilled water.
4. The PCR products were sheared by using nebulizers (Invitrogen) according to the manufacturer's instructions (*see* **Note 1**). The DNA fragments were then purified using a Wizard PCR DNA preparation kit (Promega), resuspended in 30 μL of water.

5. The 30 μL of nebulizer-sheared DNA fragment was flushed by using T4 DNA polymerase to generate blunt ended fragments in a 50-μL flushing reaction mixture according to the manufacturer's instructions.
6. The flushed fragment was subjected to single deoxyribosyladenine tailing with a single deoxyribosyladenine tailing kit from Novagen in a 160-μL reaction mixture according to the manufacturer's instructions.
7. The reaction product was then extracted with chloroform–isoamyl alcohol (24:1) and ligated to pGEM-T-easy (Promega) according to the manufacturer's instructions. Ligation was performed overnight at 16°C in a 90-μL ligation reaction mixture and the ligated products were precipitated (*see* **Note 2**) and resuspended in 20 μL of water before it was electroporated into *E. coli* DH5α and plated on agar plates containing 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside and isopropyl-β-D-1-thiogalactopyranoside (IPTG).
8. A DNA template was prepared from the resultant colonies by using a QIAprep 96 Turbo Miniprep kit and the BioRobot 9600 (Qiagen). The plasmid DNA was sequenced by using a 3700 DNA Analyzer (Applied Biosystems) with the T7 and Sp6 primers.
9. Specific primers were designed to PCR amplify any regions of the DNA in which sequence was missing (*see* **Note 3**). Each PCR was performed in a 50 μL (total volume) mixture by using *Taq* DNA polymerase as recommended by the protocol. The PCR cycles were as follows: 94°C for 2 min; 30 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 2 min; and 72°C for 5 min. Two microliters of the PCR products was electrophoresed on an agarose gel to check the amplified DNA, the DNA was subsequently purified by using a Wizard PCR DNA preparation kit (Promega) and resuspended in 35 μL of water. Three microliters of DNA was sequenced with the same primers used for PCR amplification.
10. The Phred/Phrap software package from the Genome Center of the University of Washington was used for sequencing assembling. The open reading frames were recognized and selected by using the ORF Finder from the National Center for Biotechnology Information (NCBI). The BLAST database was used to search for sequence homologies and the amino acid sequence alignments were performed using CLUSTAL W. Potential transmembrane segments were identified by TMpred. The organization and GC% of *E. coli* O86 O antigen gene cluster are shown

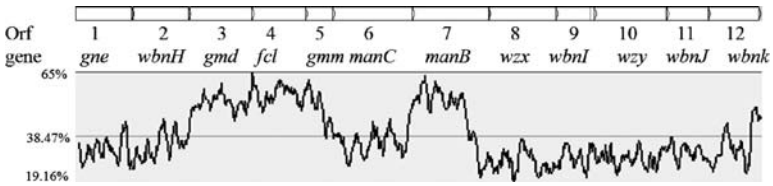


Fig. 7.3. The organization and GC% of *E. coli* O86 O antigen gene cluster.

in **Fig. 7.3**. The summary of. *E. coli* O86:H2 O antigen biosynthesis genes is shown in Table **7.1**.

3.2. Cloning, Expression, Purification, and Activity Assay of Glycosyltransferases

3.2.1. WbnH

1. *Cloning*: The *wbnH* gene was amplified by PCR from the *E. coli* O86:H2 chromosome DNA. The primers with restriction sites underlined were as follows: forward: 5′-GAGATATACATATGAAAAATGTTGGTTTTATTG (NdeI); reverse: 5′-CCGCCTCGAGTCAACCTAAAATAA TGCTTTTATATG (XhoI). PCR was performed in a volume of 50 μL: 5 μL of PCR buffer (10×), 2 μL of dNTP (each 10 mM), 1 μL of primer F (5 μM), 1 μL of primer R (5 μM), 0.5 μL of chromosomal DNA and 0.5 μL of DNA polymerase, and distilled H_2O added to a final volume of 50 μL. The PCR cycles were as follows: 94°C for 2 min; 30 cycles of 94°C for 30 s, 50°C for 45 s, and 72°C for 2 min; and then 72°C for 10 min. PCR product (50 μL) was purified by using the Wizard PCR Preps DNA Purification kit (Promega). The DNA fragments obtained were digested with restriction enzymes *Nde*I and *Xho*I in 100μL of digestion mixture: 10 μL of universal buffer (10×), 3 μL of *Nde*I, 3 μL of *Xho*I, and 50 μL of PCR product. The resulting DNA fragment was inserted into pET15b vector linearized by the same restriction enzymes to form pET15b-wbnH recombinant plasmid. The recombinant plasmid was confirmed by restriction mapping and sequencing.
2. *Expression*: The constructs were subsequently transformed into *E. coli* BL21 (DE3) competent cell following the general chemical transformation procedure for *E. coli*. *E. coli* BL21 (DE3) harboring the recombinant plasmid was grown at 37°C in 1 L Luria–Bertani (LB) medium with 100 μg/mL ampicillin antibiotic. When the cells were grown to OD 0.6–0.8, isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 0.15 mM. Expression was allowed to proceed for 12 h at 18°C. Cells were harvested, washed with 20 mM Tris–HCl (pH 7.0), and stored at –80°C until needed.

Table 7.1
The summary of *E. coli* O86:H2 O antigen biosynthesis genes

ORF[a]	Gene	No. of amino acids	Percentage of G+C content	Similar protein (Genbank accession no.)	Percentage of identical/% similar (no. of amino acid overlap)	Putative function
1	*gne*	339	33.2	Gne of *Yersinia enterocolitica* O:8 (AAC60777)	57/72 (337)	UDP-N-acetylglucosamine-4-epimerase
2	*wbnH*	339	34.2	WbcQ of *Yersinia enterocolitica* (CAA87705)	47/68 (337)	Glycosyltransferase
3	*gmd*	373	52.6	GDP-mannose 4,6-dehydratase of *E. coli* O128 (AA037691)	95/98 (370)	GDP-D-mannose dehydratase
4	*fcl*	321	56.1	GDP-fucose synthetase of *E. coli* O128 (AA037692)	99/99 (321)	GDP-fucose synthetase
5	*gmm*	167	52.4	GDP-mannose mannosyl hydrolase of *E. coli* O128 (AA037693)	94/97 (167)	GDP-mannose mannosyl hydrolase
6	*manC*	482	36.4	GDP-mannose pyrophosphorylase of *E. coli* O128 (AA037694)	85/93 (482)	GDP-mannose pyrophosphorylase
7	*manB*	462	50.5	Phosphomannomutase of *E. coli* O128 (AA037695)	92/95 (455)	Phosphomannomutase
8	*Wzx*	400	27.8	O antigen flippase of *Salmonella enterica* (AAV34509)	16/38 (218)	O antigen flippase
9	*wbnI*	134	28.4	Glycosyltransferase of *Homo sapiens* (AAQ88542)	23/42 (191)	Glycosyltransferase
10	*wzy*	466	28.8	O antigen polymerase of *E. coli* O111 (AAD46730)	25/44 (191)	O antigen polymerase
11	*wbnJ*	247	31.2	WbgO of *E. coli* O55 (AAL67559)	39/59 (217)	Glycosyltransferase
12	*wbnK*	302	32.6	Fucosyltransferase of *E. coli* O128 (AA037698)	32/53 (300)	Glycosyltransferase

3. *Purification:* the cell pellet was resuspended in 40 mL binding buffer: 50 mM Tris–HCl, pH 8.0, 0.2 M NaCl, 5 mM imidazole, and 1 mM PMSF, and sonicated on ice by using Branson Sonifier 450 (VWR Scientific) for 4 min with 2 s sonication and 2 s intervals at the energy magnitude of 100 W. The same sonication condition was repeated three times. The lysate was cleared by centrifugation (15,000×*g*, 20 min, 4°C) and loaded at a low rate of 2 mL/min onto a 5 ml Ni^{2+}-NTA chelating column (Amersham) equilibrated with binding buffer. The column was washed with 10 column volumes of wash buffer: 50 mM Tris–HCl, pH 8.0, 0.2 M NaCl, and 80 mM imidazole, and the protein was eluted with elution buffer: 50 mM Tris–HCl, pH 8.0, 0.2 M NaCl, and 250 mM imidazole. Protein eluted from the column was analyzed by SDS–PAGE following the common procedure. Protein concentration is determined by Bradford method.
4. *Activity assay:* Enzymatic assay for WbnH activity was performed with chemically synthetic acceptor GalNAcα-PP-*O*-(CH2)11-OPh.
   (1) *Chemical synthesis of acceptor substrate GalNAcα-PP-O-(CH2)11-OPh:* GalNAc phosphate (compound 1) (17) (73.0 mg, 0.281 mmol) and lipid phosphate (compound 2) (18) (65.0 mg, 0.189 mmol) were dried by co-evaporation from toluene/triethyl amine (3×7 mL/0.2 mL) separately. Compound 2 was dissolved in THF (6 mL) followed by addition of 1,10-carbonyldiimidazole (121 mg, 0.746 mmol). The reaction mixture was stirred at room temperature for 2 h and then methanol (200 μL) was added. The reaction mixture was stirred at room temperature for additional hour followed by removal of solvent. To the resulting intermediate was added the solution of compound 2 dissolved in THF (6 mL). The reaction mixture was stirred at room temperature for 36 h and concentrated. The residue was passed through Sephadex G-15 column eluted by methanol. The crude product was dried, dissolved in methanol (5 mL), and treated with sodium methoxide (3.0 mg). The reaction mixture was concentrated and purified by reverse phase C18 column chromatography by $H_2O/CNCH_3$ (10/1 then 10/2) to give GalNAc-diphosphate-lipid (compound 3) (67.1 mg, 60%).
   (2) *Enzymatic assay:* Total volume of 50 μL reaction mixture contains 5 mM GalNAcα-PP-O-(CH2)11-OPh, 5 mM UDP-GalNAc, 50 mM MES (pH 7.0), 10 mM $MnCl_2$, and 30 μg WbnH. The reaction was then

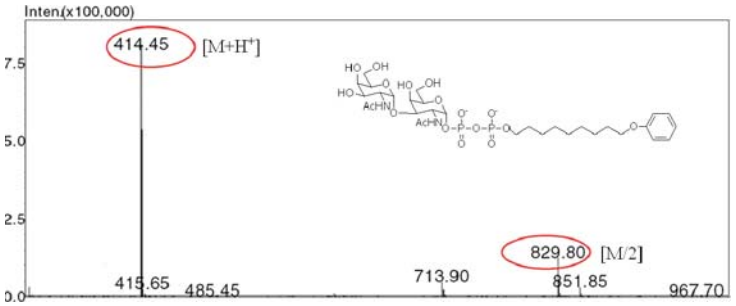


Fig. 7.4. Identification of disaccharide product catalyzed by WbnH using LC–MS.

quenched using methanol. Reaction was monitored and the product was identified using LC–MS. The mass spectrum (**Fig. 7.4**) showed a prominent peak with *m/z* ratio at 414.45 ($[M\text{-}2H]^{2-}$), and 829.80 ($[M\text{-}H]^{-}$), consistent with the formation of the GalNAc-GalNAc-PP-O(CH2)11-OPh product.

3.2.2. WbnJ

1. *Cloning, expression, and purification of WbnJ*: WbnJ was cloned, expressed, and purified using the same methods as described for WbnH with the exception that the different primers were used for amplification of *wbnJ*. The primers with restriction sites underlined for amplification of *wbnJ* gene were as follows: forward: 5′-GATATACATATGTCATTGAGAATATTAGAT (*NdeI*); reverse: 5′-CCGCCTCGAGTTATTTTATTAGTGCTTTAC (*XhoI*).
2. *Activity assay:* Enzyme assay for WbnJ was performed at 37°C for 2 h in a final volume of 100 μL containing 20 mM Tris–HCl, pH 7.0, 10 mM $MnCl_2$, 1 mM ATP, 0.3 mM radioactively labeled UDP-D-[6-^{3}H]Galactose (10,000 cpm) as sugar donor, 10 mM monosaccharide (GalNAcα-OMe) as sugar acceptor, and variable amount of enzymes. The reaction was terminated by adding 100 μL ice cold 0.1 M EDTA. Dowex 1×8-200 chloride anion exchange resin was then added in a water suspension (0.8 mL, v/v = 1/1). After centrifugation, supernatant (0.45 mL) was collected in a 20 mL plastic vial and ScintiVerse BD (10 mL) was added. The vial was vortexed thoroughly before the radioactivity of the mixture was counted in the LS-3801 liquid scintillation counter (Beckmann).
3. *Preparation of Galβ1,3-GalNAcα-OMe*: to provide the acceptor substrate for assaying the subsequent glycosyltransferase, i.e., WbnK, synthesis of disaccharide Galβ1,3-GalNAcα-OMe was performed on a larger scale by using WbnJ protein. One-pot reaction was conducted for three

days at 30 °C in a final volume of 4.0 ml containing 20 mM Tris–HCl (pH 7.5), 10 mM $MnCl_2$, 1 mM dithiothreitol (DTT), 10 mM UDP-Gal, 15 mM acceptor GalNAcα-OMe, and 15 mU WbnJ protein. The progress of the reaction was monitored by thin-layer chromatography conducted on Baker Si250F silica gel TLC plates. Products were visualized by staining solution: anisaldehyde/MeOH/H_2SO_4=1:15:2 (v/v/v). After complete conversion of donor substrate, protein was removed by brief boiling and centrifugation (12,000×*g*, 5 min). The supernatant mixture was purified by gel filtration chromatography Bio-Gel P2 (Bio-Rad). The desired fractions were pooled, lyophilized, and stored at –20°C.

3.2.3. Wbnk

1. *Cloning, expression, and purification of WbnK*: WbnK was cloned using the same procedure as described for WbnH with the exception that different primers and expression vector pGEX-4T-1 were employed. The primers with restriction sites underlined for amplification of *wbnK* were as follows: forward: 5′-CGCGGATCCATGGAAGTTAAAATTATTGGGGGGCT (*BamHI*); reverse: 5′-CGTTCTCGAGTCATAATTTTACCCACGATTCG (*XhoI*). For protein expression, a final IPTG concentration of 0.8 mM was used. For protein purification, the cell pellet was suspended in the bind buffer: 4.3 mM Na_2HPO4, 1.47 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.3 and disrupted by sonication on ice. After centrifugation, the lysate was loaded onto a 1 mL GST-Bind column (Invitrogen). After washing with the same buffer for four times, the protein was eluted with the elution buffer: 50 mM Tris–HCl, 10 mM reduced Glutathione, pH 8.0.
2. *Activity assay:* WbnK was assayed as described for WbnJ except that different sugar donor and acceptor are employed. WbnK was assayed using 0.3 mM radioactively labeled GDP-L-[U-^{14}C]fucose (7000 cpm) as sugar donor, and 10 mM synthesized disaccharide (Galβ1,3-GalNAcα-OMe) as sugar acceptor.
3. *Preparation of Fucα1,2-Galβ1,3-GalNAcα-OMe*: to provide the acceptor substrate to assay the next glycosyltransferase, i.e., WbnI, trisaccharide Fucα1,2-Galβ1,3-GalNAcα-OMe was synthesized on a larger scale by using WbnK/GST fusion protein. The reaction process was the same as mentioned before in disaccharide synthesis with the exception that the donor is 10 mM GDP-Fuc and acceptor is 13 mM disaccharide (Galβ1,3-GalNAcα-OMe) obtained previously.

3.2.4. WbnI

1. *Cloning, expression, and purification of WbnI*: WbnI was cloned, expressed, and purified using the same methods as described for WbnJ except that the different primers were used: forward: 5′-CCGCCATATGGTTATTAATATATTTT *(NdeI)*; reverse: 5′-GCACACTCGAGTTACTTCTTGATATTACCA (*XhoI*).
2. *Activity assay:* WbnI was assayed as described for WbnJ with the exception that different sugar donor and acceptor were employed. In WbnI activity assay, 0.3 mM radioactively labeled UDP-D-[6-H^3]Galactose (10,000 cpm) and 10 mM trisaccharide (Fucα1,2-Galβ1,3-GalNAcα-OMe) were used as sugar donor and acceptor, respectively.
3. *Preparation of tetrasaccharide Galα1,3(Fucα1,2)-Galβ1,3-GalNAcα-OMe:* The total reaction volume was 1 mL containing 20 mM Tris–HCl (pH 7.5), 10 mM $MnCl_2$, 1 mM dithiothreitol (DTT), 18 mM UDP-Gal, 20 mM acceptor trisaccharide (Fucα1,2-Galβ1,3-GalNAcα-OMe), and 20 mU WbnI protein.

3.3. In Vitro Reconstitution of Und-pp-Linked O-Repeating Unit of E. coli O86

With a chemically synthetic analog, GalNAc-PP-*cis*-pentaprenyl, as initial acceptor substrate, in vitro reconstitution of Und-pp-linked O-repeating unit of *E. coli* O86 was performed by sequential addition of sugar residues that was catalyzed by corresponding glycosyltransferase (**Fig. 7.5**).

1. The synthesis of GalNAc-PP-*cis*-pentaprenyl followed published procedures (17, 19).
2. Sequential synthesis of RU-PP-*cis*-pentaprenyl: The 1-mL reaction system contained 20 mM Tris–HCl (pH 7.5), 5 mM $MnCl_2$, 5 mM GalNAc-PP-*cis*-pentaprenyl, and 5 mM UDP-GalNAc. The reaction was initiated by the addition of 50 μg of the enzyme WbnH and was incubated at room temperature (*see* **Note 4**) for overnight or until the reaction was complete (ESI-Mass spectra (*see* **Note 5**) was used to monitor the reaction). The reaction was quenched by briefly boiling for 5 min. Proteins were removed by centrifugation. Subsequently, 5 mM UDP-Gal and 50 μg WbnJ were added to synthesize the trisaccharide (**Fig. 7.6A**). The reaction was incubated at room temperature overnight or until the reaction was complete. The reaction was quenched and the protein was removed as above. Similarly, Fuc and the second Gal were sequentially attached via using WbnK (**Fig. 7.6B**) and WbnI. The final product was verified by ESI-mass spectra (**Fig. 7.6C**).
3. Separation of RU-PP-*cis*-pentaprenyl: The crude residue was purified using reverse phase chromatography by

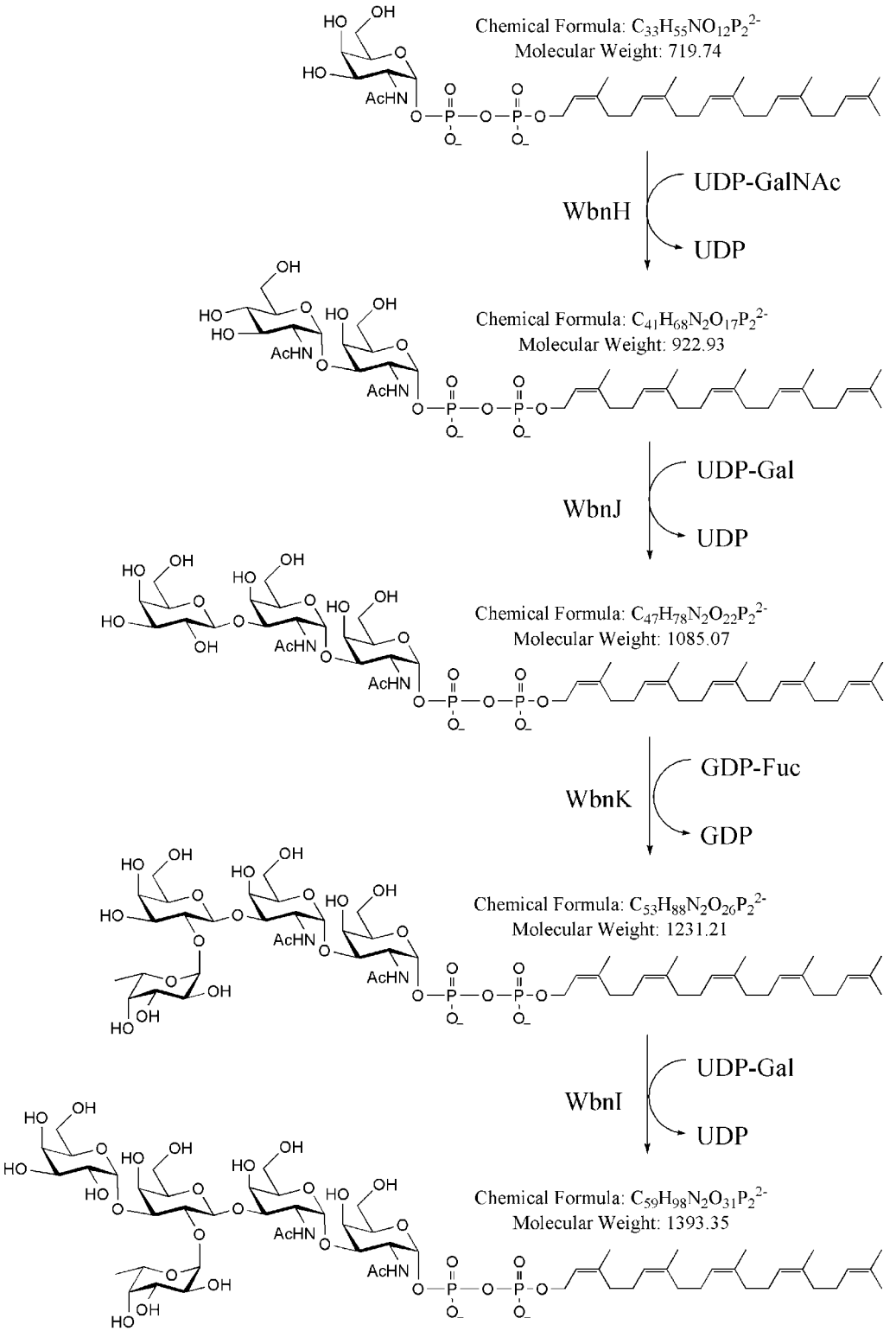


Fig. 7.5. Enzymatic synthesis of RU-PP-*cis*-pentaprenyl.

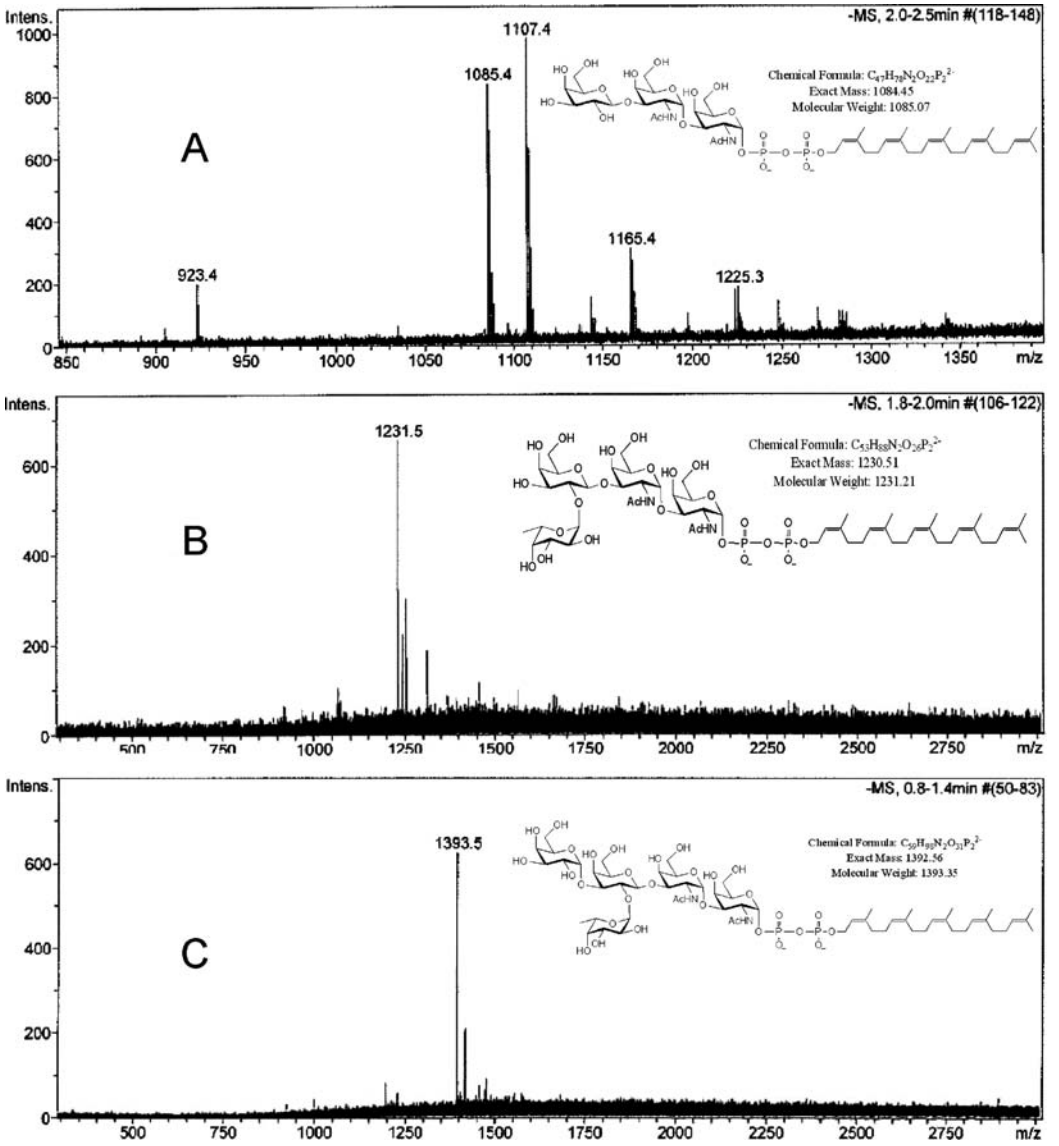


Fig. 7.6. Mass spectra of the synthesized oligosaccharide-linked pyrophosphate *cis*-pentaprenyl.

$H_2O/CNCH_3$ (10/1 then 10/2) to give the final pentasaccharide-linked lipid – RU-PP-Und (**Fig. 7.6C**).

4. Notes

1. The shearing condition needs to be optimized according to the manufacturer's instructions to generate DNA fragments, majority of which should have a size between 1.2 and 1.8 kb.

2. The precipitation is performed with addition of a 10% volume of 3 M sodium acetate, pH 5.2, and 2.5 times volume of cold ethanol followed by incubation at –70°C for at least 30 min.
3. There are often one or more DNA regions (termed gap) within the gene cluster that is not covered after initial assembly of sequence data. We need to design primers based on the DNA sequences flanking the uncovered regions and the PCR products are then sequenced to join the flanking DNA regions.
4. The large-scale reactions were always carried out at room temperature, since the glycosyltransferases tend to precipitate at 37°C.
5. ESI-mass spectra was performed at negative mode with $H_2O/CNCH_3 = 9/1$.

Acknowledgments

P. G. Wang acknowledges National Cancer Institute (R01 CA118208), NSF (CHE-0616892), Bill & Melinda Gates Foundation (51946) for financial support.

References

1. Jansson, P. E., Lindberg, B., Lindberg, A. A., and Wollin, R. (1981) Structural studies on the hexose region of the core in lipopolysaccharides from Enterobacteriaceae. *Eur. J. Biochem.* **115**, 571–577.
2. Raetz, C. R. and Whitfield, C. (2002) Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**, 635–700.
3. Amer, A. O. and Valvano, M. A. (2001) Conserved amino acid residues found in a predicted cytosolic domain of the lipopolysaccharide biosynthetic protein WecA are implicated in the recognition of UDP-N-acetylglucosamine. *Microbiology*, **147**, 3015–3025.
4. Amer, A. O. and Valvano, M. A. (2002) Conserved aspartic acids are essential for the enzymic activity of the WecA protein initiating the biosynthesis of O-specific lipopolysaccharide and enterobacterial common antigen in *Escherichia coli*. *Microbiology*, **148**, 571–582.
5. Liu, D., Cole, R. A. and Reeves, P. R. (1996) An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. *J. Bacteriol.* **178**, 2102–2107.
6. Feldman, M. F., Marolda, C. L., Monteiro, M. A., Perry, M. B., Parodi, A. J., and Valvano, M. A. (1999) The activity of a putative polyisoprenol-linked sugar translocase (Wzx) involved in *Escherichia coli* O antigen assembly is independent of the chemical structure of the O repeat. *J. Biol. Chem.* **274**, 35129–35138.
7. Daniels, C., Vindurampulle, C. and Morona, R. (1998) Overexpression and topology of the *Shigella flexneri* O-antigen polymerase (Rfc/Wzy). *Mol. Microbiol.* **28**, 1211–1222.
8. Franco, A. V., Liu, D. and Reeves, P. R. (1996) A Wzz (Cld) protein determines the chain length of K lipopolysaccharide in *Escherichia coli* O8 and O9 strains. *J. Bacteriol.* **178**, 1903–1907.
9. Heinrichs, D. E., Yethon, J. A. and Whitfield, C. (1998) Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. *Mol. Microbiol.* 3, 221–232.
10. Springer, G. F., Horton, R. E., and Forbes, M. (1959) Origin of anti-human blood

group B agglutinins in white Leghorn chicks. *J. Exp. Med.* **110**, 221–244.
11. Kochibe, N. and Iseki, S. (1968) Immunochemical studies on bacterial blood group substances. V. Structures of B determinant groups. *Jpn. J. Microbiol.* **12**, 403–411.
12. Feizi, T. (1985) Demonstration by monoclonal antibodies that carbohydrate structures of glycoproteins and glycolipids are onco-developmental antigens. *Nature* **314**, 53–57.
13. Lemieux, R. U., Bundle, D. R. and Baker, D. A. (1975) The properties of a "synthetic" antigen related to the human blood-group Lewis a. *J. Am. Chem. Soc.* **97**, 4076–4083.
14. Lemieux, R. U. and Driguez, H. (1975) The chemical synthesis of 2-O-(alpha-L-fucopyranosyl)-3-O-(alpha-D-galactopyranosyl)-D-galactose. The terminal structure of the blood-group B antigenic determinant. *J. Am. Chem. Soc.* **97**, 4069–4075.
15. Seto, N. O., Palcic, M. M., Hindsgaul, O., Bundle, D. R., and Narang, S. A. (1995) Expression of a recombinant human glycosyltransferase from a synthetic gene and its utilization for synthesis of the human blood group B trisaccharide. *Eur. J. Biochem.* **234**, 323–328.
16. Seto, N. O., Palcic, M. M., Compston, C. A., Li, H., Bundle, D. R., and Narang, S. A. (1997) Sequential interchange of four amino acids from blood group B to blood group A glycosyltransferase boosts catalytic activity and progressively modifies substrate recognition in human recombinant enzymes. *J. Biol. Chem.* **272**, 14133–14138.
17. Sim, M. M., Kondo, H., and Wong, C. H. (1993) Synthesis of dibenzyl glycosyl phosphites using dibenzyl N,N-diethylphosphoramidite as phosphitylating reagant: An effective route to glycosyl phosphates, sugar nucleotides and glycosides. *J. Am. Chem. Soc.* **115**, 2260–2267.
18. Montoya-Peleaza, P. J., Rileya, J. G., Szareka, W. A., Valvano, M. A., Schutzbachb, J. S., and Brockhausenb, I. (2005) Identification of a UDP-Gal: GlcNAc-R galactosyltransferase activity in *Escherichia coli* VW187. *Bioorg. Med. Chem. Lett.* **15**, 1205–1211.
19. Wacker, M., Linton, D., Hitchen, P. G., et al. (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli. Science* **298**, 1790–1793.

Chapter 8

Glycoprotein Characterization

Susan M. Twine, Luc Tessier, and John F. Kelly

Abstract

Increasing numbers of studies are reporting the modification of prokaryotic proteins with novel glycans. These proteins are often associated with virulence factors of medically important pathogens. Herein, we describe the steps required to characterize prokaryotic glycoproteins by mass spectrometry, using flagellin isolated from *Clostridium botulinum* strain Langeland as an example. Both "top-down" and "bottom-up" approaches will be described for characterizing the purified glycoprotein at the whole protein and peptide levels. The preliminary steps toward structural characterization of novel prokaryotic glycans by mass spectrometry and NMR are also described.

Key words: Glycoprotein, glycopeptide, collision-induced dissociation, electron transfer dissociation, nano high performance liquid chromatography–tandem mass spectrometry, glycan, oxonium ion.

1. Introduction

Glycosylation is one of the most common posttranslational modifications in eukaryotic cells and has been estimated to occur in up to half of all gene products (1). Studies have revealed a role for the combinatorial potential of carbohydrates in the modulation of protein function (2). Eukaryotic glycan expression is cell specific and regulated to allow the phenotypes of cells to change in response to environmental conditions, stage of development, etc. (3). Furthermore, glycosylation influences protein folding, subcellular targeting, and turnover (4). Eukaryotic glycoproteins have long been the focus of extensive and sophisticated research efforts worldwide. In contrast, glycosylation of prokaryotic proteins has been somewhat understudied. The accepted dogma was

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_8,

that prokaryotes were not able to glycosylate proteins and it is only relatively recently that this view has changed (5–8). The first reports of prokaryotic protein glycosylation were on the S-layer glycoproteins of Archaea (9) and there has been a steady increase in the number of publications documenting the identification and the characterization of glycoproteins, primarily surface-associated and secreted, from other bacterial species (10–22). Glycosylation of cell surface appendages such as pili and flagella has gained attention in view of their importance in motility and colonization of host cells (10, 23). Important steps in pathogenesis have been linked to the glycan substituents of bacterial proteins, indicating that prokaryotic protein glycosylation has a role to play in infection and pathogenesis and interference with host inflammatory immune responses (14).

Generally speaking, the glycosylation of proteins in eukaryotes is a highly conserved and refined process. Comparatively few monosaccharide species are used to build eukaryotic N-linked and O-linked glycan structures, which in turn are relatively conserved compared to the huge number of potential structures that could be formed. In contrast, prokaryotes produce a vast array of unusual monosaccharides and glycan structures that are often difficult to characterize using the analytical technologies developed for the more "predictable" eukaryotic glycans and glycoproteins. In this chapter we describe the mass spectrometry procedures we use to characterize prokaryotic glycoproteins and novel glycan moieties. *Clostridium botulinum* is a Gram-positive spore-forming anaerobic bacterium that produces the potent botulinum neurotoxin (BoNT). The neurotoxin is the causative agent of botulism, a descending symmetrical paralysis most commonly associated with BoNT contaminated food. *C. botulinum* BoNTF (serotype F), strain Langeland, was isolated from liver paste in Denmark in 1960 (24). In previous work, NMR and MS were used for structural assignment of a novel 417 Da glycan found on flagellins of the group I type A strain FE9909ACS Alberta and the type F strain Langeland (25). Using a combination of MS and NMR techniques, we have demonstrated that these strains of *C. botulinum* glycosylate flagellin with a novel derivative of the sialic acid-like nonulosonate sugar, legionaminic acid, αLeg5GluNMe7Ac. Herein, we describe the mass spectrometry techniques used in the characterization of the flagellin of *Clostridium botulinum*, strain Langeland.

2. Materials

2.1. Protein Samples

1. Protein samples are typically purified proteins, for example surface-associated bacterial proteins, such as flagellin or surface layer proteins. Putative glycoproteins that have been

purified using traditional column chromatography or other methods are also suitable.

2. Typically 10–20 μg of purified glycoprotein is required.

2.2. Materials and Reagents

All solvents used are of HPLC grade and are degassed before use. Water used in solvent preparations should be Milli-Q® (MQ) or higher purity.

1. Centrifugal filtration devices, such as Amicon Ultra series (Millipore, Billerica, MA, USA).
2. A reverse-phase HPLC column for nanoscale separations. For use with the nanoAcquity UPLC (Waters, Milford, MA): a Symmetry® C18, 180 μm × 20 mm, 5 μm trap column (Waters, Milford, MA), in combination with a BEH130 C18, 100 μm × 100 mm, 1.7 μm nanoAcquity UPLC column. For use with the Waters CapLC (Waters, Milford, MA): an Acclaim PepMap100 C18, 300 μm × 5 mm, 5 μm μ-Precolumn Cartridge (Dionex/LC Packings, Sunnyvale, CA), in combination with a ProteoPep™ 2 PicoFrit C18 column (New Objective, Ringoes, NJ).
3. A protein microtrap (Michrom Bioresources Inc., Auburn, CA).
4. Analytical scale reverse-phase HPLC column, for example, 4.6 mm×250 mm Jupiter C18 reverse-phase column with a SecurityGuard pre-column (Phenomenex, Torrance, CA, USA).
5. Syringe pump, capable of delivering 0.5–2.0 μL/min.
6. Hamilton syringes, 10–50 μL, compatible with syringe pump.
7. Solvent A: 0.2% (v/v) formic acid.
8. Solvent B: 0.2% (v/v) formic acid in acetonitrile.
9. Solvent C: 50% (v/v) aqueous methanol, 0.2% formic acid.
10. Solvent D: 50% (v/v) acetonitrile, 0.2% formic acid.

2.3. Instrumentation

1. A mass spectrometer (MS) equipped with an electrospray ionization source (ESI) and capable of tandem MS (MS/MS) analysis. For example, a hybrid quadrupole time-of-flight Q-TOF™ Ultima (Waters, Millford, MA) and/or an Ion Trap mass spectrometer equipped with electron transfer dissociation, ETD (Thermo Scientific, Waltham, MA).
2. An online nanoflow liquid chromatography (nanoLC) system. For example, either a CapLC capillary liquid chromatography system (Waters, Milford, MA), modified for nanoflow delivery or a nanoAcquity UPLC (Waters, Milford, MA).

3. An analytical HPLC system, for example an Agilent 1100 HPLC (Santa Clara, CA).

2.4. Software for the Analysis of Mass Spectrometry Data

1. Software for charge state deconvolution of protein mass spectra, for example MaxEnt (as part of MassLynx™ Software, Waters, Milford, MA).
2. Software for generating peak lists from nanoLC–MS/MS spectral data, such as ProteinLynx™ (Waters, Milford, MA).
3. Software for identifying peptide sequences from the nanoLC–MS/MS data using database searching (i.e., Mascot® version 2.1.0, Matrix Science Ltd., London, UK) and for de novo sequencing (i.e., PEAKS, Bioinformatics Solutions Inc., Ontario, Canada).

3. Methods

A diagrammatic overview of the techniques used in the characterization of prokaryotic glycoproteins is shown in **Fig. 8.1**.

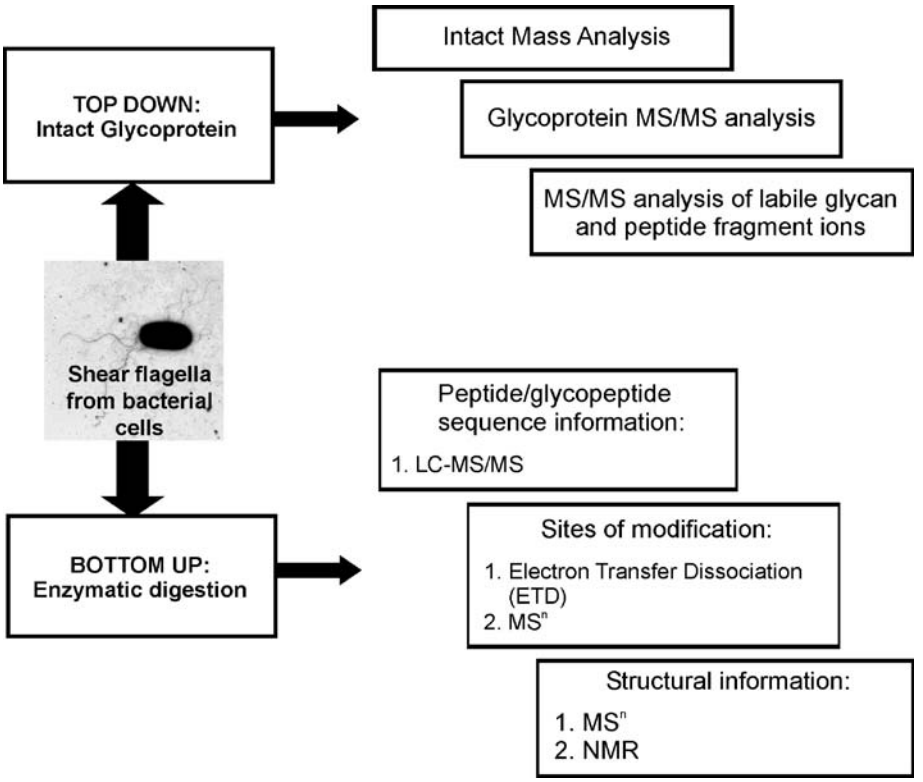


Fig. 8.1. Prokaryotic glycoprotein MS analysis workflow.

3.1. Protein Sample Preparation

It is recommended that proteins are isolated without the addition of mass spectrometry incompatible reagents, such as SDS, or other detergents and non-volatile salts. If the addition of detergent is unavoidable during protein purification steps, then we suggest the use of an acid cleavable detergent or using steps to remove such contaminants (*see* **Note 1**). For the purposes of this chapter, we describe work conducted in our laboratory to characterize the bacterial flagellin glycoprotein isolated from *Clostridium botulinum* strain Langeland. The procedure for isolating flagellin from bacterial cells is described elsewhere (24).

1. If dried or lyophilized, protein sample should be taken up in MQ-water and diluted to give a final concentration of 0.25–1.0 μg/μL.
2. Select a centrifugal filtration device, for example the Amicon Ultra series (Millipore, Billerica, MA, USA), with a molecular weight cut-off suitable to retain the protein. A molecular weight cut-off 2–3 times smaller than the solute is typically required for the solute to be fully retained by the membrane and for removal of small molecular weight contaminants. Prepare the filtration device as described in the manufacturer's instructions.
3. Pipette protein solution (10–20 μg protein) into centrifugal filtration device and dilute protein with Solvent A.
4. Spin as per the manufacturer's instructions in a bench-top centrifuge until the protein sample volume is one tenth of the total volume of the centrifugal filtration device. Add further solvent A and repeat this step four times. This serves to remove residual salts or low-molecular mass contaminants.
5. Concentrate the protein to a volume that will give a final estimated protein concentration of 0.25–1.0 μg/μL. Remove protein solution from spin column and store in a close-capped tube.
6. If possible, measure protein concentration using A_{280} or using a protein concentration assay such as the Bradford or Lowry assay.

3.2. Glycoprotein Intact Mass Analysis by Aqueous ESI-MS

1. Glycoprotein mass analysis was performed using a hybrid quadrupole time-of-flight mass spectrometer (Q-TOF Ultima, Waters). External calibration of the mass spectrometer was performed by infusing a 150-fmol/μL solution of Glu^1-fibrinopeptide B (Solvent C) at 0.6 μL/min. In our experience, if the temperature variation in the immediate environment can be minimized, then a calibration with [Glu^1]-Fibrinopeptide B, once per week, is sufficient to maintain a mass accuracy of 20–30 ppm.

2. Dilute the glycoprotein with Solvent D to yield a final solution of ~20% acetonitrile, 0.2% formic acid. Vortex briefly and centrifuge in a bench-top Eppendorf centrifuge for 5 min at 10,000×*g* to remove any insoluble protein or particulate matter. Carefully remove the supernatant and transfer into a clean tube.
3. Transfer the diluted glycoprotein sample to a 25-μL or 50-μL Hamilton syringe and place in syringe pump.
4. Connect syringe pump to electrospray interface of mass spectrometry and start infusing protein at 0.5–1.0 μL/min.
5. Mass spectrometer settings will vary. If using a Q-TOF2 instrument (Waters, Milford, MA), set the electrospray voltage to 3.5–4 kV and the cone to ~30–60 V. Acquire data between mass/charge (*m/z*) 100 and 2400 with a 1.0 s scan duration and 0.1 s inter-scan interval.
6. The length of time required for data collection will depend on the quality and intensity of the glycoprotein mass spectrum (~30 s–2 min). A series of multiply charged glycoprotein ion peaks will be observed, as shown in **Fig. 8.2a**.
7. Combine spectra to obtain an average spectrum with an acceptable signal-to-noise ratio.
8. Zoom into the specific *m/z* region that you wish to analyze and use charge state deconvolution software, such as MaxEnt (Waters Ltd.) to calculate the molecular mass of the glycoprotein species (resolution of 1 Da/channel, Uniform Gaussian width at half height, 0.75 Da, minimum intensity ratios of 33%, iterating to convergence). The results of charge deconvolution are shown in the inset in **Fig. 8.2a**.

3.3. Glycoprotein Intact Mass Analysis by Non-aqueous ESI-MS

In some cases the flagellin protein is insoluble and precipitates from aqueous solution. Therefore, we developed a non-aqueous solvent system to manage these samples.

1. Place 10–20 μg of protein in an Eppendorf tube and evaporate to dryness in a SpeedVac.
2. Add 5–10 μL of 98% formic acid and agitate the sample gently (do not vortex) until the protein dissolves.
3. Dilute the protein solution 10-fold with 1,1,1,3,3,3-hexafluoroisopropanol, 99% (HFIP).
4. Immediately infuse samples into electrospray mass spectrometer, as described above at a flow rate of 0.5–1.0 μL/min.
5. Non-aqueous intact mass experiments can be performed in the same manner as described above for aqueous solvent systems (*see* **Note 2**).

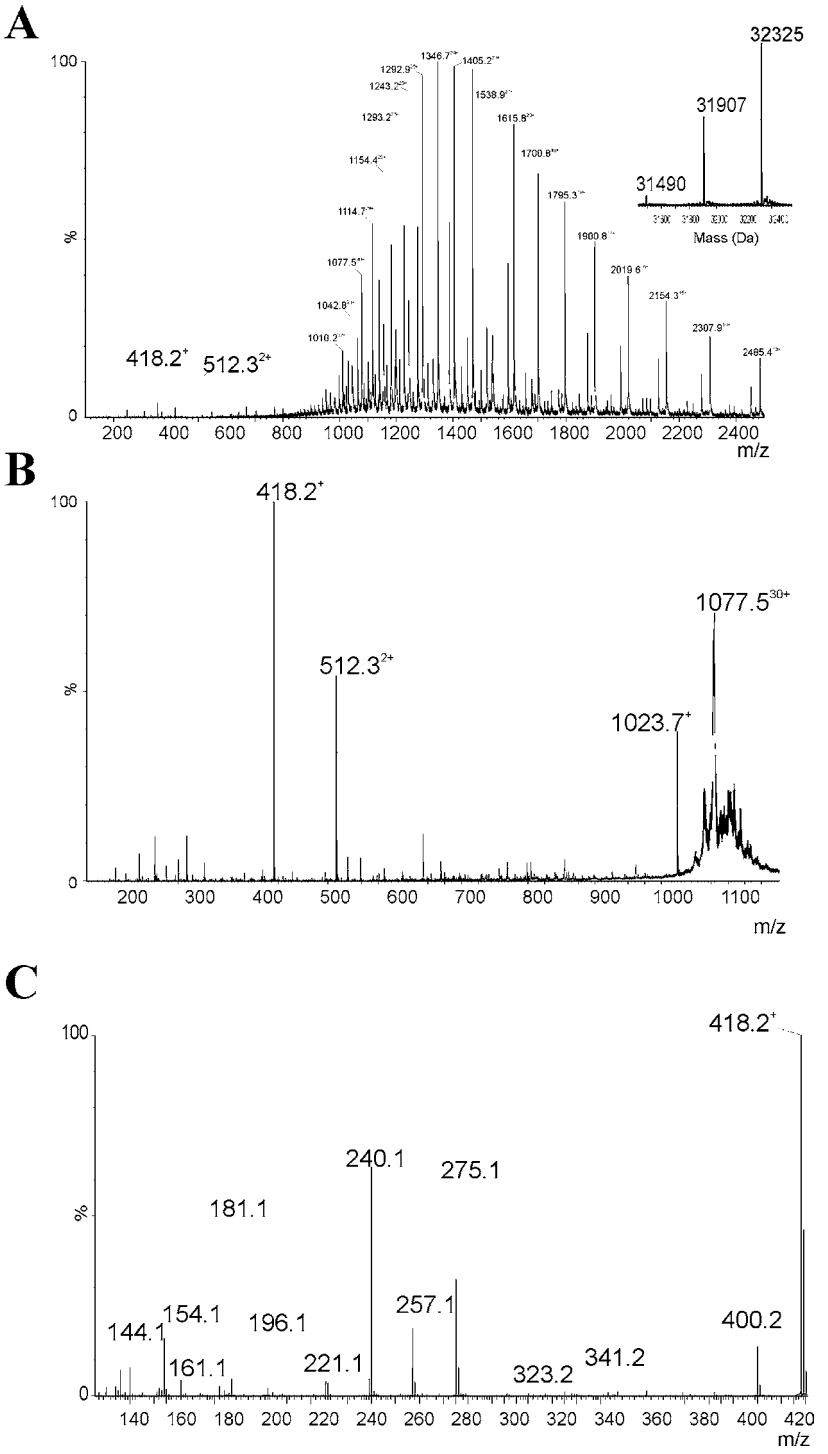


Fig. 8.2. Electrospray mass spectrometry and tandem mass spectrometry analyses of intact flagellin protein from *C. botulinum* strain Langeland. (A) Electrospray mass spectrum of intact flagellin. The inset shows the reconstructed molecular mass profile showing three major peaks at 31,490, 31,907 and 32,325 Da, respectively. Note that the separation between these glycoprotein isoforms is equivalent to one glycan (αLeg5GluNMe7Ac, 417 Da). Also visible in the lower mass region are glycoprotein fragment ions at m/z 512.32^{2+} and 418.2^{+}. (B) Tandem mass spectrometry analyses

3.4. Trap and Elute LC-ESI-MS

Trap and elute LC-ESI-MS allows on-line sample clean-up and preconcentration prior to MS analysis of intact proteins. This method may be more appropriate for the analysis of glycoproteins that cannot be isolated to sufficient purity for direct infusion. The following was conducted with a supernatant obtained from broth growth of *C. botulinum* (24). Bacterial cells were removed by centrifugation and supernatant was analyzed directly, as described below:

1. Connect a protein microtrap (Michrom Bioresources Inc., Auburn, CA) to a manual external loop injector which, in turn, is connected to a gradient HPLC pump. The mobile phase flow rate will depend on the trap used. In this instance ~60 μL/min was found to be optimal. An Agilent 1100 HPLC system was used for this experiment, so a pre-column splitter was installed to cut down the flow from the HPLC pump from 0.5 mL/min to 60 μL/min (25). The eluate from the trap apparatus was directed into the electrospray interface of the Q-TOF2 to allow real-time monitoring of ion elution profiles.
2. Place the manual injector in load mode so that the sample flow is directed through the protein microtrap before being sent to waste. Dilute sample solution 1:1 with solvent A and inject onto the trap.
3. Still in load mode, rinse the trapping column with 500 μL of solvent A to remove soluble contaminants. This step may need to be repeated, depending on the concentration of contaminants in the sample.
4. Place manual injector into inject mode such that the protein trap is brought online with the Q-TOF mass spectrometer.
5. Elute protein(s) using a gradient of 5–60% Solvent B over 30 min at 0.5 mL/min recording the LC–MS spectra using the parameters described in **Sect. 3.2**.
6. Combine spectra to obtain an average signal with acceptable level of signal-to-noise ratio.
7. Zoom into the specific *m/z* region that you wish to analyze and use deconvolution software, such as MaxEnt (Waters Ltd.) to calculate that molecular mass of the glycoprotein.

Fig. 8.2. (continued) of the multiply charged glycoprotein ion at *m/z* 1077.5^{30+} showing the same fragment ions at *m/z* 512.3^{2+} and 418.3^{+}, indicating that these ions arise from the glycoprotein. (C) Second-generation product ion spectra of the glycan oxonium ion at *m/z* 418.2. The glycan precursor ion was generated by in-source fragmentation of the multiply charged flagellin ions as described in **Sect. 3.5**. This feCID-MSMS spectrum yielded a pattern of fragment ions that resembled those generated for other pseudaminic and legionaminic acid derivatives.

3.5. "Top-Down" MS/MS Analysis to Detect Glycan PTM's and to Provide Information About the N-Terminal and/or C-Terminal Amino Acid Sequence

Multiply protonated glycoprotein precursor ions subjected to low-energy collisional activation in a quadrupole time-of-flight instrument typically show extensive and specific gas-phase deglycosylation resulting in the formation of abundant glycan oxonium ions with relatively few fragment ions from peptide bond cleavages. Structural information on individual carbohydrate residues can be obtained on a Q-TOF instrument using a second-generation product ion scan of oxonium ions formed by collisional activation in the source, a process called front-end collision-induced dissociation (feCID).

1. Infuse glycoprotein solution, as described in **Sect. 3.2 or 3.3**.
2. In MS/MS mode, select an abundant multiply charged protein ion, for example the protein ion at m/z 1077. 5^{30+} (**Fig. 8.2b**).
3. Increase the collision energy incrementally, observing the appearance of labile protein associated fragment ions. Shown in **Fig. 8.2b** is the MS/MS spectrum with a collision offset of ~50 V (laboratory frame of reference), collecting data for 1 min. Three fragment ions dominate this spectrum, m/z 418.2, m/z 512.3, and m/z 1023.7. These ions were also observed in the lower mass region of the mass spectrum of the intact protein and are indicated in **Fig. 8.2a**.
4. To generate MS/MS spectra of these fragment ions on a Q-TOF instrument, return to MS mode and adjust the cone voltage (Q-TOF2) to fragment the protein ions as they enter the mass spectrometer. Typically, the cone voltage is increased from 35 V to ~70–100 V.
5. Select the ion of interest for MS/MS analysis. Increase the collision energy from 10 to 25 V incrementally, until the fragment ions are readily visible. Shown in **Fig. 8.2c** is the MS/MS spectrum of the singly charged ion at m/z 418.2, which is composed of numerous glycan-associated fragment ions and none that could be attributed to peptide y and b fragment ions. This MS/MS spectrum is identical to that obtained for another flagellar glycan oxonium ion from a related species, *C. botulinum* strain Alberta (25). Therefore, it is very likely that this glycan is also the same novel legionaminic acid derivative, 7-acetamido-5-(N-methyl-glutam-4-yl)-amino-3,5,7,9-tetradeoxy-d-*glycero*-α-d-*galacto*-nonulosonic acid, (αLeg5GluNMe7Ac). The MS/MS spectrum of the doubly protonated ion at m/z 512.3 indicated that this fragment ion originated from the C-terminal end of the flagellin, QGVLQLLR (data not shown).

3.6. Bottom-Up LC–MS and LC–MS/MS Analysis of a Proteolytic Digest

Bottom-up studies are used to identify the location of the glycosylation sites, the nature of the linkage to the peptide (N-linkage or O-linkage), and the extent of glycosylation at each site. A proteolytic enzyme, typically trypsin, or a chemical degradation method is used to cleave the glycoprotein into medium-sized peptides, ~5–15 amino acids in length, that are suitable for LC–MS/MS analysis. If the glycoprotein contains disulfide bonds then it is usually necessary to reduce and alkylate them prior to digestion. If the purified glycoprotein does not contain cysteines, then steps 1–5 can be omitted.

1. Dissolve the glycoprotein in 6 M Guanidine HCl, 1 M Tris/HCl, pH 7.5 (~1 μg/μL) and add freshly prepared DTT to a final concentration of 4 mM. Vortex to mix and spin at 10,000×*g* for 10 s to bring solution down to the bottom of the tube.
2. Incubate samples at 95°C for 10 min to reduce the cysteine residues.
3. Vortex to mix and spin at 10,000×*g* for 10 s to bring solution down. Cool tubes at room temperature for 2 min.
4. Add freshly prepared IAA to a final concentration of 10 mM and incubate the sample at room temperature for 20 min in the dark to alkylate the cysteine residues.
5. Vortex to mix and spin at 10,000×*g* for 10 s to bring solution down. Remove buffer and reagents from the solution by dialysis or by centrifugal filtration as described in **Sect. 3.1**.
6. The desalted glycoprotein is dissolved/suspended in 50 mM ammonium bicarbonate at a concentration of ~0.2–1 mg/mL. Add trypsin (in Milli-Q® water) to a substrate to trypsin ratio of 30:1. Mix and spin at 10,000×*g* for 10 s to bring solution down. Incubate at 37°C for 4–8 h.
7. Dilute the digest 1:100 with Solvent A and inject 5 μL onto the nanoLC-Q-TOF Ultima™ MS.
8. Rinse the trapping column with the sample flow stream (Solvent A, 40 μL/min) for 3 min to ensure that all soluble contaminants are removed.
9. Place the LC-MS system into analysis mode bringing the peptide trap online with the reverse-phase nanocolumn and the electrospray mass spectrometer.
10. Use the following gradient conditions to elute the peptides: (1) 5% solvent B for 5 min, (2) 5–60% solvent B over 40 min, and (3) 60% solvent B for 6 min before returning to 5% solvent B over 5 min.

11. Set the electrospray voltage on the Q-TOF Ultima to 4 kV and set to automatic acquisition of MS/MS spectra on doubly, triply, and quadruply charged ions (data-dependent analysis) (*see* **Note 3**).
12. Submit the nLC–MS/MS file to the Mascot® search engine or any other probability-based algorithm. The following parameters can be used: (i) specify trypsin enzymatic cleavage with one possible missed cleavages; (ii) allow variable modification for oxidation (+15.99 Da) at the methionine residues; (iii) allow variable modification for carbamidomethyl (iodoacetamide derivative; +57.02 Da) at the cysteine residues if the glycoprotein was alkylated; (iv) set parent ion tolerance $\leq$ 0.5 Da; (v) set fragment ion tolerance $\leq$ 0.2 Da; and (vi) search against a database containing the sequence of the glycoprotein under investigation. Peptides not modified with glycans can be identified in this manner.
13. Glycopeptides will not be identified by traditional database searching methods and so their MS/MS spectra must be examined manually. In many cases, the glycan oxonium ion identified during "top-down" studies (**Fig. 8.2a** and **b**) will be also observed in the glycopeptide spectra, serving as a useful marker ion (**Fig. 8.2a**). De novo peptide sequencing is described elsewhere (26). See also **Note 4** for information regarding software that facilitates de novo peptide sequencing from tandem mass spectrometry data. Refer to **Note 5** for details regarding the use of enzymes other than trypsin in proteolytic digestions of glycoproteins.

3.7. Characterization of Glycopeptides

When attempting to identify sites of glycan attachment or to increase peptide coverage we recommend an HPLC fractionation step which facilitates more elaborate downstream experiments, such as MS^n, electron transfer dissociation, or beta-elimination (27).

3.7.1. HPLC Fractionation of Peptide Digests

1. Set HPLC, for example an Agilent 1100 series HPLC with a diode array detector (Agilent Technologies, Palo Alto, CA, USA) to injection mode and load 100 μL of tryptic digest (~10 μg protein).
2. Switch to inject mode and inject onto a 4.6 × 250 mm Jupiter C18 reverse-phase column with a Phenomenex precolumn (SecurityGuard, Torrance, CA, USA).
3. Use the following gradient conditions to separate peptides, with a flow rate of 1 mL/min: (1) 5% solvent B for 10 min, (2) 5–60% solvent B over 40 min, and (3) hold at 60% solvent B for 5 min prior to reducing to 5% solvent B for 5 min.

4. Collect fractions every 1 min during gradient elution.
5. Monitoring peptide elution: Option (1) – If the HPLC is equipped with a UV detector then monitor the absorbance at 204 nm to detect the peptides as they elute from the column. Option (2) – Install a post-column splitter to divert approximately 60 μL/min of column eluate to the interface of an electrospray mass spectrometer to allow real-time monitoring of peptide ion elution profiles.
6. Evaporate the fractions to dryness and store at –20°C. It is advisable to retain a small aliquot (~10–20 μL) of each fraction for screening by infusion ESI-MS and ESI-MS/MS to confirm the presence and location of the glycopeptides of interest in preparation for further analyses.

3.7.2. CAD-MS^n (MS^2 and MS^3) to Investigate Peptide Sequence and Glycan Composition

1. Place ~10–20 μL of a glycopeptide-containing HPLC fraction in a Hamilton syringe and install in a syringe pump of a mass spectrometer, equipped with an electrospray ion source, for example an LTQ-XL linear ion trap (Thermo Fisher Scientific) or a Q-TOF mass spectrometer (Waters). Set the flow rate to 0.5–1.0 μL/min.
2. Select a multiply charged glycopeptide ion for MS^2 analysis and increase the collision energy until the b and y fragment ions are clearly observed and the intensity of the glycan oxonium ion is maximized. **Figure 8.3a** shows the MS/MS spectrum obtained for the flagellar tryptic glycopeptide, IASQTQFNTK, from *C. botulinum* strain Langeland. Peptide y and b ions are indicated. The glycan oxonium ion can be observed at m/z 418.2.
3. Further structural information can be obtained by performing MS^3 analysis on the MS^2 fragment ions. For example, the MS^3 fragment ion spectrum of the glycan oxonium ion at m/z 418.2 is identical to that generated directly from the intact protein as described in **Sect. 3.5** (**Fig. 8.2c**). MS^3 analysis can easily be performed on an ion trap mass spectrometer though, as a general rule, the bottom third of a CAD fragment ion spectrum generated in an ion trap is lost. MS^3 fragment ion spectra can also be generated using a Q-TOF mass spectrometer in the manner described in **Sects. 3.5** and **3.7.4**.
4. It is often advantageous to perform further MS^n (MS^4 or greater) analyses to provide further insight into the composition of the novel glycan (15). Note that this sort of analysis can only be performed on an ion trap instrument and that the intensity of the fragment ions will decrease approximately 10-fold with each sequential step in the MS^n experiment.

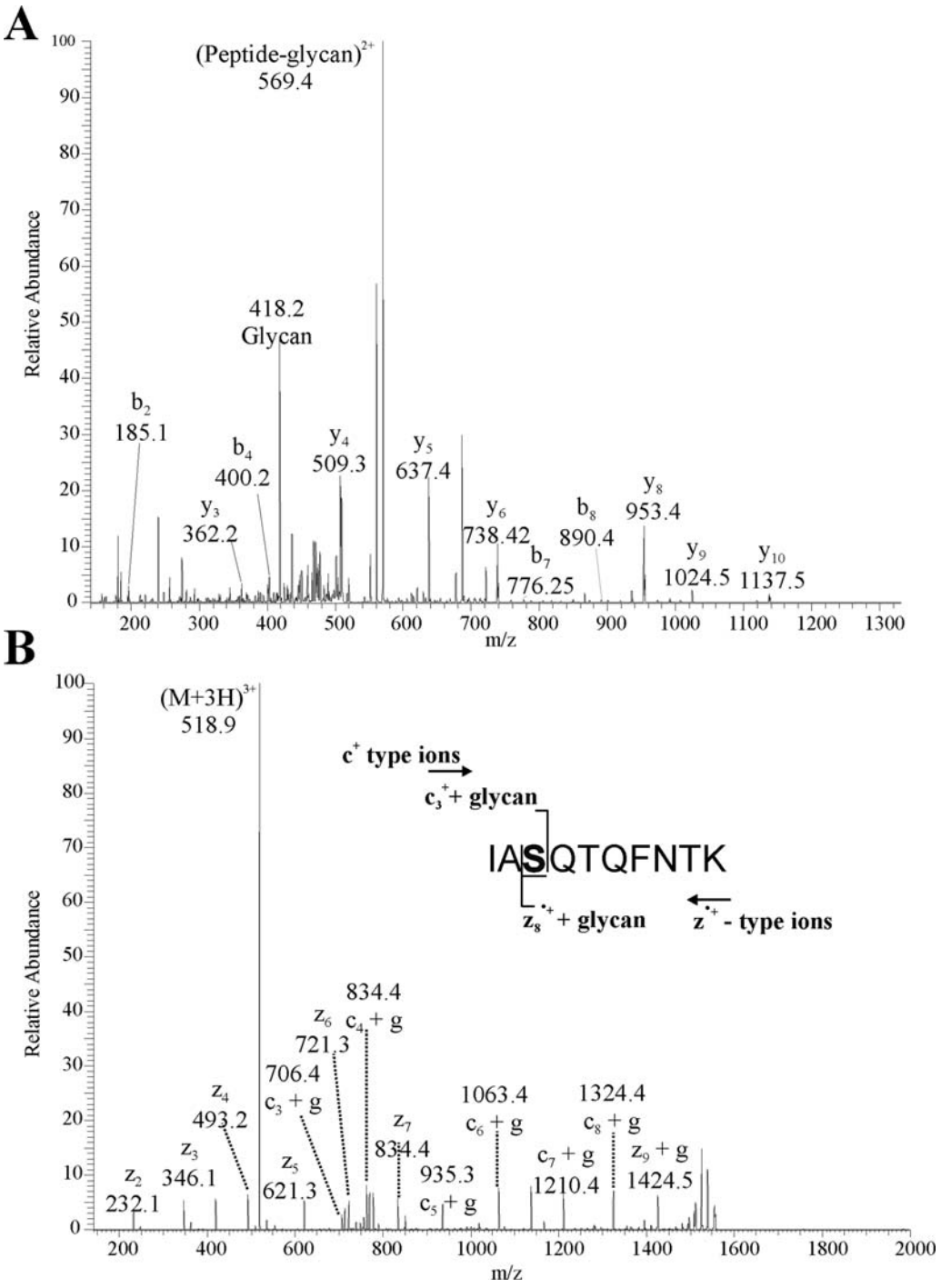


Fig. 8.3. CAD and electron transfer dissociation (ETD) MS/MS analysis of the IAS*QTQFNTK tryptic peptide (asterisk indicates modified amino acid) from the FlaA flagellin protein of *C. botulinum* strain Langeland. (A) CAD-MS/MS spectrum of triply protonated glycopeptide ion at *m*/*z* 518.9. Peptide y and b fragment ions are indicated and an intense ion at *m*/*z* 418.2 was observed, corresponding to the putative glycan oxonium ion. Additionally, a doubly charged ion corresponding to the unmodified glycopeptide was observed at *m*/*z* 569.4 (B) ETD MS/MS spectrum obtained for the same precursor ion. Observed c and z fragment ions are indicated and determined that the 417 Da glycan modification is attached to the serine residue.

3.7.3. Using Electron Transfer Dissociation (ETD) to Identify Sites of Glycosylation

ETD preserves delicate modifications during the fragmentation process and is ideal for identifying the linkage sites of O-glycans (28, 29).

1. Place 10–20 μL of a glycopeptide-containing HPLC fraction in a Hamilton syringe and place into syringe pump of an LTQ XL linear ion trap or another electrospray mass spectrometer capable of performing ETD. Set the flow rate to 0.5–1 μL/min.
2. Select the glycopeptide precursor ion to be analyzed and adjust the ETD reaction time for optimal fragmentation. The optimum ETD reaction time for the *C. botulinum* glycopeptide, IASQTQFNTK, was 35 ms (**Fig. 8.3b**) though, in our experience, this can vary significantly from one glycopeptide to another. ETD cleaves the bond between the amide nitrogen and the α-carbon generating c and z fragment ions that often cover the entire peptide sequence. The site of glycosylation can be elucidated by manually interpreting the ETD spectrum. In this example, ETD clearly demonstrated that the glycan is O-linked to the serine residue of the glycopeptide (**Fig. 8.3b**). *See* **Note 6**.
3. If ETD is not available then sites of O-glycosylation can be determined by beta-elimination. O-Linked glycans are removed by alkaline hydrolysis, a process which imparts an decrease in mass of –1 to serine or threonine residues to which the glycans were attached (30).

3.7.4. Determining Sites of N-Linked Glycosylation Using a Q-TOF Mass Spectrometer

N-Linked glycans are not as labile under CAD conditions as those that are O-linked and so it is often possible to identify N-glycosylation sites either by MS^2 or MS^3. In the example given here, feCID-MS/MS was performed on a Q-TOF mass spectrometer to determine the site of glycosylation of PEB3, a periplasmic glycoprotein from *Campylobacter jejuni* (14). This glycoprotein was purified by ion exchange and size exclusion chromatography, digested with trypsin (**Sect. 3.6**) and HPLC fractionated (**Sect. 3.7**) prior to analysis.

1. Load 20–30 μL of the glycopeptide-containing HPLC fraction in a Hamilton syringe and place into syringe pump. Infuse the solution into the microelectrospray interface of the Q-TOF2 mass spectrometer at a flow rate of 0.5–1 μL/min.
2. Select the glycopeptide precursor ion for MS/MS analysis and adjust the collision offset for optimal fragmentation. In this case the doubly protonated glycopeptide ion at m/z 1057.9 was selected and the collision offset was 25 V, laboratory frame of reference. The MS/MS spectrum is composed primarily of fragment ions resulting from the sequential loss

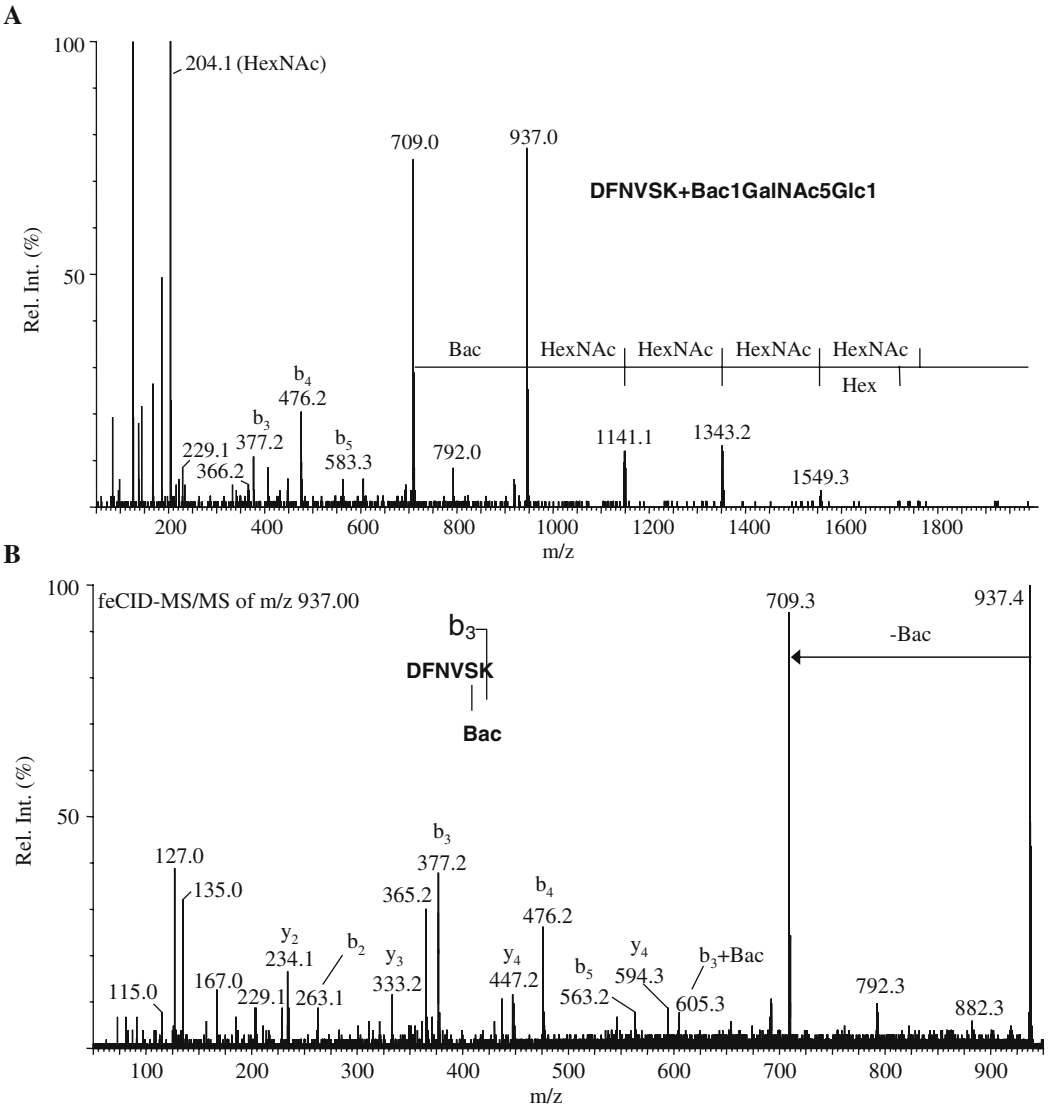


Fig. 8.4. MS/MS analysis of the PEB3 tryptic glycopeptide. (A) MS/MS spectrum of the doubly protonated glycopeptide ion at *m/z* 1057.9. The fragment ions originating from the sequential loss of oligosaccharide residues are indicated in the spectrum. The peptide sequence and glycan composition are shown in the inset. (B) Second-generation product ion spectrum of the singly charged glycopeptide fragment ion at *m/z* 937.4. The glycopeptide was fragmented by front-end collision-induced dissociation (orifice voltage 100 V) as it entered the mass spectrometer. The observation of the b_3-228 fragment ion at *m/z* 605.3 confirmed that the oligosaccharide is N-linked (reproduced from (34)).

of sugar residues from the heptasaccharide attached to the peptide though some peptide-associated fragment ions are also observed (**Fig. 8.4a**).

3. Return to MS mode and set the orifice voltage to 100 V to fragment the glycopeptide as it enters the mass spectrometer (front-end CID).

4. Perform MS/MS analysis on fragment ions composed of the peptide but with one or two sugar residues still attached. In this instance the singly charged fragment ion at *m/z* 937.4, corresponding to the peptide with just the linking sugar (bacillosamine), was chosen for MS/MS analysis (20 V offset, laboratory frame of reference). Acquire MS/MS data until the fragment ions are clearly visible (usually 1–2 min).
5. The feCID-MS/MS spectrum (**Fig. 8.4b**) of this glycopeptide fragment ion confirmed that the glycan is N-linked to the asparagine residue. The ion at *m/z* 605.3 is the b_3 peptide ion with the bacillosamine sugar. Furthermore, the amino acid sequence of this peptide has the extended N-linked consensus sequon (D/E-X-N-X-S/T, where X is any amino acid other than proline) that is unique to *Campylobacter* species (31).

3.8. Identifying Novel Glycan Modifications

The carbohydrate modifications on prokaryotic glycoproteins are often composed of novel glycans that have never been observed before. Our strategy is to combine mass spectrometry and NMR to identify these novel glycans.

3.8.1. MS^n Analysis of Novel Glycan Residue

This procedure has been described in **Sect. 3.7.2**.

3.8.2. Exact Mass and Elemental Composition Analysis of Novel Glycan Residues

1. Place 50–60 μL of a purified glycopeptide solution in a Hamilton syringe and infuse at 0.5–1.0 μL/min into the electrospray interface of a medium- to high-resolution mass spectrometer such as the Q-TOF Ultima. The amino acid sequence of the glycopeptide must be known.
2. Select the precursor ion of interest and begin MS/MS fragmentation. Adjust the collision energy until peptide type y and b ions and the oxonium ion of the novel glycan are clearly visible.
3. Collect MS/MS spectra until the fragment ions are smooth, symmetrical, and clearly stand out from the background (a signal-to-noise ratio of at least 30:1). Note that the acquisition time will depend on the intensity of glycan and peptide fragment ions and may take as long as 30–60 min.
4. Combine the MS/MS spectra and recalibrate the mass spectrum using peptide fragment ions bracketing the glycan oxonium ions as internal mass standards. A mass accuracy of at least 10 ppm can be achieved in this manner on the Q-TOF mass spectrometer allowing the mass of the unknown glycan ion to be determined accurately to greater than two places in decimals. The accurate mass of the glycan oxonium ion from *C. botulinum* Alberta strain was determined to be 418.182 Da (±0.002 Da) (25).

5. Input the glycan accurate mass with elemental composition calculators, some of which are freely available (http://library.med.utah.edu/masspec/elcomp.htm), to calculate plausible elemental formulas for the novel glycan moiety. The top-ranked plausible elemental composition for the flagellin glycan residue from *C. botulinum* Alberta strain was $C_{17}H_{28}N_3O_9$ which is the formula of the sugar identified by NMR, 7-acetamido-5-(N-methyl-glutam-4-yl)-amino-3,5,7,9-tetradeoxy-D-*glycero*-α-D-*galacto*-nonulosonic acid, (αLeg5GluNMe7Ac). All evidence suggest that this sugar is also found on the flagellin from *C. botulinum* Langeland strain.

3.8.3. Isolation for Structural Analysis by NMR

NMR is a powerful tool for identifying novel glycans (Chapter 11 in this volume). However, it is relatively insensitive compared to mass spectrometry, requiring at least 20–50 μg of purified glycan or glycopeptide. Larger glycans such as the heptasaccharide found of PEB3 from *Campylobacter jejuni* can be isolated by size exclusion chromatography after pronase/proteinase K digestion. Smaller glycans can be isolated as short glycopeptides by HPLC or by releasing the glycan from the glycoprotein by beta-elimination.

To identify the novel glycan on flagellin from *C. botulinum* Alberta strain, 20 HPLC fractions containing the tryptic glycopeptide, ^{169}VGSINIGSGK^{178}, were pooled and evaporated to dryness in preparation for NMR analysis (Chapter 11). The HPLC separation and fractionation conditions were as described in **Section 3.6**.

4. Notes

1. In some instances detergents are essential to enhance the solubility of proteins or peptides prior to MS analysis. Typically, detergents are not compatible with electrospray mass spectrometry. Recently, the use of an acid cleavable detergent, pyridinium propyl sulfonate or PPS (3-[3-(1,1-bisalkyloxyethyl)pyridin-1-yl]propane-1-sulfonate) has been reported (32). PPS is a zwitterionic surfactant that can be hydrolyzed at low pH into soluble by-products that have no residual detergent properties. The soluble by-products can be easily removed and the cleaved detergent is compatible with reverse-phase chromatography. Additionally, high concentrations of detergents or denaturants such as SDS and urea can inactivate enzymatic activity during protein digestion for bottom-up studies. If the amount of sample is not limited, these contaminants may be removed by

dialysis against water or by protein precipitation prior to protein digestion. Detergent removal cartridges offer another alternate method for the removal of excess ionic and non-ionic detergents from both peptide and protein samples prior to MS analysis and are available from vendors such as PolyLC (http://www.polylc.com/).

2. The non-aqueous ESI-MS method has proved very useful for analyzing hydrophobic proteins of mass 35 kDa and less. We have noted small mass shifts for proteins greater than 35 kDa, most likely caused by solvent adducts. These adducts have proven difficult to remove using a Q-TOF instrument. Secondly, it is important to infuse the protein solution immediately after the addition of HFIP to avoid the formation of crystals in the electrospray source, which can block capillary transfer lines and electrospray tips. The longer the protein is in the HFIP solution, the greater the likelihood of crystal formation.
3. The mass spectrometer surveys the incoming ions in MS-only mode (m/z 400–2000, one scan per second) and selects up to eight of the most intense multiply charged ions for MS/MS analysis. In this instance, four MS/MS spectra are acquired per precursor ion and the collision offset is chosen automatically by the mass spectrometer based on their size and charge state.
4. Using de novo sequencing software, such as PEAKS Studio (Bioinformatics Solutions, Waterloo, ON Canada. Good quality spectra are often left unexplained after database searching; a database search cannot reliably find peptides that are unexpectedly modified. De novo sequence information is a valuable first step in finding glycopeptides. This can be conducted by manually interpreting the unidentified MS/MS spectra or using software algorithms such as those included in PEAKS Studio which can elucidate at least some of the peptide sequence, the first step in the characterization of novel glycopeptides.
5. When using trypsin to cleave purified intact proteins in solution, we recommend limiting digestion time to less than 8 h. Typically, almost complete protein digestion can be achieved in 4–6 h. In our experience, longer digestion times result in unpredictable chymotryptic-like activity that complicates data interpretation. In some cases, there may not be sufficient trypsin cleavage sites within a protein sequence to generate peptides that are amenable to tandem mass spectrometry (~5–15 amino acids). Therefore, other proteolytic enzymes may be preferable or used in combination with trypsin to produce peptides of

desirable length and to maximize peptide/glycopeptide identification. Examine the peptide sequence of the protein of interest and perform *in silico* digests (using MassLynx Biolynx or Web-based protein digestion tools such as MS-Digest http://jpsl.ludwig.edu.au/ucsfhtml3.4/msdigest.htm or peptide cutter http://us.expasy.org/tools/peptidecutter/) to select the most appropriate enzyme. To maximize peptide coverage of flagellins from *Clostridium botulinum* we also used endoproteinase GluC. This enzyme was incubated with the flagellin protein at a ratio of 25:1 (protein:enzyme, w/w) in 100 mM phosphate buffer, pH 7.8, at 37°C overnight.

6. A glycopeptide ion may fail to fragment by ETD alone instead of generating mainly charged-reduced precursor ions. It is often possible to generate c and z fragment ions by performing a gentle CAD on the ETD generated ions ("Supplemental Activation" in the LTQ XL control software). In our experience, even better results can often be obtained by performing a gentle CAD (10 V) on an individual charge-reduced glycopeptide ion generated by ETD (a true MS^3 analysis). Production of c and z fragment ions is favored over b and y ions by reducing the Activation Q value from its default of 0.25 to 0.15 (33).

Acknowledgments

We thank Dr Catherine Paul for isolation of *C. botulinum* flagellin proteins and James Mullen for assistance with preliminary mass spectrometry characterizations of flagellin. We also thank Kelly Fulton for assistance with final manuscript preparations.

References

1. Apweiler, R., Hermjakob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8.
2. Drickamer, K. and Taylor, M. E. (1998) Evolving views of protein glycosylation. *Trends Biochem. Sci.* **23**, 321–324.
3. Varki, A. (2006) Nothing in glycobiology makes sense, except in the light of evolution. *Cell* **126**, 841–845.
4. Varki, N. M. and Varki, A. (2007) Diversity in cell surface sialic acid presentations: implications for biology and disease. *Lab. Invest.* **87**, 851–857.
5. Messner, P. (2004) Prokaryotic glycoproteins: unexplored but important. *J. Bacteriol.* 186, 2517–2519.
6. Schmidt, M. A., Riley, L. W., and Benz, I. (2003) Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol.* **11**, 554–561.
7. Logan, S. M. (2006) Flagellar glycosylation – a new component of the motility repertoire? *Microbiology* **152**, 1249–1262.
8. Szymanski, C. M., Logan, S. M., Linton, D., and Wren, B. W. (2003) Campylobacter – a tale of two protein glycosylation systems. *Trends Microbiol.* **11**, 233–238.

9. Schaffer, C. and Messner, P. (2001) Glycobiology of surface layer proteins. *Biochimie* **83**, 591–599.
10. Castric, P., Cassels, F. J., and Carlson, R. W. (2001) Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan. *J. Biol. Chem.* **276**, 26479–26485.
11. Hegge, F. T., Hitchen, P. G., Aas, F. E., Kristiansen, H., Lovold, C., Egge-Jacobsen, W., Panico, M., Leong, W. Y., Bull, V., Virji, M., Morris, H. R., Dell, A., and Koomey, M. (2004) Unique modifications with phosphocholine and phosphoethanolamine define alternate antigenic forms of *Neisseria gonorrhoeae* type IV pili. *Proc. Natl. Acad. Sci. U. S. A* **101**, 10798–10803.
12. Stimson, E., Virji, M., Makepeace, K., Dell, A., Morris, H. R., Payne, G., Saunders, J. -R., Jennings, M. P., Barker, S., Panico, M. et al. (1995) Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol. Microbiol.* **17**, 1201–1214.
13. Schirm, M., Soo, E. C., Aubry, A. J., Austin, J., Thibault, P., and Logan, S. M. (2003) Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* **48**, 1579–1592.
14. Guerry, P., Ewing, C. P., Schirm, M., Lorenzo, M., Kelly, J., Pattarini, D., Majam, G., Thibault, P., and Logan, S. (2006) Changes in flagellin glycosylation affect Campylobacter autoagglutination and virulence. *Mol. Microbiol.* **60**, 299–311.
15. Schirm, M., Schoenhofen, I. C., Logan, S. M., Waldron, K. C., and Thibault, P. (2005) Identification of unusual bacterial glycosylation by tandem mass spectrometry analyses of intact proteins. *Anal. Chem.* **77**, 7774–7782.
16. Schirm, M., Kalmokoff, M., Aubry, A., Thibault, P., Sandoz, M., and Logan, S. M. (2004) Flagellin from *Listeria monocytogenes* is glycosylated with beta-O-linked N-acetylglucosamine. *J. Bacteriol.* **186**, 6721–6727.
17. Schirm, M., Arora, S. K., Verma, A., Vinogradov, E., Thibault, P., Ramphal, R., and Logan, S. M. (2004) Structural and genetic characterization of glycosylation of type a flagellin in *Pseudomonas aeruginosa*. *J. Bacteriol.* **186**, 2523–2531.
18. Verma, A., Schirm, M., Arora, S. K., Thibault, P., Logan, S. M., and Ramphal, R. (2006) Glycosylation of b-Type flagellin of *Pseudomonas aeruginosa*: structural and genetic basis. *J. Bacteriol.* **188**, 4395–4403.
19. Thibault, P., Logan, S. M., Kelly, J. F., Brisson, J. -R., Ewing, C. P., Trust, T. J., and Guerry, P. (2001) Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *J. Biol. Chem.* **276**, 34862–34870.
20. Chaban, B., Voisin, S., Kelly, J., Logan, S. M., and Jarrell, K. F. (2006) Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked glycosylation pathways in Archaea. *Mol. Microbiol.* **61**, 259–268.
21. Voisin, S., Houliston, R. S., Kelly, J., Brisson, J. -R., Watson, D., Bardy, S. L., Jarrell, K. F., and Logan, S. M. (2005) Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. *J. Biol. Chem.* **280**, 16586–16593.
22. Voisin, S., Kus, J. V., Houliston, S., St-Michael, F., Watson, D., Cvitkovitch, D. G., Kelly, J., Brisson, J. -R., and Burrows, L. L. (2007) Glycosylation of *Pseudomonas aeruginosa* strain Pa5196 type IV pilins with mycobacterium-like alpha-1,5-linked d-Araf oligosaccharides. *J. Bacteriol.* **189**, 151–159.
23. Banerjee, A. and Ghosh, S. K. (2003) The role of pilin glycan in neisserial pathogenesis. *Mol. Cell Biochem.* **253**, 179–190.
24. Paul, C. J., Twine, S. M., Tam, K. J., Mullen, J. A., Kelly, J. F., Austin, J. W., and Logan, S. M. (2007) Flagellin diversity in *Clostridium botulinum* groups I and II: a new strategy for strain identification. *Appl. Environ. Microbiol.* **73**, 2963–2975.
25. Twine, S. M., Paul, C. J., Vinogradov, E., McNally, D. J., Brisson, J. -R., Mullen, J. A., McMullin, D. R., Jarrell, H. C., Austin, J. W., Kelly, J. F., and Logan, S. M. (2008) Flagellar glycosylation in *Clostridium botulinum*. *FEBS J.* **275**, 4428–4444.
26. Shevchenko, A., Chernushevich, I., Wilm, M., and Mann, M. (2000) De Novo peptide sequencing by nanoelectrospray tandem mass spectrometry using triple quadrupole and quadrupole/time-of-flight instruments. *Methods Mol. Biol.* **146**, 1–16.
27. Rademaker, G. J., Pergantis, S. A., Blok-Tip, L., Langridge, J. I., Kleen, A., and Thomas-Oates, J. E. (1998) Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity. *Anal. Biochem.* **257**, 149–160.
28. Coon, J. J., Shabanowitz, J., Hunt, D. F., and Syka, J. E. (2005) Electron transfer dissociation of peptide anions. *J. Am. Soc. Mass Spectrom.* **16**, 880–882.

29. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A* **101**, 9528–9533.
30. Rademaker, G. J., Pergantis, S. A., Blok-Tip, L., Langridge, J. I., Kleen, A., and Thomas-Oates, J. E. (1998) Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity. *Anal. Biochem.* **257**, 149–160.
31. Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., Mills, D. C., Watson, D. C., Hernandez, M., Kelly, J. F., Wacker, M., and Aebi, M. (2006) Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25**, 1957–1966.
32. Norris, J. L., Porter, N. A., and Caprioli, R. M. (2003) Mass spectrometry of intracellular and membrane proteins using cleavable detergents. *Anal. Chem.* **75**, 6642–6647.
33. Swaney, D. L., McAlister, G. C., Wirtala, M., Schwartz, J. C., Syka, J. E., and Coon, J. J. (2007) Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal. Chem.* **79**, 477–485.
34. Young, N. M., Brisson, J. -R., Kelly, J., Watson, D. C., Tessier, L., Lanthier, P. H., Jarrell, H. C., Cadotte, N., St, M. F., Aberg, E., and Szymanski, C. M. (2002) Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J. Biol. Chem.* **277**, 42530–42539.

Chapter 9

N-Linked Glycoprotein Analysis Using Dual-Extraction Ultrahigh-Performance Liquid Chromatography and Electrospray Tandem Mass Spectrometry

S.O. Siu, Maggie P.Y. Lam, Edward Lau, William S.B. Yeung, David M. Cox, and Ivan K. Chu

Abstract

Although reverse-phase liquid chromatography (RP-LC) is a common technique for peptide separation in shotgun proteomics and glycoproteomics, it often provides unsatisfactory results for the analysis of glycopeptides and glycans. This bias against glycopeptides makes it difficult to study glycoproteins. By coupling mass spectrometry (MS) with a combination of RP-LC and normal-phase (NP)-LC as an integrated front-end separation system, we demonstrate that effective identification and characterization of both peptides and glycopeptides mixtures, and their constituent glycan structures, can be achieved from a single sample injection event.

Key words: UPLC, 2DLC, glycoproteins, glycans, dual extraction.

1. Introduction

Although liquid chromatography–mass spectrometry (LC-MS) has been successfully implemented as a platform to study tryptic digested proteins and glycoproteins using reverse-phase (RP) and normal-phase (NP)-LC systems (1, 2), it has been restricted to only single types of analytes, i.e., those exhibiting either hydrophilic or hydrophobic properties. The commonly used RP-LC techniques perform rather poorly for the analyses of glycopeptides because the attached glycan motifs render the corresponding glycopeptides relatively hydrophilic. The exact hydrophobicity of

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_9,

a glycopeptide depends on the nature of the glycoform attached as well as the length and composition of the peptide motif. As a result, more-hydrophilic glycopeptides tend not to adsorb on the hydrophobic stationary phase of RP-LC as readily as their more-hydrophobic counterparts. The unsatisfactory separation of these glycopeptides often leads to difficulties in their detection using tandem mass spectrometry (MS/MS) because many glycopeptides co-elute, leading to ion suppression effects and limits in the duty cycle of the instrument. A separation methodology that allows the efficient, concomitant analyses of both peptides and glycopeptides would be highly desirable for the large-scale characterization of glycoproteins in complex samples.

Online coupling of RP and NP ultrahigh-performance liquid chromatography (UPLC) systems has been developed for the separation of hydrophilic and hydrophobic mixtures from a single sample injection event. We have demonstrated that the online integration of RP-LC and NP-LC is a viable approach toward characterizing glycoproteins in high-throughput bottom-up proteomics experiments. Hydrophobic peptides are effectively separated in the RP-LC system and identified through MS/MS (global proteome surveying); online NP-LC separation facilitates the subsequent identification of hydrophilic glycopeptides and provides information on the structure of the constituent glycan motif. The efficacy of the system can be demonstrated using three well-characterized proteins (bovine serum albumin. the glycoproteins horseradish peroxidase, and ribonuclease B) and comparing the results from the dual-extraction RP/NP-LC set-up with those obtained using a conventional single-dimension RP-LC system.

2. Materials

2.1. Sample Preparation

1. Protein standards: bovine serum albumin, ribonuclease B, and horseradish peroxidase (Sigma-Aldrich, St. Louis, MO); used without further purification; stored in powder form at –20°C; and 100 μg dissolved in 100 mM ammonium bicarbonate prior to use.
2. Protease: modified sequencing-grade trypsin (Promega, Madison, WI); stored at –20°C; and reconstituted with trypsin buffer prior to use.
3. Reduction and alkylation: dithiothreitol (DTT) and iodoacetamide (IAA) (Sigma-Aldrich); stored at –20°C; and dissolved in ammonium bicarbonate prior to use.
4. 100 mM ammonium bicarbonate; prepared afresh and stored in a dry place.

2.2. Setting Up the Dual-Extraction UPLC System

1. For solvent preparation: formic acid and trifluoroacetic acid (Fluka, St. Louis, MO); HPLC-grade water and acetonitrile (Tedia, Fairfield, OH).
2. Connecting parts for LC plumbing: fused-silica capillary tubing with polyimide coating [150 μm inner diameter (i.d.) for trap column and analytical columns, 20 μm i.d. for flow splitting], 2-μm stainless-steel screens, four-port two-position switching valves, six-port two-position switching valves (*see* **Note 1**), micro-tees, zero-dead-volume unions, nuts, ferrules, and closed nuts (plugs) from Valco Instruments (Houston, TX); PEEK tubing (360 μm i.d.; Upchurch Scientific, Oak Harbor, WA). When purchasing the connection, it is critical to ensure that all parts have matching fitting specifications.
3. Materials for trap and analytical columns: RP trap column, Jupiter C_{18} (5 μm particles, 300 Å pores; Phenomenex, Torrance, CA); RP analytical column, Jupiter C_{18} (3 μm particles, 300 Å pores; Phenomenex, Torrance, CA); NP trap and analytical columns, TSKgel Amide 80 (5 μm particles, 300 Å pores; Tosoh Bioscience, Montgomeryville, PA).
4. Custom-made mixer: stainless-steel chamber with Valco connections and ca. 1.5 mL space; magnetic stirrer bar; and stirrer.

2.3. Performing the LC Separation and MS/MS Experiment

1. UPLC pump system: Ultrahigh-pressure 65D syringe pump (Isco, Lincoln, NE).
2. Mass spectrometer: Applied Biosystems MDS SCIEX QStar XL (Applied Biosystems, Framingham, MA) (*see* **Note 2**).
3. ProteinPilot (Applied Biosystems).
4. Analyst QS (Applied Biosystems).

3. Methods

Recent advances in UPLC techniques, such as long-column separation, finer packing materials, and internal diameter miniaturization, have greatly advanced the applicability of LC separation in bottom-up proteomics (3–5). The integration of RP-LC and NP-LC depends on the ability of the UPLC system to sustain high pressures and elaborate plumbing. Common solvents (water and acetonitrile) are used in RP-LC and NP-LC; their opposing selectivity for the retention of compounds complements their separation power. To combine the two phases online, the configuration makes use of two trap columns and two analytical columns (one RP and one NP each). The use of trap columns allows

compounds having compatible physical properties to be retained, with non-retained compounds being diverted from the first phase (RP) to the second phase (NP) (6, 7). To achieve this result, a series of mechanical switching valves and micro-tees are used; they also divert the flow of solvents to the desired places. To ensure effective trapping for both phases, it is necessary for the solvent composition to change from highly aqueous to highly organic. This transformation is performed by varying the differential pressure and flow rate between the organic and aqueous channels.

After the samples have been trapped on both trap columns, LC-gradient elution with the online analytical column and MS/MS experiments is performed sequentially, with RP-LC followed by NP-LC. The non-glycosylated peptides are typically adsorbed onto and eluted from the RP column. Glycosylated peptides, especially those having short-peptide sequences (*see* **Note 3**), are found predominantly in the NP-LC analysis.

3.1. Sample Preparation

1. The protein standards are dissolved in 100 mM ammonium bicarbonate and combined in a microfuge tube.
2. 50 mM DTT is added to the sample, which is mixed by vortexing.
3. The sample is then incubated for 60 min at 60°C to reduce the disulfide bonds in the proteins.
4. The sample is cooled to room temperature (25°C).
5. The cysteine residues are then irreversibly blocked through incubation with 100 mM IAA in the dark at room temperature for 60 min.
6. The modified sequencing-grade trypsin is reconstituted at 1 μg/μL in the accompanying buffer. The reconstituted trypsin (3 μg) is added to the sample, which is then incubated at 37°C overnight (*see* **Note 4**).
7. Trypsin digestion is terminated by heating at 90°C on a heat plate for 10 min. When stored at –20°C, the sample can be used for several weeks.

3.2. Setting Up the Dual-Extraction UPLC System

1. Manually pack the RP and NP columns using appropriate materials (8). Trap column dimensions: 150 μm i.d., 40 mm length. Analytical columns: 150 μm i.d., 800 mm length. Long columns improve the LC dynamic range and the effective MS sensitivity.
2. The UPLC plumbing is set up as shown in **Fig. 9.1**. The system consists of nine switching valves and three micro-tees, connected using silica capillaries (*see* **Note 5**).
3. V1, V3, V5, V6, V7, and V9 are four-port, two-position valves. V2, V4, and V8 are six-port, two-position valves.

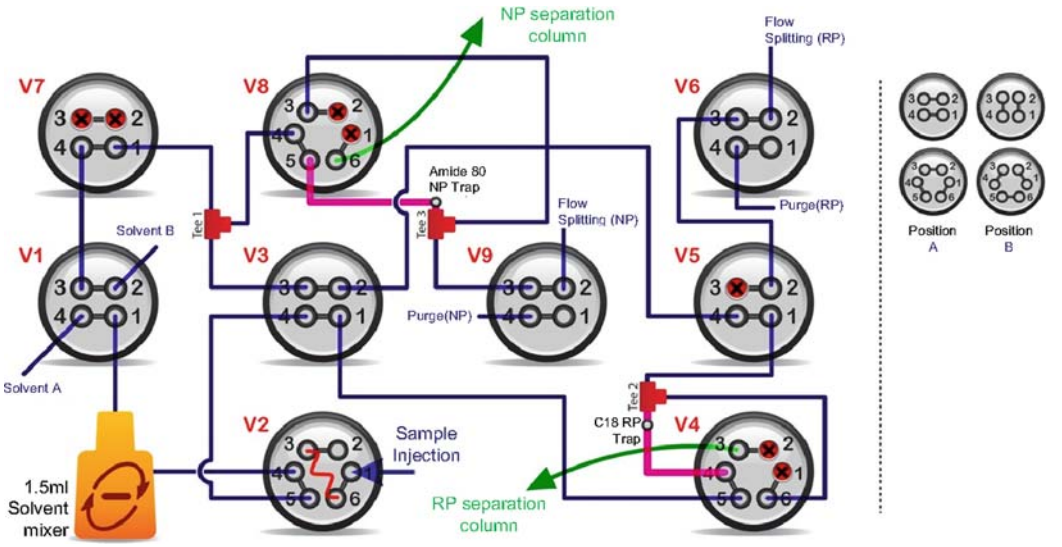


Fig. 9.1. Schematic representation of the set-up of the UPLC switches for RP/NP trapping and flow diversion.

T1, T2, and T3 are stainless-steel micro-tees for three-way splitting (*see* **Note 6**).

4. Pump A is filled with aqueous solvent [solvent A: 2% acetonitrile, 0.5% formic acid (v/v)]. Pump B is filled with organic solvent [solvent B: 98% acetonitrile, 0.5% formic acid (v/v)].
5. Connect the outlets of pumps A and B to ports 4 and 2 of V1, respectively.
6. Connect port 1 of S1 to the 1.5-mL mixer equipped with the magnetic stirrer bar and then connect port 3 of V1 to port 4 of V7. Stop ports 2 and 3 of V7 with plugs.
7. Connect the outlet of the mixer to port 4 of the six-port valve V2. Then connect ports 3 and 6 of V2 using a (5 μL) sample loop. Port 1 of V2 serves as the needle port.
8. Connect port 5 of V2 to port 4 of V3. Port 2 of V2 is left open as an outlet for excessive amounts of sample.
9. Connect port 1 of V3 to port 5 of V4. Port 4 of V4 is connected directly to the RP trap column, with a stainless-steel screen placed in the port to prevent the escape of the column material back into the switching valve channel.
10. The micro-tee T2 is placed along with another stainless-steel screen at the other end of the trap column; the other two ends of T1 should connect to port 6 of V4 and port 1 of V5. The purpose of the micro-tee is to allow "backward flushing" of the trapped sample into the RP analytical column, which is connected directly without extra capillaries to port 3 of V4. Ports 1 and 2 of V4 are stopped by plugs.

11. Connect port 2 of V5 to port 3 of V6, which serves as a flow rate control by controlling the flow of solvent to two outlet capillaries having different inner diameters. Port 4 of V6 is connected to an open stainless-steel capillary (1500 μm i.d.) for purging; whereas port 2 is connected to a restrictive 20-μm-i.d., 6-m-long silica capillary. Port 1 of V6 is left open.
12. Block port 3 of V5 with a plug and connect port 4 of V5 back to port 2 of V3 to complete the RP portion of the set-up. Connect port 3 of V3 to the micro-tee T1.
13. Connect the other two ends of T1 to port 1 of V7 and port 4 of the six-port valve V8. V8 functions for the NP portion of the set-up similar to V4 for the RP portion.
14. Connect port 5 to a stainless-steel screen and the NP trapping column.
15. Place a stainless-steel screen on and connect the micro-tee T3 to the other end of the NP trap column.
16. Connect the other two ends of T3 to port 3 of V8 and port 3 of V9, which serves as a flow control similar to V6. Ports 1 and 2 of V8 are blocked with plugs. Connect port 6 of V8 to the NP analytical column.
17. Port 4 of V9 is connected to an open stainless-steel capillary (1500 μm i.d.); port 2 is connected to a 20-μm-i.d., 6-m-long silica capillary. Port 1 is left open.
18. Connect the RP and NP analytical columns to the mass spectrometer (*see* **Note 7**). *See* **Figs. 9.2 and 9.3**.

3.3. Performing the LC Separation and MS/MS Experiment

1. Equilibrate the system and columns using the appropriate solvent. The goal of equilibration is to prime the trap and analytical columns for sample introduction and also to purge the mixer before executing the RP gradient.
2. During equilibration, V1 should be turned to position A to allow solvent A into the mixer, the RP trap, and the analytical column.
3. V2 and V3 should be set to position A; V4, V5, and V6 should be turned to position B to direct the flow through the RP set-up to the high-flow-rate splitting capillary.
4. Likewise, for NP separation, V7 should be at position A and V8 and V9 should be at position B. Run both pumps at 5000 PSI until ca. 5–10 mL of each solvent has flown through the system.
5. Turn V2 to position A; then, using a syringe, inject 5 μL of the sample to be analyzed into the sample loop at V2.
6. The pumps can then be turned on for sample trapping.

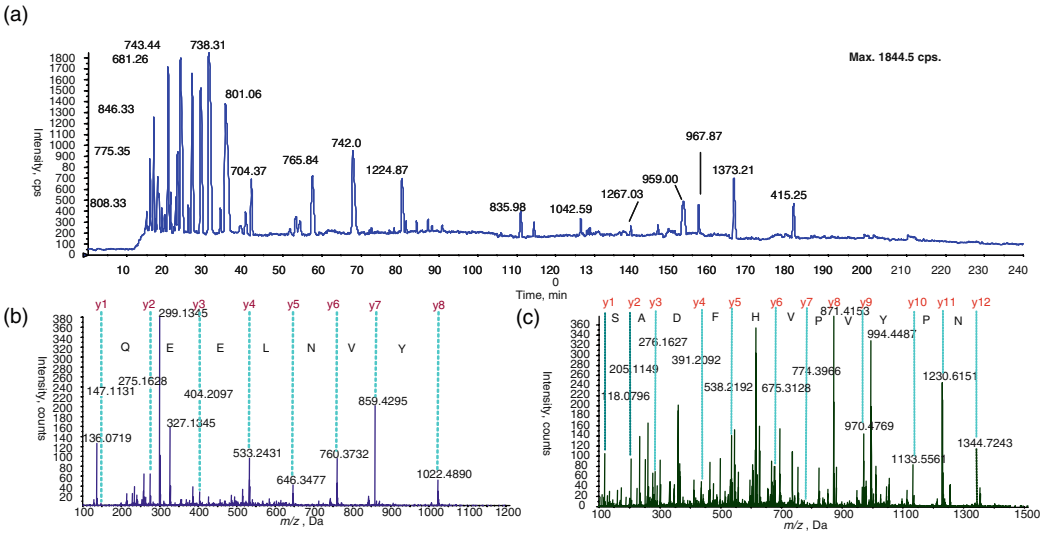


Fig. 9.2. Typical spectra acquired in the RP-LC portion of the experiment: (a) the base peak chromatograph shows clear separation of peptides in the RP column; (b) MS/MS spectrum of the ion at *m/z* 593.28 at 23.00 min; and (c) MS/MS spectrum of the ion at *m*/z 740.03 at 67.98 min. Most unmodified peptides are retained in the RP portion of the experiment; some glycopeptides possessing long peptide sequences, and which are, therefore, relatively more hydrophobic, are also detected.

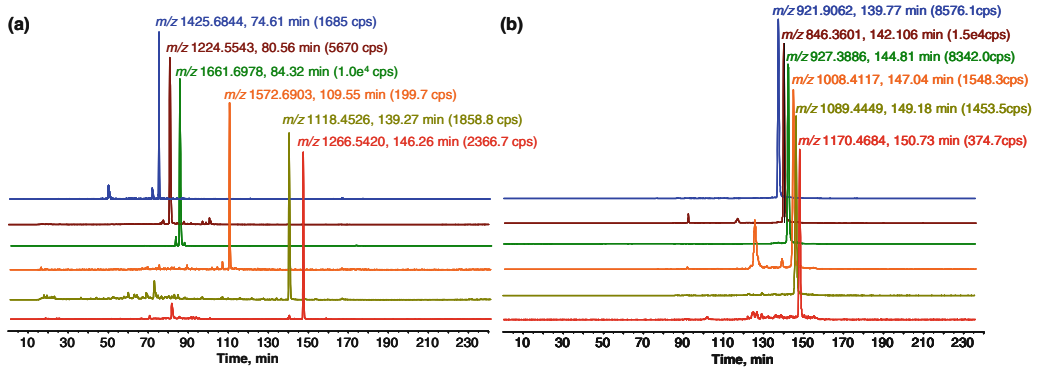


Fig. 9.3. Extracted ion chromatographs (XICs) of (a) ions of the glycopeptides from the tryptic digests of ribonuclease B and peroxidase in RP-LC; (b) ions at *m/z* 921.91, 846.36, 927.39, 1008.41, 1089.44, and 1170.47 corresponding to extra glycopeptides identified in the NP-LC portion of the experiment that eluded RP-LC trapping. In total, 58 unique, non-glycosylated tryptic peptides were identified from the product ion spectra using the dual-extraction RP/NP-LC set-up; RP-LC analysis identified 50 of the peptides, whereas the remaining eight were not retained by the RP analysis column, but were captured by the NP trap column. In addition, a total of 12 glycosylated tryptic peptides were identified, six of which were detected only in the NP portion of the experiment.

7. During dual-extraction sample trapping, V1, V3, V4, V5, V7, V8, and V9 are all set at position A; V6 is unused. For trapping purposes, operate pumps A and B at constant pressures of 7000 and 6000 psi, respectively, for 20 min to allow enough time for the injected samples to be flushed into the trap columns.

8. Ready the mass spectrometer for RP-LC–MS/MS analysis. During the RP-LC experiment, set V2, V3, and V6 to position A; V4, V5, and V7 to position B. The solvent gradient will pass through the RP trap and analytical column at V4 and any excessive volume of solvent will be diverted at V5 away from the NP portion to the outlet control valve V6. Operate both pumps at 8000 psi.
9. Initially, the mixer is filled with solvent A (from the equilibration in Step 1).
10. When the experiment starts, turn V1 to position B to stop the flow of solvent A and gradually introduce solvent B to the mixer. This process will begin the formation of the gradient. MS acquisition should be started concurrently (*see* **Note 8**).
11. When the RP-LC–MS/MS experiment is complete, the NP-LC–MS/MS experiment can be performed.
12. For the NP gradient, make sure that the mixer is filled with solvent B prior to beginning the experiment; set V2 and V9 to position A and V3, V7, and V8 to position B. At the start

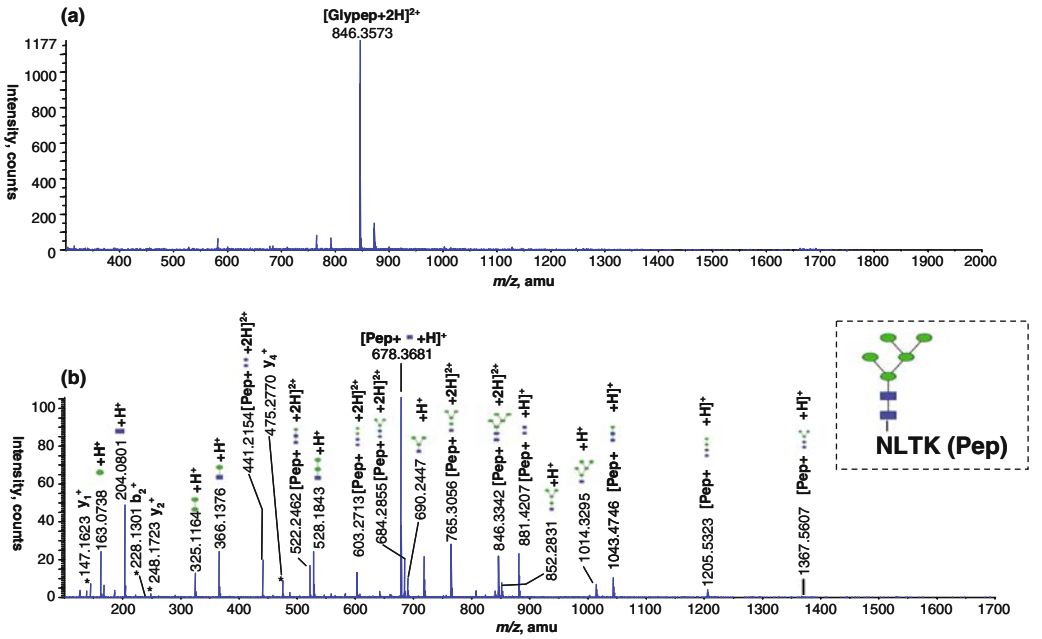


Fig. 9.4. Typical spectra obtained from the NP-LC portion of the dual-extraction experiment. (a) Survey scan spectrum at 142.11 min; (b) labeled product-ion spectrum of the ion at *m/z* 846.36, showing b and y fragment ions of the glycopeptide as well as the attached *N*-acetylglucosamine and mannose moieties. Compared with a single NP-LC experiment, the spectra of the NP portion of the dual-extraction experiment are marked by their reduced complexity and higher signal intensities, which could be attributed to the RP trap column's removal of potentially co-eluting peptides. A large proportion of the peptides and glycopeptides are retained only by a single phase, effectively reducing the sample complexity and the accompanying ion suppression effect.

of the experiment, V1 is turned from position B to position A to stop the flow of solvent B and to gradually introduce solvent A into the mixer. Similar MS/MS parameters as in the RP experiments are used.

13. After performing both the RP-LC and NP-LC experiments, the acquired spectra are analyzed to determine the identity of the peptides and glycopeptides detected through MS/MS.
14. The software ProteinPilot will automatically convert the MS/MS data into peak lists and then search (using the Paragon algorithm) against a database (*see* **Note 9**).
15. Choose the instrument type as QSTAR ESI in proteinPilot, the peptide mass and product ion tolerance are automatically set. The possible compositions of the N-glycan can be determined using the free software GlycoMod (http://www.expasy.org/tools/glycomod/) (*see* **Fig. 9.4**).

4. Notes

1. Manual valves are used in the experiment. With appropriate software, electronic valves can also be used for automation in switching.
2. The QStar XL is a hybrid quadrupole time-of-flight instrument and provides good sensitivity and resolution for shotgun proteomics. Other compatible instruments can also be used, but the MS parameters listed herein have been chosen specifically for the QStar XL.
3. The exact hydrophobicity of a glycopeptide depends on its peptide sequence, length, and glycan composition. Glycopeptides with long peptide sequences, making them relatively hydrophobic, can be found in the RP column. By the same token, short non-glycosylated peptides that are especially hydrophilic, such as those containing multiple serine and threonine residues, can be found in the NP column.
4. Trypsin digestion usually takes 6–16 h. Ensure good insulation during incubation to avoid the temperature difference on the top part of the microfuge tube causing any condensation. Trypsin digestion can be quenched with the addition of 5% formic acid at the end of the incubation process.
5. The set-up should be installed on a solid support, such as a wooden board, upon which the switching valves can be fixed. All the connections are formed using capillary tubing,

with PEEK sleeves, ferrules, and nuts at both ends. Unions connect two capillaries end-to-end. Unless otherwise specified, the capillaries connecting the valves and tees in this experiment are 75 μm i.d. and ca. 200 mm long. Sufficient length should be given to allow flexible connection, while avoiding excessive dead volume.

6. For all two-position valves, position A denotes the default channel (port 1 leads to port 6) whereas position B denotes the alternative channel (port 1 leads to port 2). V1 functions to control the flow of solvent to the 1.5-mL mixer; V2 is used for sample injection; V3 controls the flow of solvent to either the RP or NP analytical column; V4 controls the trapping and elution of the samples in the RP column; V5 serves to divert the solvent flow from the RP to NP column; V6, which controls the i.d. of the splitting capillaries, is used to control the flow rate during the conditioning and experimental phases (described in more detail in the following sections); V7 allows the flow of solvent B to be stopped; V8 controls the trapping and elution of samples in the NP column; V9 swaps the outlet of the splitting capillary to determine the flow rate of pump B under constant pressure.
7. Cap the other end of the analytical columns with a stainless-steel screen and connect through a union to an electrospray emitter. Two separate emitters for the RP and NP columns can be used to mount them separately onto the mass spectrometer's nanospray source, in which case the columns must be switched manually between the RP and NP experiments. Alternatively, the two columns can be connected through switches and plugs to a common emitter; note, however, that the increase in post-column dead volume may lead to peak broadening and a decrease in separation performance.
8. Depending on the actual flow rate, the LC gradient should last ca. 3.5 h; the MS acquisition should be set to run for the same length of time. In our experiment, the mass spectrometer parameters were set as follows: DP, 20 V; FP, 125 V; CUR, 20; GSI, 0; GS2, 0; and nanospray voltage, 3000 V. The mass spectrometer was operated under information-dependent acquisition. Each 1-s survey scan was followed by three 1-s product ion scans of the three most intense peaks having higher than 50 counts and with the charge state within the range of 2–5. Enhanced ion settings, 20-s dynamic exclusion, and rolling collision energy were used; the calibration curve for the collision energy was optimized using direct BSA infusion.

9. A public database, such as Swiss-Prot, can be used. Swiss-Prot can be downloaded from the UniProt website (http://www.uniprot.org). Alternatively, a custom database, containing just the sequences of the proteins in the experiment, can be used.

Acknowledgments

This study was supported by the Hong Kong Research Grants Council (Project No. HKU7640/07 M), Hong Kong Special Administrative Region, China.

References

1. Harvey, D. J. (2001) Identification of protein-bound carbohydrates by mass spectrometry. *Proteomics* **1 ,** 311–328.
2. Dell, A. and Morris, H. R. (2001) Glycoprotein structure determination by mass spectrometry. *Science* **291**, 2351–2356.
3. Shen, Y., Moore, R. J., Zhao, R., Blonder, J., Auberry, D. L., Masselon, C., Pasa-Tolic, L., Hixson, K. K., Auberry, K. J., and Smith, R. D. (2003) High-efficiency on-line solid-phase extraction coupling to 15–150-μm-i.d. Column liquid chromatography for proteomic analysis. *Anal. Chem.* **75**, 3596–3605.
4. Shen, Y., Tolic, N., Masselon, C., Pasa-Tolic, L., Camp, D. G., Hixson, K. K., Zhao, R., Anderson, G. A., and Smith, R. D. (2004) Ultrasensitive proteomics using high-efficiency on-line micro-SPE-NanoLC-NanoESI MS and MS/MS. *Anal. Chem.* **76**, 144–154.
5. Shen, Y. F., Zhao, R., Berger, S. J., Anderson, G. A., Rodriguez, N., and Smith, R. D. (2002) High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. *Anal. Chem.* **74**, 4235–4249.
6. Lee, H., Yi, E. C., Wen, B., Reily, T. P., Pohl, L., Nelson, S., Aebersold, R., and Goodlett, D. R. (2004) Optimization of reversed-phase microcapillary liquid chromatography for quantitative proteomics. *J. Chromatogr. B* **803**, 101–110.
7. Meiring, H. D., van der Heeft, E., ten Hove, G. J., and de Jong, A. P. J. M. (2002) Nanoscale LC-MS: technical design and applications to peptide and protein analysis. *J. Sep. Sci.* **25**, 557–568.
8. Shen, Y., Zhao, R., Belov, M. E., Conrads, T. P., Anderson, G. A., Tang, K., Pasa-Tolic, L., Veenstra, T. D., Lipton, M. S., Udseth, H. R., and Smith, R. D. (2001) Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Anal. Chem.* **73**, 1766–1775.

Chapter 10

Microarray-Based Study of Carbohydrate–Protein Binding

Zhenxin Wang and Jingqing Gao

Abstract

To develop a novel high-throughput tool for monitoring carbohydrate–protein interactions, carbohydrate or glycoprotein microarrays have been prepared for binding with lectins. The interaction events are marked by attachment of fluorescent dyes and gold nanoparticles. The attachment of the fluorescent dyes and gold nanoparticles is achieved by standard avidin–biotin chemistry. The detection principle is fluorescence or resonance light scattering (RLS). The electroless deposition of silver onto the gold particles has been employed for RLS signal enhancement. Well-defined recognition systems, three monosaccharides (Man-α, Glc-β, and Gal-β) or three glycoproteins (Asf, RNase A, and RNase B) with two lectins (ConA and RCA120), are chosen here to establish the microarray-based assay, respectively. Highly selective recognition of carbohydrate–protein down to 25.6 pg/mL for RCA120 in solution and 8 μM for Gal-β and 32 ng/mL for Asf on the microarray spots is demonstrated.

Key words: Glycan microarray, fluorescence, resonance light scattering, gold nanoparticle.

1. Introduction

DNA microarray has become essential high-throughput tool for gene discovery and mapping, gene regulation studies, disease diagnosis, drug discovery, and toxicology since the middle of 1990s (1). However, the development of similar devices for the analysis of saccharides has been hindered, primarily due to nature of these molecules (e.g., cannot be amplified by PCR and synthesized by automatic solid-phase synthesis) (2). On the other hand, understanding of the molecular basis for carbohydrate–protein interactions not only provides valuable information on biological processes in living organisms, but also aids in the development of potent biomedical agents (3, 4). It is only recently that

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_10,

glycan microarrays are developed and applied in glycomics research (2, 5–8).

Currently, these microarray techniques mainly rely on the use of fluorescent molecular dye labels which have several potential drawbacks, such as the need of amplification, i.e., the use of PCR in DNA diagnostics due to lack of sensitivity, and photoinstability of the dyes employed (1, 9). After reactions, the system is normally detected by double wavelengths laser confocal microarray scanner which also can lead to reduce experimental reproducibility because of non-uniform laser photobleaching of fluorescent labels. Metal, semiconductor, and magnetic nanoparticles can offer a unique set of physical properties that may be exploited in biological detection assays as an alternative to the use of fluorescent dyes (10, 11). In particular, the detection of resonance light scattering by metal particles represents a great step forward, toward higher sensitivity, with the eventual goal of detecting single biomolecular binding events (9, 10). As a pioneer, Mirkin and coworkers have developed a DNA-modified gold nanoparticle-based assay for highly sensitive and selective detection of DNAs and proteins on microarrays (9, 10).

Here, an RLS-based microarray format with high sensitivity and selectivity has been developed for detection of carbohydrate–lectin or glycoprotein–lectin interactions. As in previous report (12), this method is based on labeling the recognition events on a microarray with gold nanoparticles using avidin–biotin chemistry followed by silver enhancement and RLS detection (*see* **Note 1**).

2. Materials

1. Hydrogen tetrachloroaurate trihydrate ($HAuCl_4 \cdot 3H_2O$) and sodium citrate (Sigma-Aldrich).
2. Peptides CALNN and CALNNGK (biotin) G Scilight Biotechnology Ltd. (Beijing, China).
3. PBS buffer: 50 mM phosphate buffer, 0.15 M NaCl, pH 7.5, stored at 4°C.
4. Avidin-FITC solution: 50 μg/mL avidin-FITC in 50 mM phosphate buffer containing 0.15 M NaCl and 1% BSA (w/w).
5. 4-Aminophenyl α-D-mannopyranoside (Man-α), 4-Aminophenyl β-D-galactopyranoside (Gal-β) and 4-Aminophenyl β-D-glucopyranoside (Glc-β), Ribonuclease B (RNase B), Ribonuclease A (RNase A) and Asialofetuin (Asf), ethanolamine, silver enhancer, and bovine serum albumin (BSA) (Sigma-Aldrich).

6. Aldehyde and epoxide-modified glass microscope slides (CapitalBio, Beijing, China).
7. Biotinylated Concanavalin A from *Canavalia ensiformis* (ConA-biotin) and fluorescein isothiocyanate-labeled avidin (avidin-FITC) (Sigma-Aldrich). Biotinylated *Ricinus communis* agglutinin (RCA120-biotin) (Vector Laboratories Ltd., Burlingame, CA, USA).
8. Spotting buffer: 0.3 M phosphate buffer, 0.15 M NaCl, pH 8.5, supplemented with 0.005% (v/v) Tween-20.
9. Washing buffer: 50 mM phosphate buffer, pH 7.5, supplemented with 0.05% (v/v) Tween-20.
10. Blocking buffer: 50 mM phosphate buffer, 0.15 M NaCl, pH 7.5, supplemented with 1% (w/w) BSA and 0.1 M ethanolamine.
11. Probe buffer: 50 mM phosphate buffer, 0.15 M NaCl, pH 7.5 supplemented with 1% BSA.
12. Other chemicals were of analytical grade and used as received. Milli-Q water (18.2 MΩ.cm) was used in all experiments.

3. Methods

3.1. Preparation of Gold Nanoparticle Probes

The citrate-stabilized 13 nm gold nanoparticles were synthesized by traditional Frens–Turkevich method (13, 14). Briefly, 4 mL of 1% sodium citrate aqueous solution was added to 100 mL of 0.01% boiling $HAuCl_4$ solution. Then, the reaction mixture was refluxed for 30 min. The color of the suspension will change from a dark blue to a deep wine red as the monodisperse colloidal gold particle is formed (*see* **Note 2**). Peptide-stabilized nanoparticles were prepared by our previously reported peptide capping procedure (15). Generally, desired amount of aqueous solution of peptide mixture (peptides CALNN and CALNNGK (biotin) G) was added to the solution of 2.7 nM 13 nm gold nanoparticles to give a final concentration of total peptides of 1.5 mM. The ratio of CALNN and CALNNGK (biotin) G in the mixture is 9:1. Excess peptides were removed by repeated centrifugation at 13,000 rpm, ($\sim$16,100$\times g$, 3 times) using an Eppendorf centrifuge. After centrifugation, the purified gold nanoparticles were re-suspended in PBS and stored at 4°C.

3.2. Carbohydrate Microarrays Fabrication and Binding with Lectins

Multiple copies of three amino-modified monosaccharides, Man-α, Gal-β, and Glc-β, with desired concentration were dissolved in 15 μL spotting buffer with 15% (v/v) glycerol and spotted on the

aldehyde-modified glass microscope slides by a standard robotic procedure (*see* **Note 3**). After an overnight incubation under 75% humidity at 25°C, the slides were rinsed with 50 mL washing buffer (3 times) and then immersed in 10 mL blocking buffer at 30°C for 1 h to remove remaining free aldehyde groups. Then, each array was reacted with biotin-modified lectins, ConA-biotin and RCA120-biotin, which were diluted to the desired concentration with 20 μL probe buffer. Following 1 h incubation at 25°C, the slides were subjected to a series of rinsing steps: (1) 50 mL probe buffer for 3 min (3 times); (2) 50 mL washing buffer for 3 min (3 times); (3) 50 mL PBS buffer for 3 min (3 times); and (4) 50 mL Milli-Q water for 3 min (3 times), then dried by centrifugation (480×*g* for 1 min). Through this reaction, together with the lectin, a biotin site was transferred to the carbohydrate on the microarray spot. After recognition of the immobilized monosaccharides by biotinylated lectins, avidin-conjugated fluorescein isothiocyanate (avidin-FITC) was used to label the recognition event (*see* **Note 4**). At this step, the microarrays were ready for fluorescence detection. In a subsequent step, avidin-FITC was reacted with peptide-stabilized gold nanoparticles followed by the conventional silver enhancement step and finally detection by RLS (*see* **Note 5**).

3.2.1. Preparation of Assays for Fluorescence Detection

1. The monosaccharides, Man-α, Gal-β, and Glc-β, on the microarray spot were recognized by biotinylated lectins, RCA120 and ConA, respectively (*see* **Note 6**).
2. Then array was incubated with 20 μL avidin-FITC solution for 1 h at 37°C.
3. The arrays were subjected to series rinsing steps: (1) 50 mL washing buffer for 3 min (3 times); (2) 50 mL PBS buffer for 3 min (3 times); and (3) 50 mL Milli-Q water for 3 min (3 times), and dried by centrifugation (480×*g* for 1 min).
4. Upon labeled by avidin-FITC, the slides were imaged with a conventional fluorescence microarray scanner (Luxscan-10 K/A, CapitalBio Ltd., China) and analyzed with the Array Analyser software (*see* **Note 7**). Fluorescent images and corresponding analysis of peptide microarrays are shown in **Fig. 10.1a** and **1b**.

3.2.2. Preparation of Assays for RLS Detection

1. Following Step 2 of **Section 3.2.1**, the array was treated with 20 μL of a solution of peptide-stabilized gold nanoparticles (2.5 nM) in probe buffer for 1 h at 37°C, then washed and dried as described in Step 3 of **Section 3.2.1**.
2. One milliliter of silver enhancement reagent was applied to each slide (*see* **Note 8**).
3. Upon signal amplification by silver deposition, the slides were imaged with ArrayIt SpotWare Colorimetric

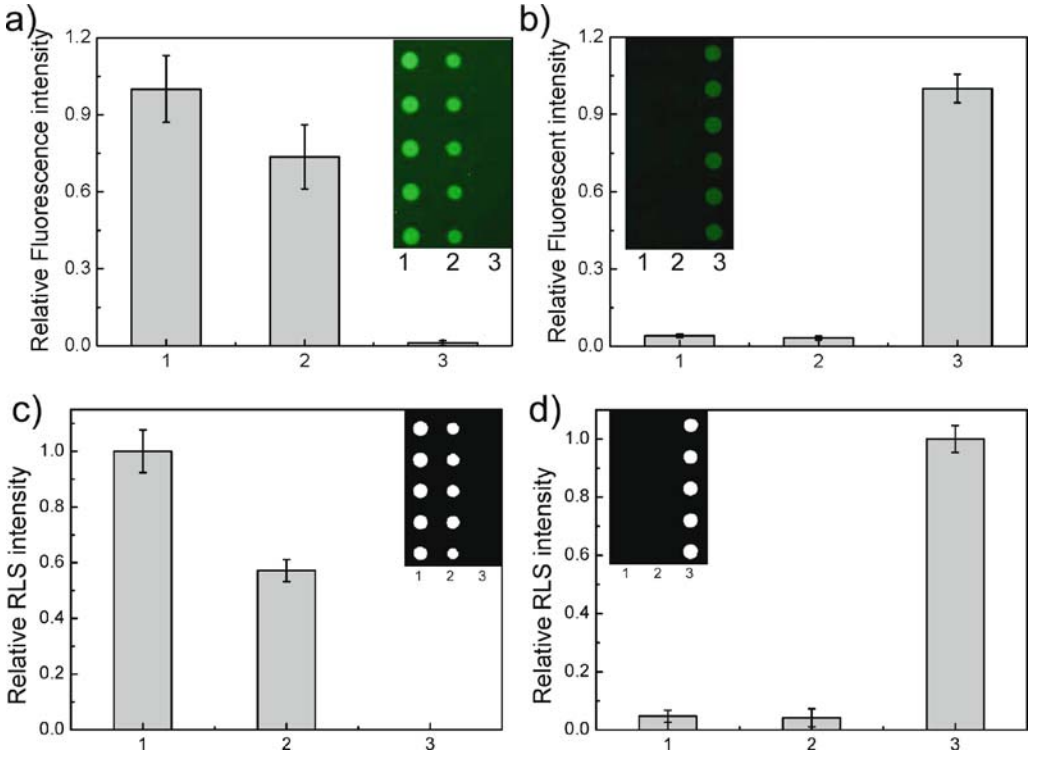


Fig. 10.1. Fluorescent (a and b) and RLS (c and d) images and corresponding analysis of monosaccharide microarrays after probing with lectins, labeling positives with gold nanoparticles, and enhancement by silver deposition. Man-α was spotted on column 1, Glc-β was spotted on column 2, and Gal-β was spotted on column 3. Each column contains five replicate spots. The microarrays were probed with ConA-biotin (a and c) and RCA120-biotin (b and d). The concentration of spotted monosaccharide is 50 mM and probe lectin is 50 μg/mL, respectively. (Reproduced from (8) with permission from ACS.)

Microarray Scanner (TeleChem International, Inc., USA) using white light and analyzed with the Array Analyser software. Light scattering images and corresponding analysis are shown in **Fig. 10.1 c** and **1d**.

3.3. Glycoprotein Microarrays Fabrication and Detection

Multiple copies of RNase B, RNase A, Asf, and BSA were spotted on epoxide functionalized slides in PBS spotting buffer with 40% glycerol (v/v) included to prevent evaporation of the nanodroplets (*see* **Note 9**). After an overnight reaction at 25°C in a vacuum environment, the slides were washed, blocked, and reacted with biotinylated lectins, ConA-biotin and RCA120-biotin, as previously described (*see* **Note 10**). Then array was subjected to label with 20 μL avidin-FITC solution, treated with 20 μL 2.5 nM peptide-stabilized gold nanoparticles, silver enhanced, and detected as described in **Section 3.2**. Corresponding images and analysis are shown in **Fig. 10.2** (*see* **Note 11**).

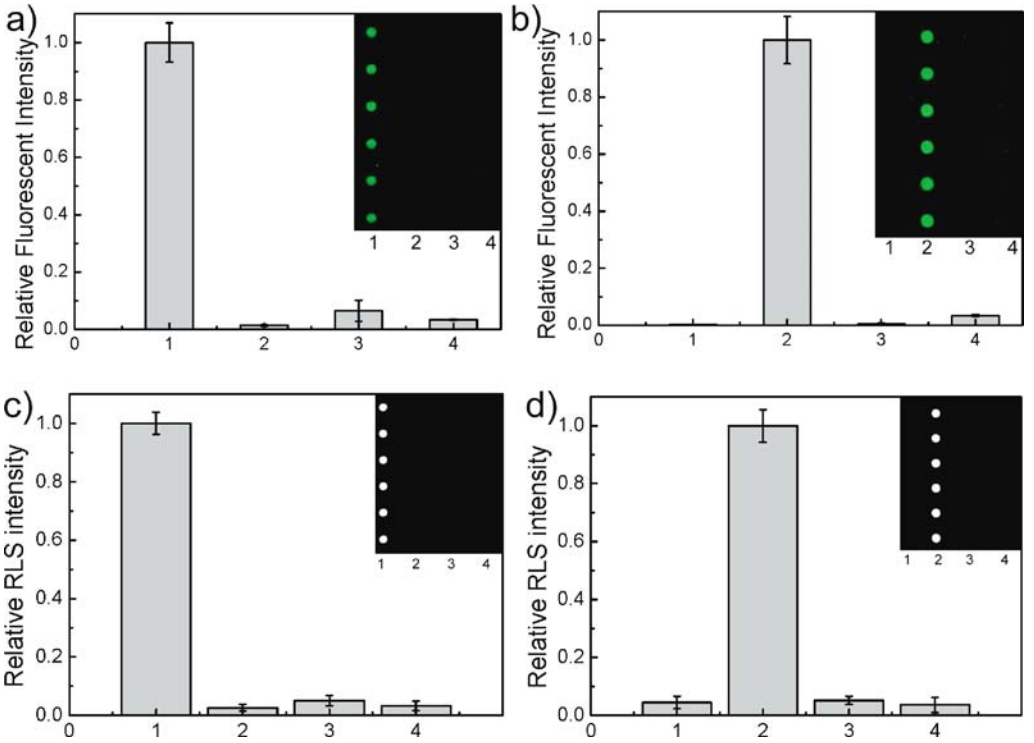


Fig. 10.2. Fluorescent (a and b) and RLS (c and d) images and corresponding analysis of glycoprotein microarrays after probing with lectins, labeling positives with gold nanoparticles, and enhancement by silver deposition. RNase B was spotted on column 1, Asf was spotted on column 2, RNase A was spotted on column 3, and BSA was spotted on column 4. Each column contains six replicate spots. The microarrays were probed with ConA-biotin (a and c) and RCA120-biotin (b and d). The concentration of spotted glycoprotein is 0.5 mg/mL and probe lectin is 50 μg/mL, respectively. (Reproduced from (8) with permission from ACS.)

4. Notes

1. Gold nanoparticles exhibit the ability to resonantly scatter visible and near-infrared light (generally named as resonance light scattering, RLS). This property is the result of the excitation of surface plasmon resonances (SPR) and is extremely sensitive to the size, shape, and aggregation state of the particles. The RLS of small gold nanoparticles (<40 nm in diameter) can be enhanced after silver deposition.
2. The sodium citrate solution should be freshly prepared. The maximum absorption band of the as-prepared gold nanoparticles is about 518 nm. The average size of the gold nanoparticles is 13 ± 1.8 nm in diameter.

3. The aldehyde groups on the glass surface react readily with the primary amines of the amino-modified monosaccharides to form a Schiff base linkage.
4. Biotin (also known as vitamin H) is a small organic molecule found in every cell. Avidin ($pI = 10$) or streptavidin (*Streptomyces avidinii*, $pI = 6$) is a much larger protein (molecular weight is more than 60 kDa) that has four high-affinity biotin-binding sites (Ka = 10^{15} mol/L).
5. We have found that the detection sensitivity and dynamic range are critically related to the amount of silver deposited. No spots can be detected when exposure time is less than 1 min. For short exposure times (2–4 min), there was poor sensitivity, while long exposure times (>15 min) had significant background noise and saturated signal output which also reduced sensitivity and dynamic range. With the optimum exposure time of ~8 min, good features of the spots and a relatively higher signal-to-background noise ratio can be achieved.
6. RCA120 binds only to Gal-β, while ConA binds to both Man-α and Glc-β. In addition, the affinity of ConA with Man-α is stronger than that of ConA with Glc-β.
7. The background originating from the slide was recorded and subtracted from each image prior to evaluation. The mean value and standard deviation of the signal were determined for the 25 or 36 spot replicates per sample, respectively. The detection limit was determined to be the concentration where signal/standard deviation = 3 was reached. For the determination of the linear ranges of the curves, the range of concentrations that best fitted the linear equation $y = mx + b$ was specified.
8. For comparison, the detection limits and dynamic ranges of the assays are summarized in **Table 10.1**. Generally, the detection limit and dynamic range of the RLS assay are better than that of fluorescent assay. These results indicate that the prospects for clinical application of our approach in monitoring lectins response to immunization are extremely promising since a natural immune response typically yields specific lectin concentrations over 2 ng/mL.
9. The primary amines on the glycoprotein surface act as nucleophiles, attacking epoxy groups and coupling the protein covalently to the surface.
10. In this case, RCA120 binds to Asf, while ConA binds to RNase B.
11. For further comparison, the detection limits and dynamic ranges are summarized in **Table 10.2**. These results

Table 10.1
The detection limits and dynamic ranges of the RLS or Fluorescence assay with monosaccharides microarray format

	RLS assay		Fluorescence assay	
Analytes	**Detection limit**	**Dynamic range**	**Detection limit**	**Dynamic range**
Man-α[a]	100 μM	0.5–50 mM	500 μM	0.5–50 mM
ConA[b]	100 pg/mL	0.5–500 ng/mL	1 μg/mL	1–50 μg/mL
Gal-β[a]	8 μM	0.2–100 mM	200 μM	1–50 mM
RCA120[b]	100 pg/mL	400 pg/mL–1.5 μg/mL	300 ng/mL	0.3–100 μg/mL

[a]The detection limit and dynamic range of monosaccharide (Man-α or Gal-β) are based on the initial concentration of spotting solution and the concentration of the lectin (RCA120 or ConA) in the probe solution is 1.5 μg/mL.
[b]The concentration of monosaccharide (Man-α or Gal-β) is 25 mM in spotted solution.

Table 10.2
The detection limits and dynamic ranges of the RLS or fluorescence assay with glycoprotein microarray format

	RLS assay		Fluorescence assay	
Analytes	**Detection limit**	**Dynamic range**	**Detection limit**	**Dynamic range**
Asf[a]	32 ng/mL	0.16–500 μg/mL	4 μg/mL	0.02–1 mg/mL
RCA120[b]	25.6 pg/mL	0.003–10 μg/mL	400 ng/mL	2–100 μg/mL
RNase B[a]	0.16 μg/mL	0.8–1000 μg/mL	0.8 μg/mL	4–500 μg/mL
ConA[b]	0.5 ng/mL	1–1000 ng/mL	100 ng/mL	0.5–10 μg/mL

[a]The detection limit and dynamic range of glycoprotein (Asf or RNase B) are based on the initial concentration of spotting solution and the concentration of the lectin (RCA120 or ConA) in the probe solution is 2 μg/mL.
[b]The concentration of glycoprotein (Asf or RNase B) is 250 μg/mL in spotted solution.

indicate that our method could be used to detect glycoproteins in various body fluids which typically yield specific glycoprotein (e.g., RNase B) concentrations ranging from approximately 0.1 μg/mL to over 1 μg/mL.

Acknowledgments

The author would like to thank the NSFC (Grant No. 20675080 and 20875087) for financial support.

References

1. Schena, M. (2003) *Microarray Analysis*. Hoboken, NJ: Wiley-Liss.
2. Wang, D., Liu, S., Trummer, B. J., Deng, C., and Wang, A. (2002) Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nat. Biotechnol.* **20**, 275–281.
3. Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97–130.
4. Feizi, T. (2000) Progress in deciphering the information content of the 'glycome'-a crescendo in the closing years of the millennium. *Glycoconj. J.* **17**, 553–565.
5. Blixt, O., Han, S., Liao, L., Zeng, Y., Hoffmann, J., Futakawa, S., and Paulson, J. C. (2008) Sialoside analogue arrays for rapid identification of high affinity siglec ligands. *J. Am. Chem. Soc.* **130**, 6680–6681.
6. Huang, C. Y., Thayer, D. A., Chang, A. Y., Best, M. D., Hoffmann, J., Head, S., and Wong, C. H. (2006) Carbohydrate microarray for profiling the antibodies interacting with Globo H tumor antigen. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15–20.
7. Liang, P. H., Wang, S. K., and Wong, C. H. (2007) Quantitative analysis of carbohydrate-protein interactions using glycan microarrays: determination of surface and solution dissociation constants. *J. Am. Chem. Soc.* **129**, 11177–11184.
8. Gao, J., Liu, D., and Wang, Z. (2008) Microarray-based study of carbohydrate-protein binding by gold nanoparticle probes. *Anal. Chem.* **80**, 8822–8827.
9. Taton, T. A., Mirkin, C. A., and Letsinger, R. L. (2000) Scanometric DNA array detection with nanoparticle probes. *Science* **289**, 1757–1760.
10. Nam, J. M., Thaxton, C. S., and Mirkin, C. A. (2003) Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins. *Science* **301**, 1884–1886.
11. Katz, E., and Willner, I. (2004) Integrated nanoparticle-biomolecule hybrid systems: synthesis, properties, and applications. *Angew. Chem. Int. Ed.*, **43**, 6042–6108.
12. Sun, L., Liu, D., and Wang, Z. (2007) Microarray-based kinase inhibition assay by gold nanoparticle probes. *Anal. Chem.* **79**, 773–777.
13. Turkevich, J., Stevenson, P. C., and Hillier, J. (1951) A study of the nucleation and growth processes in the synthesis of colloidal gold. *Discuss. Faraday Soc.* **11**, 55–75.
14. Frens, G. (1973) Controlled nucleation for the regulation of the particle size in monodisperse gold suspensions. *Nat. Phys. Sci.* **241**, 20–22.
15. Wang, Z. X., Levy, R., Fernig, D. G., and Brust, M. (2005) The peptide route to multifunctional gold nanoparticles. *Bioconjugate Chem.* **16**, 497–500.

Chapter 11

The Application of NMR Spectroscopy to Functional Glycomics

Jean-Robert Brisson, Evgeny Vinogradov, David J. McNally, Nam Huan Khieu, Ian C. Schoenhofen, Susan M. Logan, and Harold Jarrell

Abstract

Glycomics which is the study of saccharides and genes responsible for their formation requires the continuous development of rapid and sensitive methods for the identification of glycan structures. It involves glycoanalysis which relies upon the development of methods for determining the structure and interactions of carbohydrates. For the application of functional glycomics to microbial virulence, carbohydrates and their associated metabolic and carbohydrate processing enzymes and respective genes can be identified and exploited as targets for drug discovery, glyco-engineering, vaccine design, and detection and diagnosis of diseases. Glycomics also encompasses the detailed understanding of carbohydrate–protein interactions and this knowledge can be applied to research efforts focused toward the development of vaccines and immunological therapies to alleviate infectious diseases.

Key words: Glycomics, glycans, polysaccharides, glycoanalysis, protein–carbohydrate interactions, HR-MAS, NMR, structural analysis, molecular modeling.

1. Introduction

Determination of the functional role of carbohydrates in biology requires the development of methodologies that can analyze carbohydrate structures with increasing levels of sensitivity and simplicity. Recent efforts by our laboratory to characterize the glycome of a number of important bacterial pathogens led to the elucidation of many surface glycan structures (LOS, CPS, N-linked, and O-linked glycoproteins) and the development and

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_11, © Humana Press, a part of Springer Science+Business Media, LLC 2010

adaptation of rapid methods for glycan detection and study of carbohydrate–protein interactions (1–7).

The presence of novel monosaccharides in bacterial glycans presents unique challenges for structural analysis. Mass spectrometry plays an integral part in the discovery of these novel glycans due to its high sensitivity and mass accuracy. With recent advances in NMR technology, detailed structural characterization from nanomolar amounts of purified glycans is now possible (**Fig. 11.1**). Of all the spectroscopic methods currently available, high-resolution NMR spectroscopy offers the most complete approach for structure determination of saccharides which includes the determination of the stereochemistry of sugars, sequence, non-carbohydrate substituents, as well as conformation and dynamics. In addition, high-resolution NMR can facilitate the study of sugar–protein interactions. Another advantage of NMR is that labile compounds can be analyzed without purification of individual molecules, a process which can often lead to the loss of labile components. In comparison with mass spectrometry, NMR is, however, a relatively insensitive technique in terms of the amount of material required. Recent advances in equipment and continuing developments in techniques are reducing the threshold in the amount of material required.

Functional glycomics of bacterial glycoproteins requires the identification of novel glycans on glycoproteins and the genes associated with their synthesis. One approach is to complete the structural characterization of a novel glycan and to use this information to analyze bioinformatically the corresponding genome

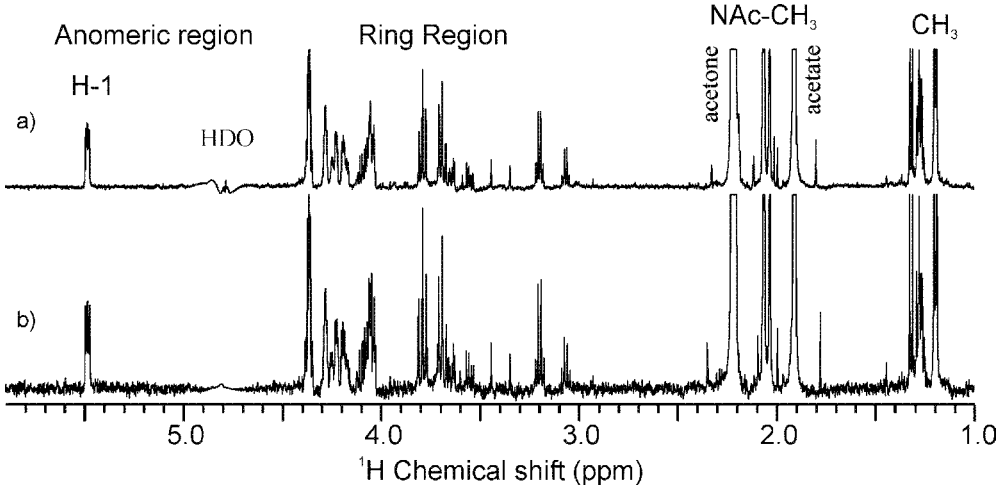


Fig. 11.1. Increase in signal to noise obtained using a cryogenically cooled probe. The anomeric region of the NMR spectrum for a sugar metabolite sample (200 μL 99% D_2O) isolated from *C. jejuni* 81–176 (8) is shown. (a) ^{1}H NMR spectrum (128 scans) acquired at 600 MHz using a cold probe ($S/N = 100:1$). (b) ^{1}H NMR spectrum (128 scans) acquired at 500 MHz using a standard probe ($S/N = 39:1$).

to identify putative glycan biosynthetic genes which share homology with genes responsible for the biosynthesis of similar sugars in other organisms. These genes can then be mutated to confirm their role and/or the encoded proteins expressed recombinantly for detailed functional characterization and elucidation of biosynthetic pathways. For structural characterization of novel glycans a complete mass spectrometric analysis is completed (Chapter 8) and then purification of the glycan is scaled up to do NMR analysis and determine the stereochemistry of the glycan as well as the identity and location of substituents (**Fig. 11.2**).

Functional glycomics of polysaccharides from bacterial pathogens has also benefited from the ability of HR-MAS NMR to rapidly identify glycan structures from 10^8 to 10^{10} bacterial cells directly without recourse to isolation and purification (7). This method can be applied to investigate expression of glycans

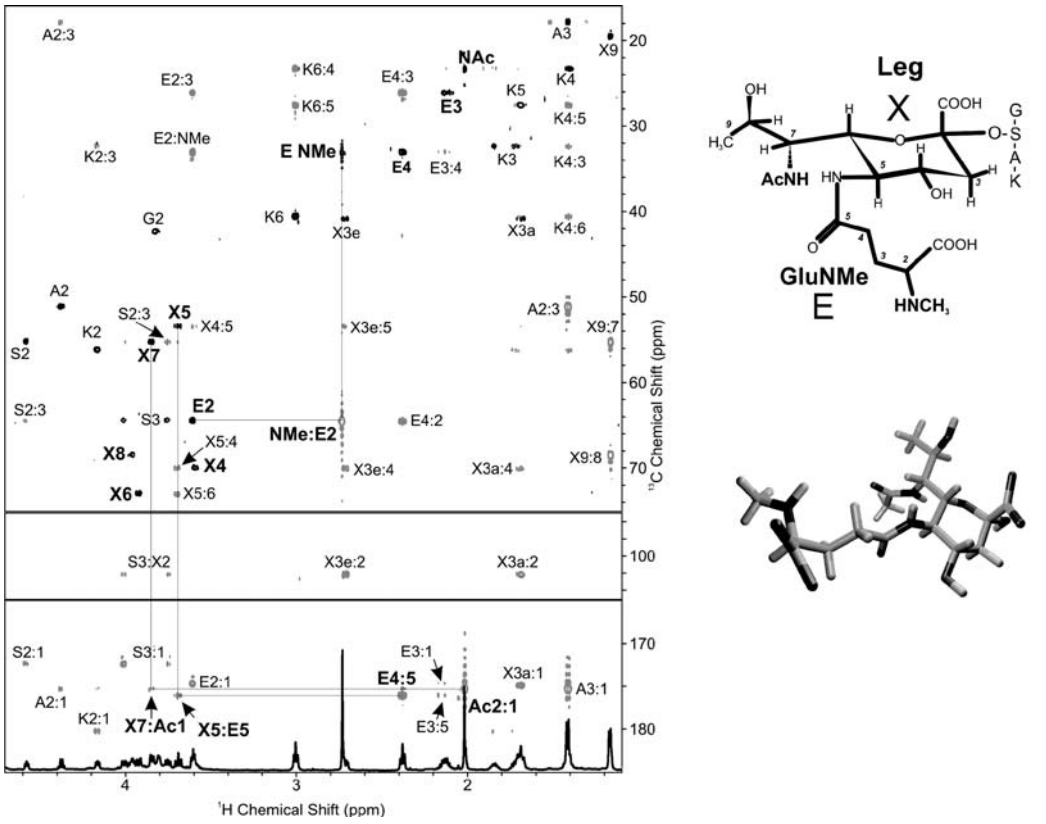


Fig. 11.2. Heteronuclear NMR experiments for the O-linked glycan Leg5GluNMe7Ac for glycopeptide GSAK from *C. botulinum*. In the overlap of the ^{1}H–^{13}C HSQC and HMBC spectra, the single bond H–C correlations are indicated by a residue code and atom number using the single letter code for the amino acids, X for Leg, Ac for the acetyl group, E for Glu and NMe for the *N*-methyl group. The multiple bond correlations are indicated as H:C. Correlations used to determine the location of the NAc and GluNMe groups are joined by *solid lines*. The proton spectrum is shown on the bottom. The structure is shown on the left along with a molecular model.

under different laboratory growth conditions and directly from the natural environments in which the pathogen is found. Examination of mutants allows the assignment of genes involved in the biosynthetic pathways of these glycans and their modifications and helps to determine the importance of structural phase variability in survival and pathogenesis (3). Minor glycan structures and labile components, which had previously been undetected due to isolation procedures of purified material, could be detected by examining the intact cells (6,8). Characterization of

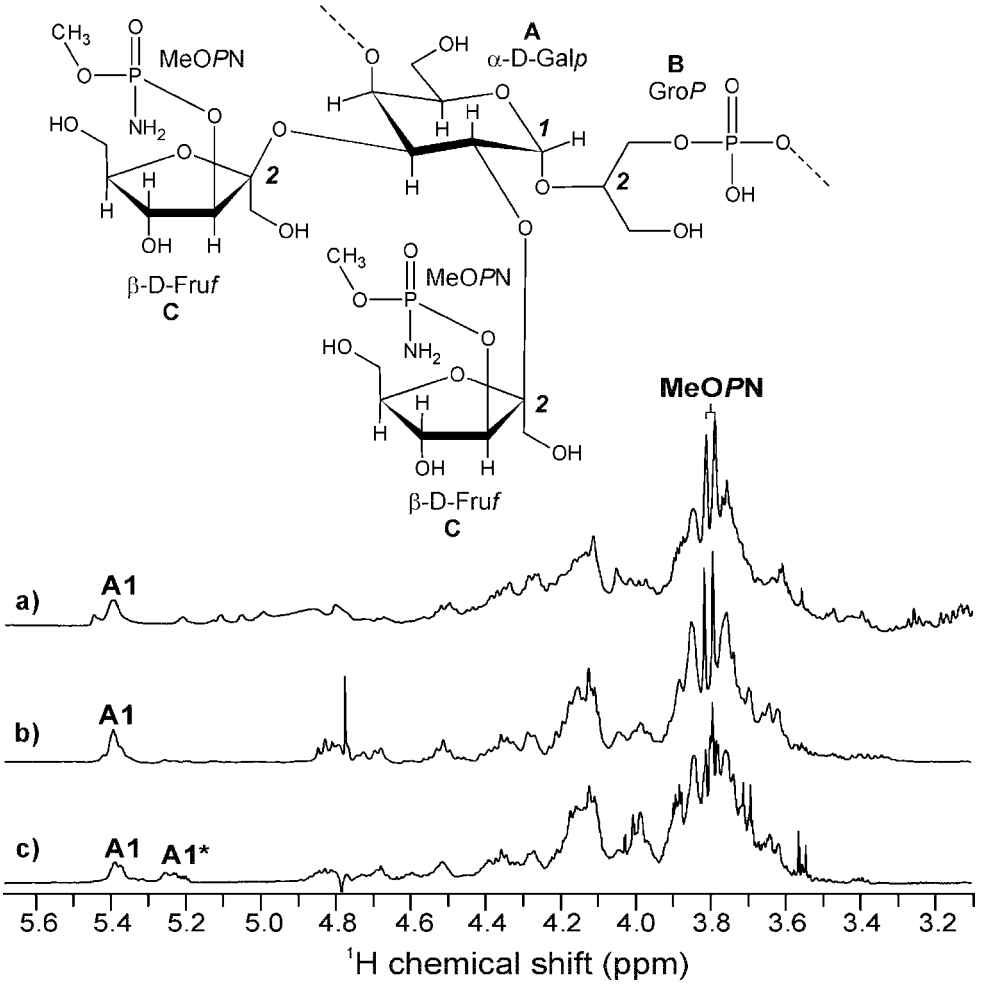


Fig. 11.3. Structure for the complete repeating unit of *C. jejuni* HS:1, [-4)-α-D-Gal*p*-(1-2)-Gro-(1-P-]$_n$ with MeO*P*N-3-β-D-fructofuranose branches at C-2 and C-3 of Gal. Structural heterogeneity is due to variable phosphoramidate groups on non-stoichiometric fructose branches. Residue A is α-D-Gal*p*, α-D-galactopyranose; residue B is Gro*P*, glycerol-phosphate; residue C is β-D-Fru*f*, β-D-fructofuranose; and MeO*P*N is *O*-methyl phosphoramidate, $CH_3OP(O)(NH_2)(OR)$. The proton NMR spectrum of cell-bound and purified and *C. jejuni* HS:1 CPS indicates that the anomeric resonance and the labile MeO*P*N resonance observed in the HR-MAS spectrum are still present in the ^{1}H NMR spectrum of an enzyme purified CPS sample (b). In the ^{1}H NMR spectrum of a hot water/phenol purified CPS sample (c), the MeOPN doublet at 3.8 ppm is absent and the sample has degraded due to the presence of an additional A1∗ anomeric resonance due to loss of labile fructose residues.

these novel labile modifications was then possible by the use of mild enzymatic extraction methods instead of harsher chemical ones (**Fig. 11.3**). Since the ^{1}H NMR shifts of the anomeric protons are sensitive to the conformation about the glycosidic bond, a further consequence of examining glycans on cells surfaces is that the resulting spectra provide insight into the similarity between their in situ structure/conformation and that of the isolated glycan.

In combination with metabolomics and genomics, NMR was also successfully used to identify the complete biosynthetic pathways of novel carbohydrate substituents on glycoproteins in *H. pylori* and *C. jejuni* and elucidate the reaction products of individual biosynthetic enzymes involved in the pseudaminic acid (Pse) biosynthetic pathway (2,5). The ability to carry out the enzymatic reactions in the NMR tube and observe in real-time the products being made led to the discovery of novel labile reaction products and to a revision of the biosynthetic steps required for synthesis of this novel monosaccharide (**Fig. 11.4**). Previously, the use of conventional structural analysis procedures of isolating each reaction product by chromatography prior to analysis led to erroneous conclusions on individual reaction product structure since the labile compounds were lost during the purification procedures.

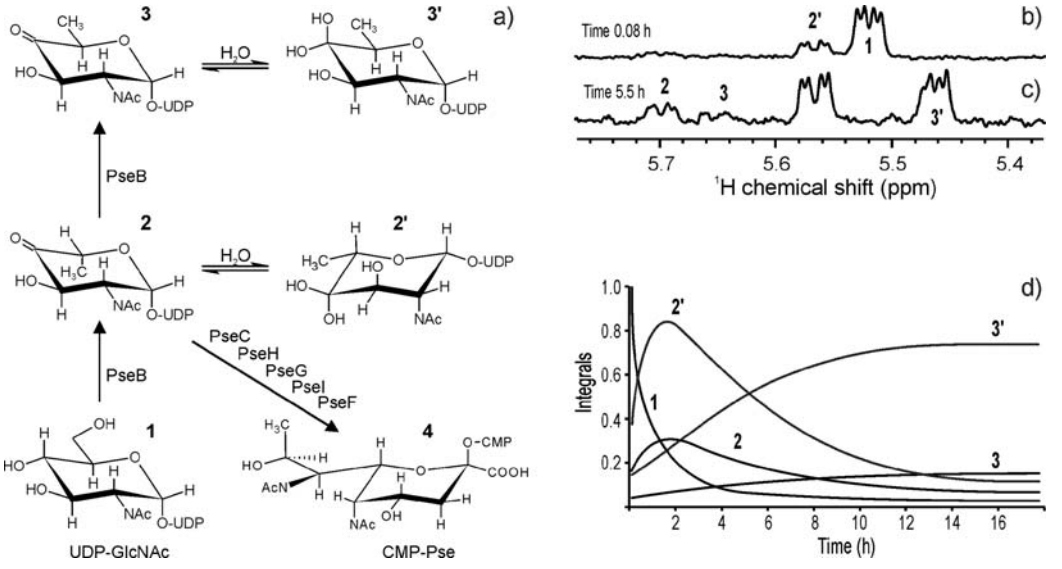


Fig. 11.4. Products of the PseB enzymatic reactions in the CMP-Pse biosynthetic pathway (a). **1** is UDP-α-D-GlcNAc; **2** is UDP-2-acetamido-2,6-dideoxy-β-L-*arabino*-hexos-4-ulose; **2'** is the *gem*-diol form of **2**; **3** is UDP-2-acetamido-2,6-dideoxy-α-D-*xylo*-hexos-4-ulose; **3'** is the *gem*-diol form of **3**; and **4** is CMP-pseudaminic acid. Proton spectra at 500 MHz of the anomeric region (b, c) for the PseB enzymatic reaction in aqueous reaction buffer (25°C, 10% D_2O, 25 mM UDP-a-D-GlcNAc, 638 μg PseB, 25 mM NaCl, 25 mM NaPO4, pH 7.2) from a time course of spectra acquired every 10 min. (d) Integrals for the anomeric proton resonances plotted over time.

NMR can also be used to study protein–carbohydrate interactions of the first enzyme in the Pse pathway, PseB. STD NMR was utilized to examine interactions within the active site of PseB with the substrate UDP-GlcNAc as well as the final reaction product of the pathway, CMP-Pse, which is a potent natural inhibitor of this enzyme (**Table 11.1**, **Figs. 11.5** and **11.6**).

Table 11.1
Saturation transfer difference enhancements for CMP-Pse with PseB. Comparison of calculated results using CORCEMA to those observed experimentally using STD NMR

	STD enhancement (%)	
Resonance	**Calculated**	**Observed[a]**
Pse H3ax	27	32
Pse H3eq	26	29
Pse H7	36	27
Pse H8	45	9
Pse H9	25	24
Pse 5NAc-CH_3	70	25
Pse 7NAc-CH_3	77	22
Ribose H1	100	100
Cytosine H5	56	44
Cytosine H6	91	32

[a]STD enhancements for Pse H4, H5 and Ribose H2, H3, H4, and H5/5' were not determined experimentally due to spectral overlap that prevented accurate integration of individual resonances.

2. Materials

2.1. Bacterial Cells and Carbohydrates

1. Bacterial cells are grown under various conditions often specific to the bacterium and isolated as previously described widely in numerous literature articles.
2. Carbohydrate samples are isolated from glycoprotein extracts or the CPS of bacteria as previously described.

2.2. Chemicals and Reagents

1. Pronase E (type XIV from *Streptomyces griseus*, EC 3.4.24.31, 5.2 units/mg, Sigma, Oakville, Canada).
2. Protease K (type XXIV from *Bacillus licheniformis*, EC 3.4.21.62, 8.8 units/mg, Sigma, Oakville, Canada).
3. Potassium-buffered 98% D_2O (pD 7.0) (Cambridge Isotopes Laboratories, Inc., Andover, USA).

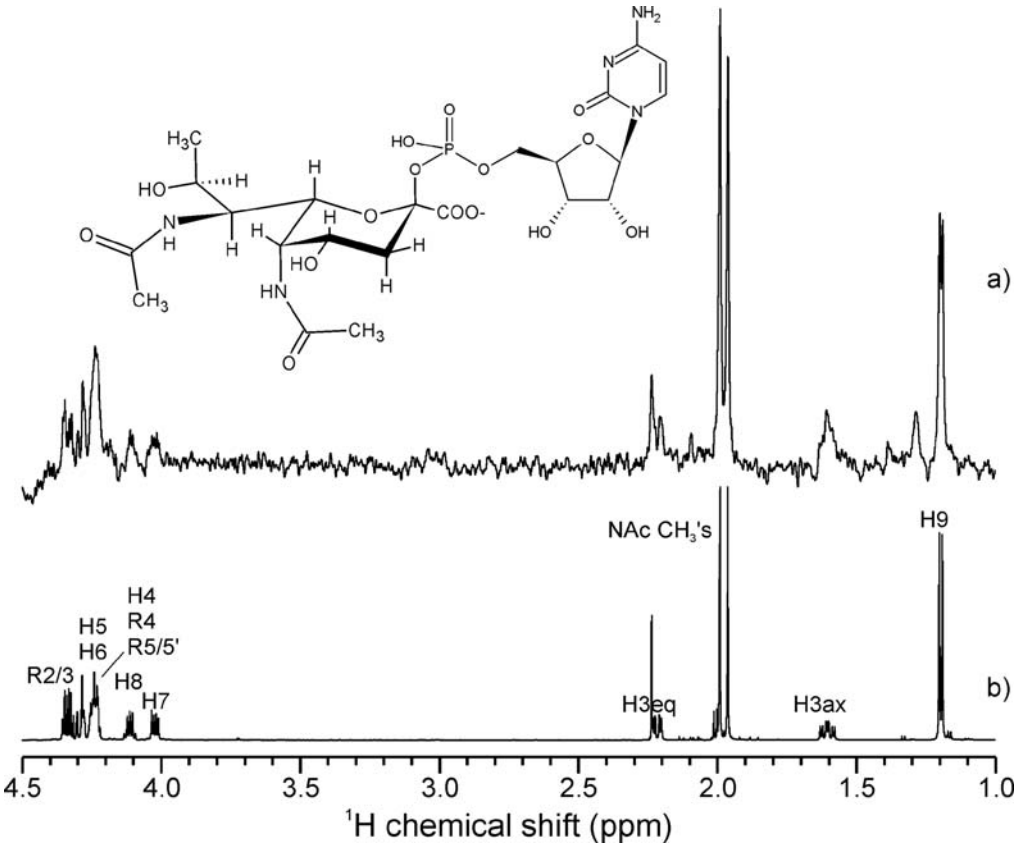


Fig. 11.5. Detection of binding of CMP-Pse to PseB using STD NMR (600 MHz, ^{1}H) (25°C, 3% H_2O/97% D_2O, 25 mM $NaPO_4$, 25 mM NaCl, pH 7.3, 4.4 mM CMP-Pse, 22.3 μM PseB, substrate:protein 200:1). (a) STD NMR spectrum (2 s saturation, 3000 scans). (b) Proton reference spectrum (128 scans). STD effects were not determined for R2, 3, 4, 5/5' as well as for Pse 4–6 resonances due to spectral overlap that prevented accurate integration. R represents ribose.

4. Lysozyme (Sigma, Oakville, Canada).
5. Mutanolysin (Sigma, Oakville, Canada).
6. DNAse I and RNAse (Sigma, Oakville, Canada).
7. Sephadex® superfine G-50 column (Sigma, Oakville, Canada).
8. Waters differential refractometer (model R403, Waters, Mississauga, Canada).
9. Sephadex® superfine G-15 column (Sigma, Oakville, Canada).

2.3. Instrumentation

1. Gilson liquid chromatograph (model 306 and 302 pumps, 811 dynamic mixer, 802B manometric module, Gilson, Middleton, WI, USA) with a Gilson UV detector (220 nm) (model UV/Vis-151 detector, Gilson, Middleton, WI)

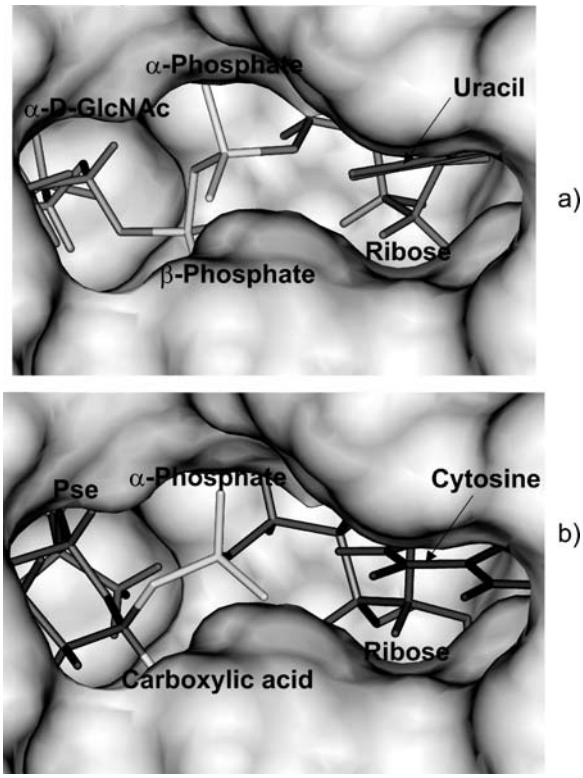


Fig. 11.6. Molecular model showing UDP-α-D-GlcNAc (a) docked within the active site of PseB and of the energy minimized structure for CMP-Pse (b) docked within the active site of PseB. The pyridine rings, ribose and sugar, and the negatively charged groups for both nucleotide sugars occupy similar space.

equipped with a tandem QHP HiTrap™ ion exchange column (Amersham Biosciences, Piscataway, NJ, USA).

2. A 500 or 600 MHz (^{1}H) NMR spectrometer equipped with a Z gradient 3 or 5 mm triple resonance (^{1}H, ^{13}C and ^{31}P or ^{15}N) probe. Due to the tendency of carbohydrate NMR spectra to exhibit extensive overlap of signals, the use of as high a magnetic field as possible is desirable to reduce spectral crowding and thus facilitate analysis. A cryogenically cooled probe (cold probe) is used for optimized sensitivity. For high-resolution magic angle spinning experiments, a magic angle spinning probe is required.

2.4. Software for the Analysis of NMR Data

1. NMR software for analysis of NMR data is usually supplied by the NMR vendor.
2. Molecular modeling software (Insight, Sybyl, Amber, MM3, Hyperchem) for generation of minimized protein and carbohydrate structures. Distances extracted from minimized carbohydrate structures are used to analyze experimental distances obtained from measurements of nuclear

Overhauser enhancements (NOE) NMR experiments. The dihedral angles can also be used to analyze proton coupling constants obtained from ^{1}H NMR experiments.

3. Software for performing automated docking of carbohydrate and protein (AUTODOCK (9) or similar software).
4. Software (CORCEMA) (10–12) for calculating theoretical enhancements from STD NMR experiments and docking studies and comparing with experimental ones (**Table 11.1**).

3. Methods

Sample preparation is the most crucial step for NMR spectroscopy. The analyzed compound ideally must be free of impurities or at least has not too much overlap with non-related cross-peaks in two-dimensional spectra. Viscosity leads to signals broadening; this problem can be solved by temperature increase, addition of deuterated detergents (SDS), pH, or salt content variation. For the glycopeptide analysis, the method of protein digestion is important. Ideally one can obtain carbohydrate linked to a single amino acid, but this is still out of control. Proteinase digestion most often leads to the production of sugar part linked to several short peptides, which hinder the isolation of pure compounds, but on the other hand can be useful for the localization of particular glycoforms on a protein.

Proton spectra are acquired to determine if the sample is suitable for further NMR experiments. If the anomeric resonances (4–6 ppm) or any other isolated resonances can be observed at a suitable signal to noise (>100:1), then further experiments can be performed for complete structural analysis. To increase the signal to noise (sensitivity) by a factor of 2–3, a cryogenically cooled probe can be used (**Fig. 11.1**). To observe the glycan from the CPS present on intact bacterial cells, high-resolution magic angle spinning can be used (**Fig. 11.3**). In this case, observation of the carbohydrate resonances depends on the mobility of the molecule as well as its abundance in the sample.

Most often, structural analysis of glycans requires the complete analysis of all the resonances in the spectrum. Several reviews deal with structural analysis of carbohydrate (13–16). If all the resonances cannot be identified, additional experimental data obtained from chemical methods and mass spectrometry must be obtained to support the conclusions reached from NMR analysis. The NMR analysis for the glycopeptide shown in **Fig. 11.2** required a complete assignment of the carbohydrate and peptide resonances since the modification at C5 of the nonulosonic acid

was a novel N-methylated amino acid that had to be distinguished from the other amino acids in peptide part. In this case mass spectral data were critical for the identification of a methyl group, since similar signals often belong to impurities.

NMR experiments are usually performed with suppression of the HOD signal. The position of the HOD signal varies with temperature and its chemical shift can be adjusted so that it does not overlap sugar resonances. Standard two-dimensional homonuclear experiments, COSY, TOCSY, NOESY, are performed for assignment of the proton signals. The spin lock for TOCSY experiments can vary from 30 to150 ms to establish the complete homonuclear scalar coupling pathways. Mixing times for NOESY experiments can vary from 100 to 800 ms depending on the size of the molecule being studied. The ROESY pulse sequence gives excellent results as it is not dependent on the molecular mass of analyzed compound. Two-dimensional ^{1}H–^{13}C experiments are performed for assignment of ^{13}C resonances. If the glycan is phosphorylated ^{1}H–^{31}P heteronuclear experiments are performed for the assignment of ^{31}P resonances. Depending on the concentration of the samples, all these experiments can be acquired in a matter of hours to overnight.

Glycopolymers residing on bacterial surface have sufficiently slow relaxation to be observed by high-resolution magic angle spinning method (HR-MAS). In this case, observation of the carbohydrate resonances depends on the mobility of the molecule as well as its abundance in the sample. HR-MAS experiments of bacterial cells (**Fig. 11.3**) are performed using a magic angle spinning probe. ^{1}H NMR spectra from 40 μL samples are spun at 3 kHz and acquired with the Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence [90-(τ-180-τ)$_n$-acquisition] to remove broad lines arising from lipids and solid-like material. The total duration of the CPMG pulse ($n*2\tau$) was 10 ms with τ set to (1/MAS spin rate) was typically 10 ms although the duration should be optimized for each bacterial system. We have found that for glycans that are phosphorylated both one-dimensional and two-dimensional ^{1}H–^{31}P heteronuclear correlation experiments under HR-MAS conditions are practical for whole cells. Indeed, this method served as a basis for screening bacterial strains for a genetic biomarker (3).

One-dimensional selective TOCSY experiments and selective NOESY can be used instead of two-dimensional experiments for higher resolution and accurate measurements of proton coupling constants. For use under MAS conditions, the TOCSY sequences were modified so that the DIPSI-2 mixing sequence was replaced with the adiabatic WURST-2 pulses (8) which reduces the interference effects between sample rotation and the spin lock. Typically, proton spectra of bacterial cells could be obtained using 256–1024 transients (15 min to 1 h). For the selective

experiments focusing on glycan resonances of minor components in the bacterial cells, the time for each TOCSY and NOESY varied from 1 to 8 h.

3.1. Preparation of Glycopeptides Using Proteinase K Digestion

A purified glycoprotein sample is digested by pronase/proteinase K digest followed by size exclusion chromatography for isolation of glycopeptides (1).

1. Take 8 mg of glycoprotein in water and adjust the pH to 8.5 (optimal for proteinase K) by the addition of Na_2HPO_4.
2. Add 10 mg of proteinase K (10 mg) and incubate the solution at 50°C for one week (*see* **Note 1**).
3. Filter the resulting digest through a Sep-Pak C18 column (pre-washed with methanol and then water) for the removal of precipitate and hydrophobic peptides.
4. Desalt the sample on a Sephadex G 15 column (1.6×80 cm) in pyridine (0.4%)acetic acid (1%) in H_2O, monitoring elution with refractive index detector.
5. Collect eluate in 5 mL fractions. Lyophilize and analyze each fraction. Lyophilize and analyze by NMR and MS to identify fractions that contain glycopeptides (**Fig. 11.2**) (*see* **Note 2**).

3.2. Preparation of Bacterial Cells for HR-MAS NMR

Cells should be resuspended in a buffer, such as a phosphate buffer, that does not contain strong proton signal which would interfere with proton-detected NMR experiments.

1. Overnight growth from one MH agar plate was harvested and placed in 1 mL of 10 mM potassium-buffered 98% D_2O (pD 7.0) containing 10% sodium azide (w/v) for 1 h at room temperature to kill cells.
2. The cells were pelleted by centrifugation (8900×*g* for 2 min) and washed once with 10 mM potassium-buffered saline in D_2O.
3. Cells can be frozen at this point for use at a later time.
4. The pellet was resuspended by adding 20 μL of D_2O for a final volume of about 40 μL. Approximately 10 μL of 1% (w/v) TSP in D_2O was then added as an internal standard (0 ppm) to the cell suspension prior to being loaded into an NMR rotor for HR-MAS using a long tipped pipette cut diagonally approximately 1 cm from the end (*see* **Note 3**).

3.3. Enzymatic Isolation of CPS

An enzymatic method of isolating CPS from bacterial cells was developed based on the methodologies of Darveau and Hancock (17), Huebner et al. (18), and Hsieh et al. (19) to preserve labile groups.

1. Cells harvested from 6 L of BHI broth were suspended in PBS buffer (pH 7.4).
2. Lysozyme was then added to a final concentration of 1 mg/mL prior to the addition of mutanolysin to a final concentration of 67 U/mL.
3. The bacterial cell suspension was then incubated for 24 h at 37°C with agitation at 100 rpm.
4. The mixture was then emulsiflexed twice (21,000 psi) to lyse cells and DNAse I and RNAse (130 μg/mL each) were added prior to being incubated for 4 h at 37°C with agitation at 100 rpm.
5. Pronase E and protease K were both added to a final concentration of 200 μg/mL followed by incubation at 37°C overnight with agitation at 100 rpm.
6. The crude CPS extract was then dialyzed against running water for 72 h, ultracentrifuged for 2 h (140,000×*g*, 15°C), and the supernatant was lyophilized.
7. Crude CPS was resuspended in H_2O and purified using a Sephadex® superfine G-50 column equipped with a Waters differential refractometer.
8. Fractions containing CPS were combined and lyophilized. Semi-purified CPS was then resuspended in H_2O and purified using a Gilson liquid chromatograph.
9. Fractions containing CPS were combined and lyophilized. Purified bacterial CPS was then de-salted using a Sephadex® superfine G-15 column lyophilized and stored at –20°C until further analysis.

3.4. Protein Sample Preparation for NMR Experiments

For functional characterization, the isolation of enzyme products and study of protein–carbohydrate interactions, it is recommended that proteins are isolated in a buffer, such as a phosphate buffer, that does not contain strong proton signal which would interfere with proton-detected NMR experiments. The PseB protein-coding sequence was cloned from *H. pylori* DNA and constructed with C-terminal His_6-tag (5).

1. The expression strain was grown in 500 ml 2× YT medium with either kanamycin (50 μg/mL) or ampicillin (50 μg/mL) for selection. The cultures were grown at 30°C, induced at an OD_{600} of 0.6 with 0.1 mM IPTG, and harvested 2.75 h later.
2. Cell pellets were resuspended in lysis buffer [50 mM sodium phosphate pH 7.7 (unless otherwise stated), 400 mM NaCl, 10 mM β-mercaptoethanol] containing 10 mM imidazole and complete protease inhibitor cocktail, EDTA-free.

3. After addition of 10 μg/mL of RNaseA and DNase I, the cells were disrupted by two passes through an Emulsiflex C5 (20,000 psi). Lysates were centrifuged at 100,000×*g* for 1 h at 4°C and the supernatant fraction was applied to a 1 ml Ni-NTA column equilibrated in 10 mM imidazole lysis buffer, using a flow rate of 0.5 mL/min.
4. After sample application, the column was washed with 20 column volumes (CV) of 10 mM imidazole lysis buffer. To elute the protein of interest, a linear gradient from 10 to 100 mM imidazole, in lysis buffer, over 60 CV was applied to the column prior to a final pulse of 40 CV of 200 mM imidazole lysis buffer.
5. Fractions containing the purified protein of interest, as determined by SDS-PAGE (12.5%) and Coomassie staining, were pooled and dialyzed against dialysis buffer [25 mM sodium phosphate, pH 7.7 (unless otherwise stated), 50 mM NaCl] overnight at 4°C. PseB protein concentration was measured spectrophotometrically using an A_{280} 0.1% value of 0.536.

3.5. Sample Preparation for Observation of Labile Products from an Enzymatic Reaction

1. Enzymatic reactions were carried out in 3 mm NMR tubes at 25°C in 90% aqueous buffer (10% D_2O, 25 mM $NaPO_4$, 50 mM NaCl, pH 7.2). The sugar concentration was set at 25 mM to obtain in a few scans a proton spectrum with excellent signal to noise for detection of minor components. The PseB protein concentration was adjusted (638 μg) so that the reaction could be monitored in an overnight run (**Fig. 11.4**).
2. The ligand was dissolved in a stock solution and 10 μL was added to the NMR tube for a final concentration of 25 mM. The sugar concentration was set at 25 mM to obtain in a few scans a proton spectrum with excellent signal to noise for detection of minor components. Immediately after combining the enzyme with its ligand, NMR tubes were centrifuged to eliminate air bubbles (5 min, 2000 rpm).
3. The reaction was monitored continuously through the acquisition of ^1H spectrum at regular time intervals (128 scans, 10 min) using a 500-MHz (^1H) spectrometer with a 3-mm probe. In order that the results be quantitative, data were acquired with a $\pi/4$ RF pulses and a total delay of 3.5 s between the acquisitions being averaged for each spectrum. For the anomeric protons, these conditions would give intensities that are >98% of their actual relative intensities.
4. Integrals for anomeric protons which express their relative concentrations were plotted over time (*see* **Note 4**).

5. To perform structural characterization of reaction products, scaled-up reactions performed concurrently in the laboratory (1–3 mL) were stopped when the concentration of each respective product was highest. Enzymes were removed by membrane filtration and samples were exchanged into D_2O using several partial lyophilization steps with each step exchanging approximately 25% of the total sample volume with D_2O.
6. Samples were then concentrated during the final lyophilization step for structural analysis by NMR.

3.6. STD NMR Experiments to Determine Protein–Carbohydrate Interactions

The STD NMR experiment is widely used in detecting/screening ligand binding to proteins where the dissociation constant (K_d) is in the range of $\sim 10^{-3}$ to 10^{-7} M^{-1}. The method relies upon intermolecular cross-relaxation between protein protons and ligand protons. Thus signal amplitude is dependent primarily upon the distance between protein protons and ligand protons, but it is sensitive to a number of other factors. For short saturation times, protons that are closer to the protein surface upon binding are revealed by larger STD NMR signal amplitudes. In optimal cases the measured enhancement of ligand resonances can be related to the conformation and orientation of the ligand in the binding site. *See* references in McNally et al. (2) for a complete list of the use of STD NMR.

1. Recombinant *H. pylori* PseB was expressed and purified as previously reported. CMP-Pse was synthesized and purified from a one-pot reaction of PseB, PseC, PseH, PseG, PseI, and PseG with UDP-GlcNac.
2. STD NMR spectra were acquired at 13°C or 25°C in 3 mm NMR tubes on a 600-MHz (^1H) spectrometer with a cryogenically cooled probe (*see* **Note 5**).
3. STD experiments were carried out using an excess of ligand:protein (200:1) in buffered D_2O (8% H_2O/92% D_2O, 25 mM $NaPO_4$, 25 mM NaCl, pH 7.3). Substrate concentrations and conversions were determined using the molar extinction coefficients for UDP (ε_{260}=10,000) and CMP (ε_{260}=7400) (*see* **Note 6**).
4. Saturation of PseB resonances was achieved using a train of Gaussian-shaped pulses, with bandwidths of 300 Hz centered at –1.0 ppm.
5. Phase cycling was used to subtract reference spectra (off-resonance pulses centered at 30 ppm) from those where PseB resonances were excited. STD spectra were acquired using 2 s saturation time, a relaxation delay of 2.5 s, an acquisition time of 1.9 s, and the number of transients varied from 1024 to 3072 (**Fig. 11.5**) (*see* **Note 7**).

6. A spin-lock filter of 10 ms was used to suppress the protein background (*see* **Note 8**).
7. STD enhancements are frequently reported as a percentage of the resonance that received the largest absolute amount of saturation transfer, which was normalized to 100% (**Table 11.1**) (*see* **Note 9**).

3.7. Molecular Modeling and Calculation of STD Enhancements

The underlying theory for CORCEMA-ST calculations is described elsewhere (10–12). To reduce the computational time, only protons from the protein and NAD for which the distance to at least one CMP-Pse proton was less than or equal to 5 Å were used. Since STD NMR experiments were performed in D_2O, OH and NH ligand protons as well as amino acid protons in the active site were assumed to have exchanged and were not included in the calculations.

1. The Autodock 3.0 software was used to dock ligands into the PseB monomer. Coordinates for PseB were extracted from the X-ray structure reported for the PseB dimer (20). Ligand structures were constructed using Insight II software package (Accelrys, San Diego, USA) (*see* **Note 10**). All molecules were further processed in ADT, a Graphical User Interface of AutoDock, where atomic charges were assigned using a Gasteiger method (21) then nonpolar hydrogens were merged.
2. Grid maps were calculated at the active site of receptor using AutoGrid 3.0. All ligand torsions except those of amides and sugar rings were allowed to move. Step sizes were 0.2 Å for ligand translations and 5° for ligand orientations and torsions. A Lamarckian genetic algorithm (9) was used to perform 100 docking experiments each with an initial population of 50 random individuals and a maximum number of 1.5×10^6 energy evaluations (*see* **Note 11**). **Figure 11.6** shows the lowest energy conformation for UDP-GlcNAc and CMP-Pse docked within the PseB active site. Figures for molecular models were generated using the Insight II software.
3. For intramethyl group relaxation, the generalized order parameter S^2 was set to 0.25, while an S^2 of 0.85 was used for methyl-X relaxation.
4. In keeping with experimental conditions, the concentration of the ligand was set to 8 mM with a ligand-to-protein ratio of 200:1.
5. The value for K_{on} was set to 10^8 and the equilibrium constant K_{eq} was varied between 10^3 and 10^6 with 2×10^6 giving the best fit when compared to experimental results (saturation time 2 s) (*see* **Note 12**).

6. The correlation time of the free ligand was set 0.3 ns which gave calculated longitudinal relaxation rates that were consistent with those measured experimentally for Pse (not shown) and those reported for UDP-Glc (22).
7. A correlation time of 80 ns was used for the PseB/CMP-Pse complex (hexameric form) since it gave the best fit when compared to the experimental STD signal observed for ribose H1 (saturation time 2 s).
8. STD signal intensities were evaluated as $100\times[(I(0)_j - I(t)_j)/I(0)_j]$ with $I(t)_j$ and $I(0)_j$ corresponding to the intensity of proton j with and without saturation transfer during a saturation time *t*. Relative saturation transfer intensities are expressed as the ratio of the STD intensity of proton k in the ligand to that of the ligand ribosyl H1 (**Table 11.1**).

4. Notes

1. Duration of the reaction can be easily monitored by MS. Additional amount of proteinase can be used in case of slow progress of reaction.
2. The elution profile recorded by detector is irrelevant, it just indicates non-empty fractions that should be further analyzed. Often this is enough to obtain reasonably pure compound for NMR analysis. Further purification can include RP, IE, or HILIC type HPLC.
3. When filling the rotor it is important to minimize or eliminate the formation of an air bubble in the active volume to maximize sensitivity. **Note 2**: In cases where the sample will not achieve a stable spin rate or will only spin at rates lower than that desired, changing the turbine often helps.
4. If the enzymatic reaction proceeds too quickly at ambient temperature, one can change the protein:substrate ratio or reduce the temperature to reduce the reaction rate to a value more amenable for NMR.
5. It is prudent to verify, that under conditions used to perform the STD NMR experiments with protein, ligand alone exhibits no STD signal.
6. The absolute STD signal intensity typically reaches a plateau at ligand/enzyme ratios >200:1.
7. It is prudent to verify that signal cancellation is adequate under conditions being used for the STD NMR experiment. This may be done by performing the experiment with irradiation only “off resonance” and/or only

"on-resonance" which should result in residual signals that a significantly less than the true STD signals. Frequently, but not necessarily always, incomplete signal cancellation gives rise to signals that differ in "sign" or "phase" to that of the true STD signals.

8. An alternate method of removing potentially interfering protein resonances is to acquire an STD spectrum of the protein alone under identical conditions and to subtract this spectrum from that with the ligand present.
9. If long saturation times (>2 s) are used to obtain adequate (acceptable signal to noise) spectra, it is helpful to measure the ligand's ^{1}H spin lattice relaxation times for its various protons. If the ligand exhibits positional relaxation rates that differ considerably (>2-fold), the relative STD signal intensity for each position may no longer reflect only its proximity to the protein. Caution should be exercised in interpreting results.
10. For receptor and ligands alike, hydrogen atoms were added automatically. A short energy minimization of the ligands using Discover-3 program in Insight II was performed to alleviate bad atomic contacts.
11. Docked structures of ligands were clustered using a tolerance of 2.0 Å then the lowest binding energy conformation was selected as the best docked structure.
12. If is helpful in these calculations to minimize the number of parameters that must be varied in fitting the STD NMR data. Preferably parameters such as K_{on}, K_{eq}, and ligand correlation time may be determined independently.

Acknowledgments

We thank Scott Houliston and Dennis Whitfield for helpful discussion and Dr. N.R. Krishna (University of Alabama at Birmingham) for generously providing the CORCEMA-ST program.

References

1. Twine, S. M., Paul, C. J., Vinogradov, E., McNally, D. J., Brisson, J. -R., Mullen, J. A., McMullin, D. R., Jarrell, H. C., Austin, J. W., Kelly, J. F., and Logan, S. M. (2008) Flagellar glycosylation in *Clostridium botulinum*. *FEBS J.* **275**, 4428–4444.
2. McNally, D. J., Schoenhofen, I. C., Houliston, R. S., Khieu, N. H., Whitfield, D. M., Logan, S. M., Jarrell, H. C., and Brisson, J. -R. (2008) CMP-pseudaminic acid is a natural potent inhibitor of PseB, the first enzyme of the pseudaminic acid pathway in *Campylobacter jejuni* and *Helicobacter pylori*. *ChemMedChem.* **3**, 55–59.
3. McNally, D. J., Lamoureux, M. P., Karlyshev, A. V., Fiori, L. M., Li, J., Thacker, G., Cole-

man, R. A., Khieu, N. H., Wren, B. W., Brisson, J. -R., Jarrell, H. C., and Szymanski, C. M. (2007) Commonality and biosynthesis of the O-methyl phosphoramidate capsule modification in *Campylobacter jejuni*. *J. Biol. Chem.* **282**, 28566–28576.

4. McNally, D. J., Schoenhofen, I. C., Mulrooney, E. F., Whitfield, D. M., Vinogradov, E., Lam, J. S., Logan, S. M., and Brisson, J. -R. (2006) Identification of labile UDP-ketosugars in *Helicobacter pylori*, *Campylobacter jejuni* and *Pseudomonas aeruginosa*: key metabolites used to make glycan virulence factors. *Chembiochem.* **7**, 1865–1868.
5. Schoenhofen, I. C., McNally, D. J., Vinogradov, E., Whitfield, D., Young, N. M., Dick, S., Wakarchuk, W. W., Brisson, J. -R., and Logan, S. M. (2006) Functional characterization of dehydratase/aminotransferase pairs from Helicobacter and Campylobacter: enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. *J. Biol. Chem.* **281**, 723–732.
6. McNally, D. J., Jarrell, H. C., Li, J., Khieu, N. H., Vinogradov, E., Szymanski, C. M., and Brisson, J. -R. (2005) The HS:1 serostrain of *Campylobacter jejuni* has a complex teichoic acid-like capsular polysaccharide with nonstoichiometric fructofuranose branches and O-methyl phosphoramidate groups. *FEBS J.* **272**, 4407–4422.
7. Szymanski, C. M., St Michael, F., Jarrell, H. C., Li, J., Gilbert, M., Larocque, S., Vinogradov, E., and Brisson, J. -R. (2003) Detection of conserved N-linked glycans and phase-variable lipooligosaccharides and capsules from campylobacter cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy. *J. Biol. Chem.* **278**, 24509–24520.
8. McNally, D. J., Hui, J. P., Aubry, A. J., Mui, K. K., Guerry, P., Brisson, J. -R., Logan, S. M., and Soo, E. C. (2006) Functional characterization of the flagellar glycosylation locus in *Campylobacter jejuni* 81-176 using a focused metabolomics approach. *J. Biol. Chem.* **281**, 18489–18498.
9. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662.
10. Moseley, H. N., Curto, E. V., and Krishna, N. R. (1995) Complete relaxation and conformational exchange matrix (CORCEMA) analysis of NOESY spectra of interacting systems; two-dimensional transferred NOESY. *J. Magn. Reson. B* **108**, 243–261.
11. Jayalakshmi, V. and Krishna, N. R. (2002) Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J. Magn. Reson.* **155**, 106–118.
12. Jayalakshmi, V. and Rama, K. N. (2004) CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities. *J. Magn. Reson.* **168**, 36–45.
13. Brisson, J. -R., Sue, S. C., Wu, W. G., McManus, G., Nghia, P. T., and Uhrin, D. (2002) NMR of carbohydrates: 1D homonuclear selective methods, in *NMR Spectroscopy of Glycoconjugates* (Jimenez-Barbero, J. and Peters, T., Eds.) pp. 59–93, Wiley-VCH, Weinhem, Germany.
14. Duus, J. O., Gotfredsen, C. H., and Bock, K. (2000) Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem. Rev.* **100**, 4589–4614.
15. Uhrin, D. and Brisson, J. -R. (2000) Structure determination of microbial polysaccharides by high resolution NMR spectroscopy, in *NMR in Microbiology: Theory And Applications* (Barbotin, J. N. and Portais, J. C., Eds.) pp. 165–190, Horizon Scientific Press, Wymondham, U.K.
16. Kogan, G. and Uhrin, D. (2000) Current NMR methods in the structural elucidation of polysaccharides, in *New Advances in Analytical Chemistry* (Atta, u. R., Ed.) pp. 73–134, Harwood Academic, Amsterdam.
17. Darveau, R. P. and Hancock, R. E. (1983) Procedure for isolation of bacterial lipopolysaccharides from both smooth and rough *Pseudomonas aeruginosa* and *Salmonella typhimurium* strains. *J. Bacteriol.* **155**, 831–838.
18. Huebner, J., Wang, Y., Krueger, W. A., Madoff, L. C., Martirosian, G., Boisot, S., Goldmann, D. A., Kasper, D. L., Tzianabos, A. O., and Pier, G. B. (1999) Isolation and chemical characterization of a capsular polysaccharide antigen shared by clinical isolates of *Enterococcus faecalis* and vancomycin-resistant *Enterococcus faecium*. *Infect. Immun.* **67**, 1213–1219.
19. Hsieh, Y. C., Liang, S. M., Tsai, W. L., Chen, Y. H., Liu, T. Y., and Liang, C. M. (2003) Study of capsular polysaccharide from *Vibrio parahaemolyticus*. *Infect. Immun.* **71**, 3329–3336.
20. Ishiyama, N., Creuzenet, C., Miller, W. L., Demendi, M., Anderson, E. M., Harauz,

G., Lam, J. S., and Berghuis, A. M. (2006) Structural studies of FlaA1 from *Helicobacter pylori* reveal the mechanism for inverting 4,6-dehydratase activity. *J. Biol. Chem.* **281**, 24489–24495.
21. Gasteiger, J. and Marsili, M. (1980) Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. *Tetrahedron.* **36**, 3219–3228
22. Monteiro, C., Neyret, S., Leforestier, J., and du Penhoat, C. H. (2000) Solution conformation of various uridine diphosphoglucose salts as probed by NMR spectroscopy. *Carbohydr. Res.* **329**, 141–155.

Chapter 12

Metabolomics in Glycomics

Evelyn C. Soo and Joseph P. M. Hui

Abstract

Metabolomics is essentially the study of all low molecular weight molecules in a biological system under defined conditions. In glycomics, there is much potential to gain insight into the biosynthesis of novel glycoconjugate structures by probing the metabolome for substrates that are suspected, or known, to be involved in the biosynthetic processes. Recently, we employed the use of hydrophilic interaction liquid chromatography–mass spectrometry (HILIC–MS) in a focused metabolomic study of sugar-nucleotides relevant to the biosynthesis of highly novel carbohydrate modifications on the flagellin of *Campylobacter* sp. We exploited the unique selectivity of the HILIC–MS method for discriminating between closely related sugar-nucleotide intermediates and allowed their subsequent structural identification using a combination of high-resolution mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy. In addition, the HILIC–MS method permitted screening of selected isogenic mutants for sugar-nucleotide intermediates to determine a role for the corresponding genes on the flagellin glycosylation locus in the biosynthesis of the novel carbohydrate modifications.

Key words: Flagellin, hydrophilic interaction liquid chromatography, glycosylation, mass spectrometry, metabolomics, precursor ion scanning, sugar-nucleotides.

1. Introduction

Over the last few years, metabolomics has become recognised as an integral tool alongside more established "omics" technologies, namely genomics and proteomics, for gaining a comprehensive understanding of biological systems (1–4). The metabolome contains the end products of all metabolic processes and can respond profoundly to environmental changes and genetic perturbations. Accordingly, the comprehensive analysis of all the metabolites in a defined biological system, i.e. metabolomics, could yield tremendous information on cellular function and the physiological state

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_12, © Humana Press, a part of Springer Science+Business Media, LLC 2010

of a biological system. However, the metabolome is an inherently complex sample due to large numbers of metabolites that are present at any given time, the diverse physicochemical properties that are exhibited by the metabolites, and their presence at trace-level concentrations (5, 6). This sample complexity poses significant technological challenges when it comes to performing a true comprehensive analysis of the metabolome, where every small molecule metabolite is identified and quantified, as there is still no single analytical tool which provides the selectivity and sensitivity that is required for such an ambitious task. While the ability to perform a comprehensive analysis of the metabolome remains the ultimate goal in metabolomics it has become apparent, particularly when studying a selected metabolic pathway, that there is much to be gained from using a focused metabolomics approach to target metabolites that are suspected to be involved (7–10). In glycomics, there is significant opportunity to employ focused metabolomics approaches to study the biosynthesis of novel glycoconjugate structures since there are several classes of metabolites that are common to many glycoconjugate biosynthesis pathways (11–14).

The fields of metabolomics and glycomics have both grown significantly in recent years but the specific application of metabolomics in glycomics is still relatively unique (15–17). However, with the tremendous effort that is being undertaken to characterise the structures and functions of novel glycans, it is inevitable that there will be an increasing convergence of these two disciplines as the metabolic pathways leading to the biosynthesis of novel glycan structures are investigated. We herein report the use of hydrophilic interaction liquid chromatography (HILIC) and mass spectrometry in targeted analysis of sugar-nucleotides involved in the flagellin glycosylation process of *Campylobacter*, an important human pathogen and the leading cause of bacterial gastroenteritis worldwide (18, 19).

The flagellin of many bacterial pathogens is an important virulence factor that confers motility and allows colonisation of host cells. In *Campylobacter*, the major structural protein FlaA must be glycosylated for flagellar filament assembly and subsequent motility (20, 21). Therefore, the potential to interrupt the flagellin glycosylation process and prevent motility could offer tremendous opportunities for the development of novel antimicrobial therapies (22). Earlier studies revealed *Campylobacter* flagellin to be extensively glycosylated by the novel sialic acid-like sugar, pseudaminic acid, and a number of other related derivatives, but the precise metabolites and genes leading to the biosynthesis of these novel sugars were poorly understood at the time (23). However, given the structural similarities between pseudaminic acid and sialic acid, as well as the presence of genes that display significant homology to sialic acid biosynthesis genes within the

flagellin glycosylation locus of *Campylobacter*, it was expected that the biosynthesis of pseudaminic acid and related derivatives would involve a similar pathway to that of sialic acid. The biosynthesis of sialic acid in prokaryotes is well established and is known to involve UDP-N-acetylglucosamine (UDP-GlcNAc) and CMP-*N*-acetylneuraminic acid (CMP-Neu5Ac) as key precursors (24). Accordingly, it was decided to target sugar-nucleotides as a likely class of metabolites to be involved in the biosynthesis of pseudaminic acid and its derivatives in *Campylobacter*. To that end, hydrophilic interaction liquid chromatography (HILIC) and mass spectrometry (MS) were used to probe the metabolome of *Campylobacter* for sugar-nucleotides relating to the biosynthesis of pseudaminic acid and its derivatives (16, 25, 26).

HILIC has demonstrated exceptional selectivity towards hydrophilic, polar compounds and has shown tremendous promise in metabolomics research for separating complex mixtures of highly polar compounds including carbohydrates (27–30). Acetonitrile-based mobile phases are generally used in HILIC and this greatly facilitates its use with mass spectrometry for detecting trace-level metabolites as well as the opportunity to employ highly selective scan modes for targeted analysis of pre-defined metabolites or classes of metabolites. In the study of flagellin glycosylation, HILIC–MS and precursor ion scanning for fragment ion(s) characteristic of the nucleotide carriers CMP-, UDP-, GDP- and ADP- was employed to compare the intracellular pool of sugar-nucleotides in wild-type cells and selected isogenic mutants, prepared by site-directed mutagenesis of individual genes on the flagellin glycosylation locus (25, 26, 31). This focused metabolomics approach led to the identification of novel sugar-nucleotides and genes involved in the biosynthesis of unique carbohydrates that glycosylate the flagellin protein of two *Campylobacter* species, namely *Campylobacter jejuni* and *Campylobacter coli* (25, 26). We also demonstrated the potential to employ the focused metabolomics approach as an alternative to proteomics-based methods for structural determination of novel glycans on the flagellar protein (25, 26, 31).

2. Materials

2.1. Molecular Weight Filtering and Solid-Phase Extraction (SPE) of Intracellular Sugar-Nucleotides

1. Microcon YMC 3 kDa MWCO centrifugal filter devices (Millipore, Bedford, MA) (500 μL) or Amicon Ultra 5 kDa MWCO centrifugal filter devices (Millipore) (4 mL). The filters are rinsed with deionised water prior to use (*see* **Note 1**).
2. Visiprep™ Standard Vacuum Manifold (Supelco, Bellefonte, CA).

3. ENVI-Carb™ solid-phase extraction (SPE) cartridges (12 mL, 1 g) (Supelco).
4. The ENVI-Carb™ SPE cartridges are activated before use, as follows:
   a. *SPE activation solvent 1* consisting of acetonitrile (HPLC grade) and 0.1% trifluoroacetic acid (80:20, v/v).
   b. *SPE activation solvent 2* consisting of deionised water obtained from a Milli-Q system.
5. The ENVI-Carb™ SPE cartridges are washed after sample loading, with the following:
   a. *SPE wash solvent 1* consisting of deionised water.
   b. *SPE wash solvent 2* consisting of 5% acetonitrile (HPLC grade).
   c. *SPE wash solvent 3* consisting of 10 mM triethylammonium acetate buffer (TEAA) (Fluka, Buchs, Switzerland) adjusted to pH 7.0 using 0.01% ammonium hydroxide.
6. *SPE elution solvent* consisting of acetonitrile-50 mM triethylammonium acetate buffer (adjusted to pH 7.0 using 1% acetic acid) (35:65, v/v) is used to elute the sugar-nucleotides.

2.2. Hydrophilic Interaction Liquid Chromatography–Mass Spectrometry (HILIC–MS)

1. A *dissolving solvent* consisting of acetonitrile–6.5 mM ammonium acetate buffer (adjusted to pH 5.5 using 1% acetic acid) (70:30, v/v) is used to reconstitute the cell lysates.
2. The HILIC separations are performed using an Agilent 1100 liquid chromatography system (Agilent Technologies, Santa Clara, CA).
3. A TSK-gel Amide80, 250×4.6 mm ID column, 5 μm (Tosoh Bioscience, Montgomeryville, PA) is used for the HILIC separation of the intracellular sugar-nucleotides.
4. HPLC grade acetonitrile (LC solvent B) and 6.5 mM ammonium acetate buffer (adjusted to pH 5.5 using 1% acetic acid, LC solvent A) are used in the preparation of the mobile phase for the HILIC separations. A linear gradient of 20–30% LC solvent A in 12 min is applied and held at 30% LC solvent A for additional 5 min.
5. A QTRAP 4000 mass spectrometer equipped with a Turbo V electrospray source (AB/Sciex, Concord, ON, Canada) is used for selective detection of the intracellular sugar-nucleotides.
6. A QToF Premier mass spectrometer equipped with an Electrospray Ionization (ESI). LockSpray™ modular source

(Waters Corporation, Milford, MA) is used for accurate mass measurements.

7. 10 μg/mL CMP-Neu5Ac prepared in the dissolving solvent and used as a quality control solution (*see* **Note 2**).
8. 1 pmol/μL [Glu^1]-Fibrinopeptide B (Sigma-Aldrich, St. Louis, MO) is used as a calibration solution (positive mode) for high-resolution mass spectrometry experiments.
9. 0.1 μg/mL caesium iodide solution (ULTRA SCIENTIFIC Analytical Solutions, North Kingstown, RI) is used as a calibration solution (negative mode) for high-resolution mass spectrometry experiments.
10. 100 pg/μL leucine enkephalin (Sigma-Aldrich) is used as a lock mass solution.

3. Methods

Sugar-nucleotides are a common feature of many glycoconjugate biosynthesis pathways and were therefore selected as the predefined class of metabolites to target in our focused metabolomics study of the flagellin glycosylation process in *Campylobacter*. Earlier studies had shown that the flagellin protein FlaA of *Campylobacter* is extensively glycosylated by a novel sialic acid-like sugar identified as pseudaminic acid and a number of other related derivatives. However, the precise metabolites and genes leading to the biosynthesis of pseudaminic acid and its derivatives were unknown. Given the structural similarities between pseudaminic acid and sialic acid, as well as the presence of genes that display significant homology to sialic acid biosynthesis genes within the flagellin glycosylation locus of *Campylobacter*, it was expected that the biosynthesis of pseudaminic acid and related derivatives would be similar to that of sialic acid. It is well known that UDP-N-acetylglucosamine (UDP-GlcNAc) and CMP-N-acetylneuraminic acid (CMP-Neu5Ac) are key metabolites in the biosynthetic pathway of sialic acid in prokaryotes. Accordingly, the HILIC–MS and precursor ion scanning method were designed to probe specifically for biosynthetic sugar-nucleotides, with an emphasis on CMP-linked and UDP-linked sugars as potential metabolites relating to the flagellin glycosylation pathway in *Campylobacter*. In addition to screening the metabolome for potential biosynthetic sugar-nucleotides relating to pseudaminic acid and related derivatives, changes to the sugar-nucleotide complement following insertional inactivation of genes on the flagellin glycosylation locus were selectively monitored using HILIC–MS and multiple reaction monitoring.

In addition, structural information on the sugar-nucleotides was gained using high-resolution mass spectrometry to facilitate their identification.

3.1. Solid-Phase Extraction (SPE) of Intracellular Sugar-Nucleotides

1. Cell lysates from wild-type cells and/or isogenic mutants are reconstituted in deionised water (*see* **Note 3**).
2. Cell lysates are filtered through Microcon YMC 3 kDa MWCO (14,000×*g* for 70 min) or Amicon Ultra 5 kDa MWCO (3180×*g* for 60 min) centrifugal filter devices by centrifugation (*see* **Note 4**).
3. ENVI-Carb™ solid-phase extraction (SPE) cartridges are activated using:
   a. 6 mL SPE activation solvent 1
   b. 4 mL SPE activation solvent 2.
4. 2 mL of the filtered cell lysates (samples) is loaded onto the ENVI-Carb™ cartridges.
5. ENVI-Carb™ cartridges are washed using:
   a. 4 mL SPE wash solvent 1
   b. 4 mL SPE wash solvent 2
   c. 4 mL SPE wash solvent 3.
6. Sugar-nucleotides are eluted from the ENVI-Carb® cartridges using:
   a. 2×4 mL SPE elution solvent.
7. Eluents are evaporated to dryness using a SpeedVac concentrator and stored at –20°C until HILIC–MS analysis.

3.2. Hydrophilic Interaction Liquid Chromatography–Mass Spectrometry (HILIC–MS)

3.2.1. Selective Screening of Cell Lysates for Sugar-Nucleotides Involved in the Flagellin Glycosylation Process

1. Samples are reconstituted in 500 μL dissolving solvent and stored in a refrigerator until HILIC–MS analysis. The resolubilisation is performed by sonication for 10 min and vortexing for 3 min.
2. HILIC separations are performed using the Agilent 1100 LC system on a TSK-gel Amide80 (250×4.6 mm ID) column. Prior to the analysis of the cell lysates, the HILIC column is conditioned using a mobile phase consisting of acetonitrile–6.5 mM ammonium acetate buffer pH 5.5 (80:20, v/v) for at least 30 min or when the pressure is equilibrated.
3. The QTRAP 4000 mass spectrometer is used for selective screening of cell lysates for novel sugar-nucleotides relating to the flagellin glycosylation process in *Campylobacter*. For the HILIC–MS experiments, the conditioned HILIC column is connected to the QTRAP 4000 mass spectrometer and the mobile phase effluent is split using a stainless steel

tee such that 200 μL of the solvent is directed to the mass spectrometer and the remaining 800 μL to waste.

4. Prior to sample analysis, the HILIC–MS conditions are optimised using the quality control (QC) sample (*see* **Note 2**). For example, the mass spectrometer is set for precursor ion scanning of *m/z* 322 across a mass range of *m/z* 300–1000 in the negative mode to detect for the CMP-Neu5Ac standard. The position of the electrode, the nebulizer gas, temperature, and electrospray voltage are adjusted until a stable baseline can be achieved. Typically the conditions used are curtain gas 12 psi; electrospray voltage –4.0 kV; temperature 450°C; GS 1 of 50 psi and GS 2 optimised between 45 and 50 psi; declustering potential (DP) –70 V; collision energy (CE) –35 eV; collision cell exit potential (CXP) –15 V; and mobile phase, linear gradient of 20–30% LC solvent A in 12 min is applied and held at 30% LC solvent A for additional 5 min.

5. Samples are assayed specifically for intracellular sugar-nucleotides by HILIC–MS and precursor ion scanning for fragment ion(s) characteristic of the nucleotide carriers CMP-, UDP-, GDP- and ADP-. To that end, the analysis is carried out in the negative ion mode using the optimised conditions. For CMP-linked sugars, precursor ions relating to the fragment ion of CMP (precursor of *m/z* 322) are scanned. For UDP-, GDP- and ADP-linked sugars, three characteristic fragment ions are used i.e. *m/z* 403, 385, 323 for UDP-, *m/z* 442, 424, 362 for GDP-, and *m/z* 426, 408 and 346 for ADP-. Typically, the precursor ions are scanned over *m/z* 300–1000 mass range (**Figs. 12.1a and 12.2a**)i

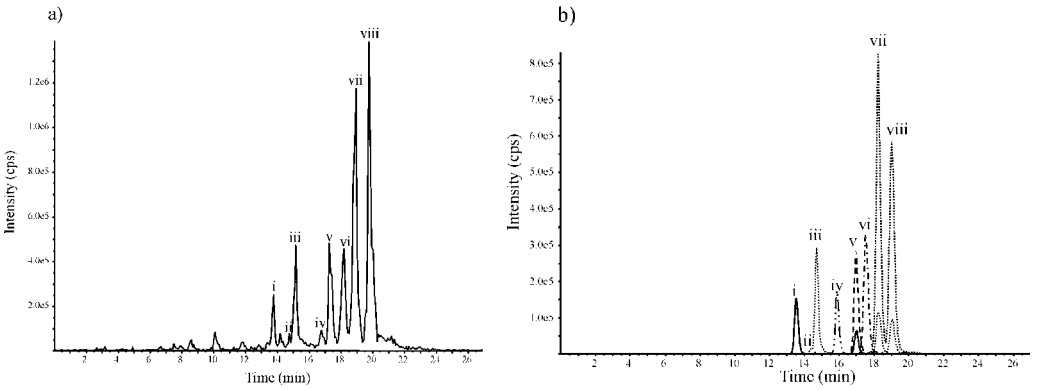


Fig. 12.1. (a) HILIC–MS and precursor ion scanning for fragment ions related to CMP- (*m/z* 322) in cell lysates from *C. jejuni* 11168 and (b) corresponding HILIC–MS and MRM detection of the intracellular CMP-linked sugars in *C. jejuni* 11168. (i) CMP-Pse5(dimethylglycerol)7Ac (*m/z* 712); (ii) CMP-Leg5Ac7Ac (*m/z* 638); (iii) CMP-Pse5Ac7Ac (*m/z* 638); (iv) CMP-Neu5Ac (*m/z* 613); (v) CMP-Pse5(dimethylglycerol)7Am (*m/z* 711); (vi) CMP-Z/E-Leg5MeAM7Ac (*m/z* 651); (vii) CMP-Leg5Am7Ac (*m/z* 637); and (viii) CMP-Pse5Ac7Am (*m/z* 637).

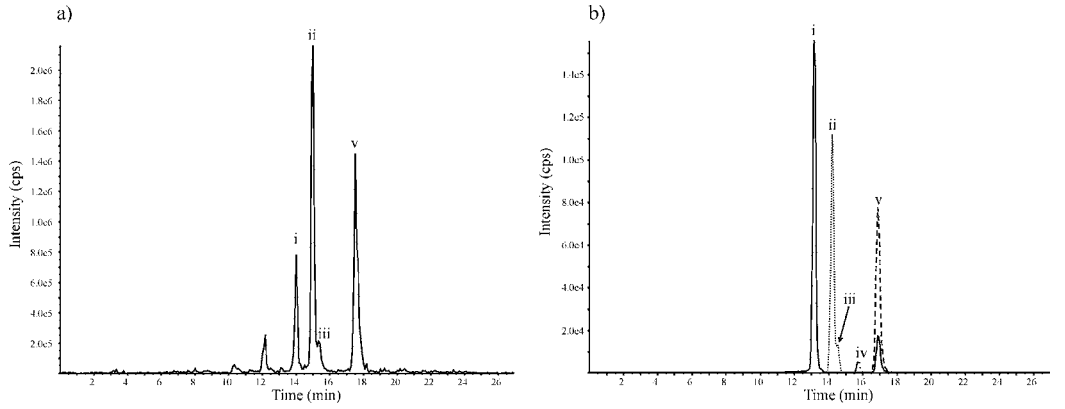


Fig. 12.2. (a) HILIC–MS and precursor ion scanning for fragment ions related to CMP- (*m/z* 322) of cell lysates isogenic mutant Cj1324 and (b) corresponding HILIC–MS and MRM detection of the intracellular CMP-linked sugars in isogenic mutant Cj1324. (i) CMP-Pse5(dimethylglycerol)7Ac (*m/z* 712); (ii) CMP-Leg5Ac7Ac (*m/z* 638); (iii) CMP-Pse5Ac7Ac (*m/z* 638); (iv) CMP-Neu5Ac (*m/z* 613); and (v) CMP-Pse5(dimethylglycerol)7Am (*m/z* 711).

6. For specific comparison of the sugar-nucleotide complement between wild-type cells and isogenic mutants, cell lysates are also assayed for intracellular sugar-nucleotides using the multiple reaction monitoring (MRM) mode. This mode scans for a specific transition from the precursor ion (Q1) to a major fragment ion (Q3) and therefore adds an extra level of selectivity on the targeted metabolites. In the case of sugar-nucleotides, the nucleotide carriers are the major fragment ions resulting from tandem MS. The HILIC–MS conditions are optimised using a QC sample consisting of a mixture of five sugar-nucleotide standards (*see* **Note 5**). Typically the conditions used are: electrospray voltage –4.0 kV; temperature 450°C; GS 1 of 50 psi and GS 2 of 45–50 psi; CE of 35 eV; and a 400 ms dwell time for each MRM transition (**Figs. 12.1b and 12.2b**).
7. For large-scale purification of the intracellular sugar-nucleotides (*see* **Note 3**) the mobile phase effluent is split using a stainless steel tee such that 200 μL of the solvent is directed to the mass spectrometer and the remaining 800 μL to a vial for collection. The purified cell lysates are pooled, evaporated to dryness on a SpeedVac concentrator and stored at –20°C prior to precise structural analysis by NMR.

3.2.2. Structural Elucidation of Novel Biosynthetic Sugar-Nucleotides by High-Resolution Mass Spectrometry

1. Accurate mass measurements and tandem mass spectrometry experiments of novel sugar-nucleotides found to be involved in the flagellin glycosylation process are performed on a Q-ToF Premier mass spectrometer (Waters, Milford, MA) to verify their structures. Prior to sample analysis, calibration of the Q-ToF Premier mass spectrometer is performed using the MS/MS fragment ions of [Glu^1]-Fibrinopeptide

B in the positive mode and a lock mass solution of leucine enkephalin. A typical mass accuracy of 2 ppm or less is generally obtained and deemed acceptable. In the negative mode, calibration is carried out using a solution of sodium caesium iodide and a lock mass solution of leucine enkephalin.

2. Typical HILIC–MS conditions in the positive mode are as follows (1): electrospray voltage 3.0 kV; sample cone 30 V; extraction cone 3 V; ion guide 2 V; source temperature 120°C; desolvation temperature of 350°C; cone gas flow of 25 L/h; desolvation gas flow of 600 L/h; collision energy of 4 eV; and LM resolution and HM resolution of 4.7 V and 15 V, respectively. The HILIC solvents, gradient and flow-split conditions are the same as reported in the case of 4000 QTRAP mass spectrometer.

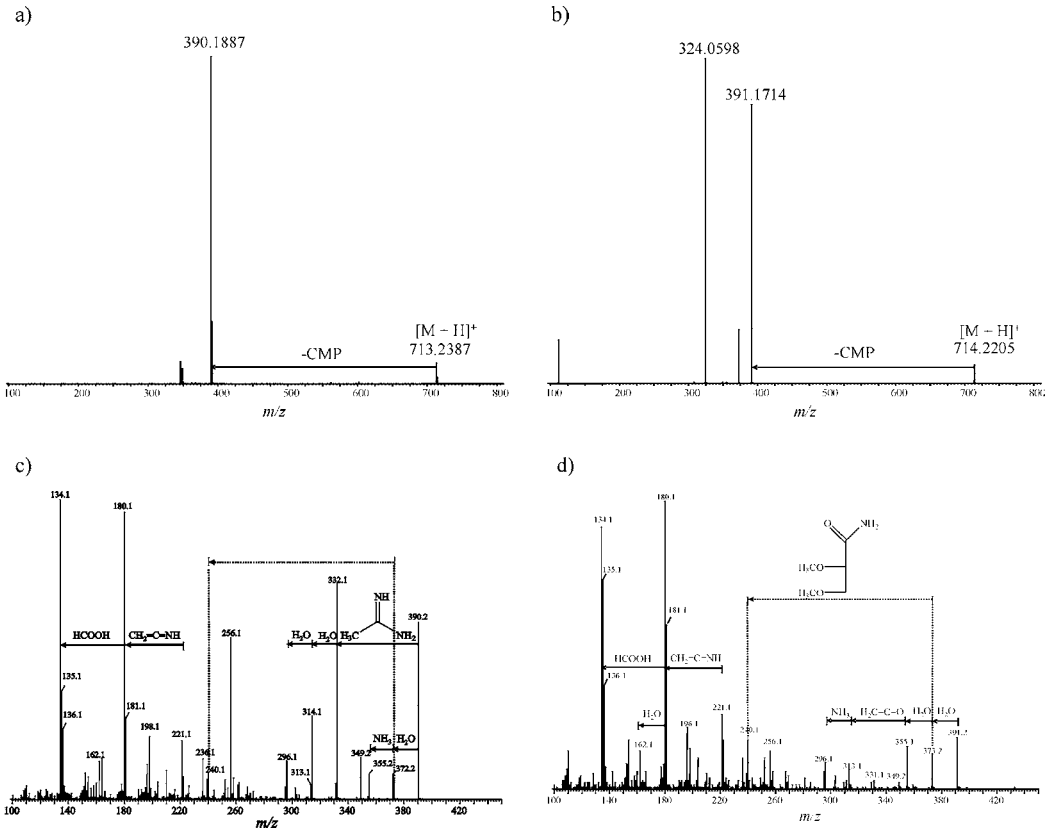


Fig. 12.3. Accurate mass measurements of two unknown CMP-sugars detected in *C. jejuni* 11168. (a) MS/MS of the first unknown CMP-sugar detected as $[M+H]^+$ ions *m/z* 713 showing a novel glycan moiety at *m/z* 390; (b) MS/MS of the second unknown CMP-sugar detected as $[M+H]^+$ ions *m/z* 714 showing a novel glycan moiety at *m/z* 391; (c) second-generation product ion spectrum of the glycan moiety at *m/z* 390 showing possible neutral loss of a 2,3-di-*O*-methylglyceramide ($C_5H_{11}NO_3$); and (d) second-generation product ion spectrum of the glycan moiety at *m/z* 391 showing possible neutral loss of a 2,3-di-*O*-methylglyceramide ($C_5H_{11}NO_3$).

3. Accurate mass measurements of the intact sugar-nucleotide ions are performed in both the negative and positive ion modes. For structural information on the carbohydrate moiety of the sugar-nucleotides, experiments are performed in the positive mode. The cone voltage is increased to 50 V to promote in-source fragmentation and the resulting sugar oxonium ions are selected as precursor ions for tandem MS (CE = 20 eV) (**Fig. 12.3**).
4. The MS data are acquired using centroid mode and all data acquisition is done using MassLynx version 4.1.
5. Elemental composition is performed using a mass tolerance of 10 ppm and is sorted by i-FIT score assigned by MassLynx based on the isotopic patterns of the target ions. The elements used for searching are limited to C, H, N, O and P.

4. Notes

1. Deionised water was obtained from a Milli-Q system and has a resistivity of 18.2 MΩ-cm.
2. 10 μg/mL UDP-glucose was employed as a quality control solution during selective screening of the metabolome for novel UDP-linked precursors. Acceptable criteria for the QC are as follows: CMP-Neu5Ac, retention time 16.2 min (±0.2 min), intensity of 1.0×10^5 cps; UDP-Glc, retention time of 14.8 min, intensity of 1.0×10^5 cps.
3. Dried cell lysates are typically obtained from a 500-mL culture of wild-type cells or isogenic mutants and reconstituted in 10 mL deionised water. For large-scale purification of the sugar-nucleotide intermediates, up to 10 L of culture of the bacterial cells was necessary and cell lysates were reconstituted in 200 mL deionised water. The cell lysates were stored at –20°C until further analysis.
4. The cell lysates were molecular weight filtered using either the Microcon YMC 3 kDa MWCO centrifugal filter devices for volumes up to 500 μL, or the Amicon Ultra 5 kDa MWCO centrifugal filter devices for volumes up to 4 mL. For *Campylobacter*, 1 mL of the reconstituted cell lysates from wild-type cells or the isogenic mutants is generally sufficient for the selective analysis of intracellular sugar-nucleotides. Depending on the concentration of the intracellular sugar-nucleotides in other bacterial pathogens, it may be necessary to use more cell lysates for the analysis. After centrifugation, the cell lysate is further concentrated

to another 2.5-fold, i.e. 500 μL of filtrate in the Microcon YMC centrifugal filter device is dried and reconstituted in 200 μL of *dissolving solvent*.

5. The mixture of five sugar-nucleotide standards consisting of TDP-Glc, ADP-Glc, UDP-Glc, CMP-Neu5Ac and GDP-Man is monitored by MRM in the negative mode. The MRM transitions, expected retention times and minimum intensities are: (1) TDP-Glc, m/z 563 > 383, 12.9 min, 1.0×10^4 cps; (2) ADP-Glc, m/z 588 > 346, 14.1 min, 1.0×10^5 cps; (3) UDP-Glc, m/z 565 > 385, 14.8 min, 1.0×10^4 cps; (4) CMP-Neu5Ac, m/z 613 > 322, 16.2 min, 5.0×10^4 cps and (5) GDP-Man, m/z 604 > 424, 17.6 min, 1.0×10^4 cps. A variation of ±0.2 min of the retention times is deemed acceptable.

Acknowledgements

The authors wish to thank Dr. Susan M Logan for providing cell lysates from wild-type *Campylobacter* sp. and isogenic mutants for the focused metabolomics studies.

References

1. Fiehn, O. (2002) Metabolomics–the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155–171.
2. Griffin, J. L. (2004) Metabolic profiles to define the genome: can we hear the phenotypes? *Philos. Trans. R Soc. Lond. B Biol. Sci.* **359**, 857–871.
3. Oliver, S. G., Winson, M. K., Kell, D. B., and Baganz, F. (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**, 373–378.
4. Rochfort, S. (2005) Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research. *J. Nat. Prod.* **68**, 1813–1820.
5. Dunn, W. B., Bailey, N. J., and Johnson, H. E. (2005) Measuring the metabolome: current analytical technologies. *Analyst* **130**, 606–625.
6. Lenz, E. M. and Wilson, I. D. (2007) Analytical strategies in metabonomics. *J. Proteome Res.* **6**, 443–458.
7. Gao, P., Shi, C., Tian, J., Shi, X., Yuan, K., Lu, X., and Xu, G. (2007) Investigation on response of the metabolites in tricarboxylic acid cycle of *Escherichia coli* and *Pseudomonas aeruginosa* to antibiotic perturbation by capillary electrophoresis. *J. Pharm. Biomed. Anal.* **44**, 180–187.
8. Guan, X. L. and Wenk, M. R. (2006) Mass spectrometry-based profiling of phospholipids and sphingolipids in extracts from *Saccharomyces cerevisiae*. *Yeast* **23**, 465–477.
9. Sabatine, M. S., Liu, E., Morrow, D. A., Heller, E., McCarroll, R., Wiegand, R., Berriz, G. F., Roth, F. P., and Gerszten, R. E. (2005) Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* **112**, 3868–3875.
10. Soga, T., Ueno, Y., Naraoka, H., Ohashi, Y., Tomita, M., and Nishioka, T. (2002) Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal. Chem.* **74**, 2233–2239.
11. Metzler, D. E. (2003) Some pathways of carbohydrate metabolism. In: Biochemistry: The Chemical Reactions of Living Cells (Metzler, D. E. and Metzler, C. M. Eds), Academic Press, San Diego, pp. 1129–1178.
12. Turnock, D. C. and Ferguson, M. A. (2007) Sugar nucleotide pools of *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major*. *Eukaryot Cell* **6**, 1450–1463.

13. Zhu, B. C., Drake, R. R., Schweingruber, H., and Laine, R. A. (1995) Inhibition of glycosylation by amphomycin and sugar nucleotide analogs PP36 and PP55 indicates that *Haloferax volcanii* beta-glucosylates both glycoproteins and glycolipids through lipid-linked sugar intermediates: evidence for three novel glycoproteins and a novel sulfated dihexosyl-archaeol glycolipid. *Arch. Biochem. Biophys.* **319**, 355–364.
14. Pfoestl, A., Hofinger, A., Kosma, P., and Messner, P. (2003) Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-galactose in *Aneurinibacillus thermoaerophilus* L420-91T. *J. Biol. Chem.* **278**, 26410–26417.
15. Schirm, M., Soo, E. C., Aubry, A. J., Austin, J., Thibault, P., and Logan, S. M. (2003) Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* **48**, 1579–1592.
16. Soo, E. C., Aubry, A. J., Logan, S. M., Guerry, P., Kelly, J. F., Young, N. M., and Thibault, P. (2004) Selective detection and identification of sugar nucleotides by CE-electrospray-MS and its application to bacterial metabolomics. *Anal. Chem.* **76**, 619–626.
17. Liu, X., McNally, D. J., Nothaft, H., Szymanski, C. M., Brisson, J. -R., and Li, J. (2006) Mass spectrometry-based glycomics strategy for exploring N-linked glycosylation in eukaryotes and bacteria. *Anal. Chem.* **78**, 6081–6087.
18. Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M., and Tauxe, R. V. (1999) Food-related illness and death in the United States. *Emerg. Infect. Dis.* **5**, 607–625.
19. Blaser, M. J. (1997) Epidemiologic and clinical features of *Campylobacter jejuni* infections. *J. Infect. Dis.* **176**(Suppl 2), S103–105.
20. Guerry, P., Doig, P., Alm, R. A., Burr, D. H., Kinsella, N., and Trust, T. J. (1996) Identification and characterization of genes required for post-translational modification of *Campylobacter coli* VC167 flagellin. *Mol. Microbiol.* **19**, 369–378.
21. Logan, S. M. (2006) Flagellar glycosylation – a new component of the motility repertoire? *Microbiology* **15 2**, 1249–1262.
22. Alksne, L. E. (2002) Virulence as a target for antimicrobial chemotherapy. *Expert Opin. Investig. Drugs* **11**, 1149–1159.
23. Thibault, P., Logan, S. M., Kelly, J. F., Brisson, J. -R., Ewing, C. P., Trust, T. J., and Guerry, P. (2001) Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *J. Biol. Chem.* **276**, 34862–34870.
24. Angata, T. and Varki, A. (2002) Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem. Rev.* **102**, 439–469.
25. McNally, D. J., Aubry, A. J., Hui, J. P., Khieu, N. H., Whitfield, D., Ewing, C. P., Guerry, P., Brisson, J. -R., Logan, S. M., and Soo, E. C. (2007) Targeted metabolomics analysis of *Campylobacter coli* VC167 reveals legionaminic acid derivatives as novel flagellar glycans. *J. Biol. Chem.* **282**, 14463–14475.
26. McNally, D. J., Hui, J. P., Aubry, A. J., Mui, K. K., Guerry, P., Brisson, J. -R., Logan, S. M., and Soo, E. C. (2006) Functional characterization of the flagellar glycosylation locus in *Campylobacter jejuni* 81-176 using a focused metabolomics approach. *J. Biol. Chem.* **281**, 18489–18498.
27. T'Kindt, R., Storme, M., Deforce, D., and Van Bocxlaer, J. (2008) Evaluation of hydrophilic interaction chromatography versus reversed-phase chromatography in a plant metabolomics perspective. *J. Sep. Sci.* **31**, 1609–1614.
28. Antonio, C., Larson, T., Gilday, A., Graham, I., Bergstrom, E., and Thomas-Oates, J. (2008) Hydrophilic interaction chromatography–electrospray mass spectrometry analysis of carbohydrate-related metabolites from *Arabidopsis thaliana* leaf tissue. *Rapid Commun. Mass. Spectrom.* **22**, 1399–1407.
29. Bajad, S. U., Lu, W., Kimball, E. H., Yuan, J., Peterson, C., and Rabinowitz, J. D. (2006) Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J. Chromatogr. A* **1125**, 76–88.
30. Kiefer, P., Portais, J. C., and Vorholt, J. A. (2008) Quantitative metabolome analysis using liquid chromatography-high-resolution mass spectrometry. *Anal. Biochem.* **382**, 94–100.
31. Logan, S. M., Hui, J. P., Vinogradov, E, Aubry, A. J., Melanson, J. E., Kelly, J. F., Nothaft, H., and Soo, E. C. (2009) Identification of novel carbohydrate modifications on *Campylobacter jejuni* 11168 flagellin using metabolomics-based approaches. *FEBS J.* **276**, 1014–1023.

Chapter 13

Characterization of Lipid-Linked Oligosaccharides by Mass Spectrometry

Christopher W. Reid, Jacek Stupak, and Christine M. Szymanski

Abstract

N- Glycosylation of proteins is recognized as one of the most common post-translational modifications. Until recently it was believed that N-glycosylation occurred exclusively in eukaryotes until the discovery of the general protein glycosylation pathway (Pgl) in *Campylobacter jejuni*. We have developed a new glycomics strategy based on lectin-affinity capture of lipid-linked oligosaccharides (LLOs) coupled to capillary electrophoresis mass spectrometry. The LLO intermediates of the *C. jejuni* Pgl pathway were used to validate the methodology and to better characterize the bacterial model system for protein N-glycosylation. This method provides a rapid, non-radioactive approach for the characterization of intermediates in polysaccharide biosynthesis and is a useful tool for glycoengineering efforts in bacteria.

Key words: *Campylobacter jejuni* , affinity-capture, capillary electrophoresis, lipid-linked oligosaccharide, protein N-glycosylation, metabolomics, glycobiology.

1. Introduction

Traditionally believed to occur solely in Eukarya, it is now evident that both Bacteria and Archaea are capable of protein glycosylation (1–4). In prokaryotes, it appears that glycosylation leads to a greater diversity of glycan compositions and structures than those found in eukaryotic cells. There is a burgeoning interest in protein glycosylation in part due to the increasing frequency with which this modification is seen in pathogenic bacterial species (5). Glycans are generally attached to serine or threonine (O-glycosylation) or to asparagine (N-glycosylation) residues. In the last decade, protein N-glycosylation has been identified in

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_13,

the intestinal pathogen *Campylobacter jejuni* (6), and O-linked protein glycosylation in pathogens such as *Neisseria meningitidis* (7), *N. gonorrhoeae* (7), *Helicobacter pylori* (8), and *Pseudomonas aeruginosa* (9). Two different mechanisms for protein glycosylation can be distinguished by the mode in which the glycans are transferred to proteins. The first mechanism involves the transfer of carbohydrates directly from nucleotide-activated sugars to acceptor proteins. This mechanism is typical of O-glycosylation in eukaryotes (10) and flagellar glycosylation in several bacterial species (9–11). In the alternate mechanism, an oligosaccharide is pre-assembled on a polyisoprene carrier before being transferred en bloc to protein acceptors by an oligosaccharyltransferase. This mechanism is utilized in N-glycosylation in eukaryotes, *C. jejuni* (4), and archaea (12), as well as O-glycosylation in *Pseudomonas aeruginosa* and *Neisseria meningitidis* (13).

The N-glycosylation system of *C. jejuni* is encoded by the *pgl* operon (6). The glycan in the *C. jejuni* system is a conserved heptasaccharide with the structure GalNAc-α1,4-GalNAc-α1,4-[Glc-β1,3]-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac2,4diNAc-β1, where Bac2,4diNAc is 2,4-diacetamido-2,4,6-trideoxyglucopyranose (**Fig. 13.1**) (14). The similarities between the eukaryotic and bacterial N-glycan pathways include the assembly of an oligosaccharide on a polyisoprenyl carrier anchored in the membrane and transfer of the oligosaccharide *en bloc* to asparagine residues of proteins with sequons containing Asn-X-Ser/Thr, where X is any amino acid except for Pro, where bacteria also have a requirement for Asp/Glu in the -2 position (15). Furthermore, the *C. jejuni* oligosaccharyltransferase PglB shows modest homology to Stt3p, which is an essential subunit in the eukaryotic OST complex, and both contain the conserved catalytic domain, WWDYG (4). Since the N-glycosylation pathway in *C. jejuni* is not required for bacterial viability and possesses many similarities with the eukaryotic counterpart, it provides an ideal system to study protein N-glycosylation.

Traditionally, the detection of lipid-linked oligosaccharides (LLOs) has been performed using a variety of techniques including metabolic radiolabeling (16), solid-state analysis of LLOs with conjugated lectins (17), an OST assay with ^{125}I-labeled acceptor peptides, or fluorophore-assisted carbohydrate electrophoresis (FACE) (18).

Capillary electrophoresis (CE) is a high-resolution technique for the separation of complex biological mixtures and has been widely applied to biological analyses (19). The coupling of CE with mass spectrometry provides a powerful approach for rapid identification of target analytes present at trace levels in biological matrices and for structural characterization of complex biomolecules (20–22). Here we report a method for the detection of LLOs from bacterial samples. Affinity capture-mass

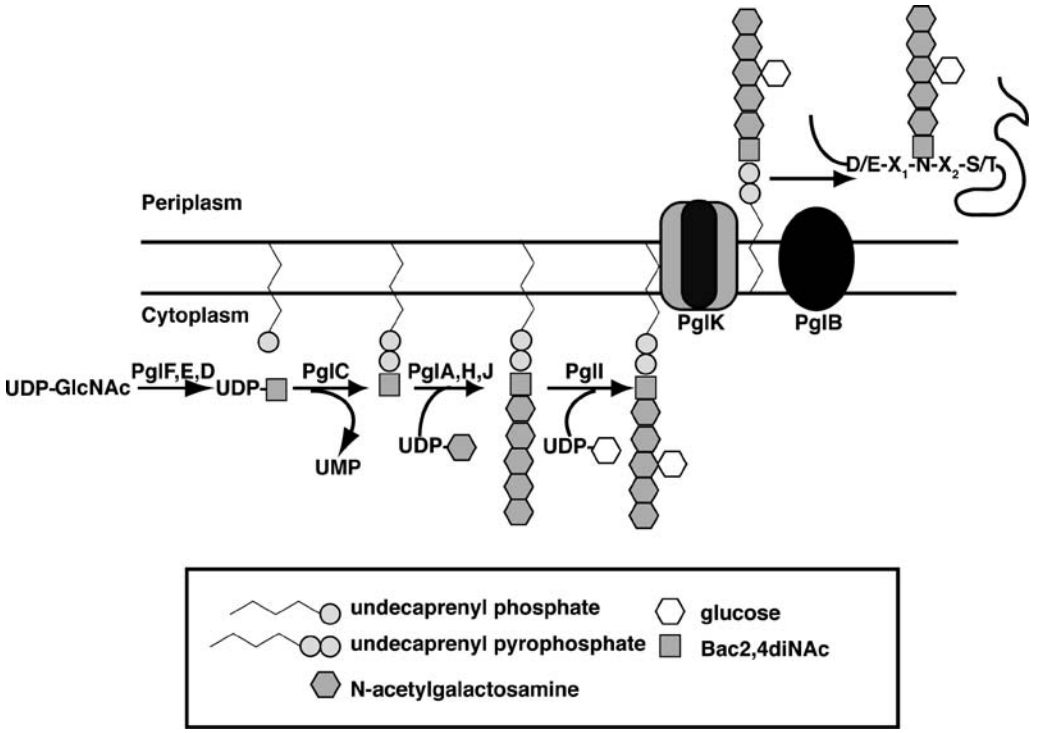


Fig. 13.1. The *Campylobacter jejuni* N-glycosylation pathway. The heptasaccharide is synthesized on the cytoplasmic face of the plasma membrane on undecaprenyl-phosphate. Once synthesis is complete, the heptasaccharide is then flipped into the periplasmic space by the flippase, PglK, where the heptasaccharide is transferred to acceptor proteins by the action of the oligosaccharyltransferase, PglB.

spectrometry has been applied to the analysis of LLOs from the *C. jejuni* N-glycan pathway. Affinity-capture with agarose immobilized soybean agglutinin (SBA) was chosen for selective enrichment of LLOs. Lectins have been used successfully to characterize glycolipids from a variety of biological sources (23–26), and SBA has previously been shown to effectively bind the *C. jejuni* N-glycan (27).

2. Materials

2.1. Cell Culture and Lipid-Linked Oligosaccharide Extraction

1. Brain-Heart Infusion (BHI) (Gibco/BRL, Bethesda, MD).
2. Mueller–Hinton (MH) supplemented with 1.5% agar (Gibco/BRL).
3. 2× YT medium (Sigma, Oakville, Canada).
4. Kanamycin (Sigma) dissolved in water to 30 mg/mL and filter sterilized for stock solution.

5. Chloroform (EMD Chemicals, Gibbstown, NJ).
6. Methanol and 95% Ethanol (EMD Chemicals).
7. Milli-Q purified water (Millipore, Billerica, MA).
8. Omni-Mixer (Thermo Scientific, Waltham, MA).

2.2. Enrichment of LLOs from Biological Samples

1. Agarose-immobilized soybean agglutinin (SBA) (EY Laboratories, San Mateo, CA).
2. Ethanol:methanol (9:1).
3. Lectin-binding buffer: 10 mM Hepes, pH 7.4, 3.4 mM EDTA, 150 mM NaCl.
4. Lectin elution buffer: 25 mM NaH_2PO_4, 0.2 M galactose.

3. Methods

3.1. Culture Conditions

3.1.1. Growth of C. jejuni

1. Each flask of BHI broth (5×1.5 L) is inoculated with one confluent plate of cells as previously described (28, 29) (*see* **Note 1**).
2. Cultures are grown shaking at 37°C under microaerobic conditions (5% O_2, 10% CO_2, 85% N_2) for 18 h (*see* **Note 2**).
3. Cells are harvested by centrifugation (7000×*g*, 15 min, 4°C) and washed twice with methanol to drive out residual water and then either extracted immediately or frozen at –80°C for later use.

3.1.2. Lipid Extraction

1. The cell pellets are extracted with 2:1 $CHCl_3$:MeOH (25 mL) (*see* **Note 3**).
2. The cell debris is removed by centrifugation in Corex glass tubes (500×*g*, 15 min, 4°C) and the supernatants retained.
3. Repeat Steps 1–2 (1×).
4. Cell pellets are extracted with 10:20:3 $CHCl_3$:MeOH:H_2O (25 mL). The cell debris is removed by centrifugation in Corex glass tubes (500×*g*, 15 min, 4°C), and the supernatants are retained and pooled with the organic extracts from Step 2.
5. Repeat Step 4 (1×).
6. The combined organic extracts are evaporated to dryness on a Buchi rotary evaporator and immediately subjected to affinity-capture enrichment of LLOs.

3.1.3. Quantitative Assessment of SBA Affinity for Polyisoprene Oligosaccharides

1. Radioactively labeled Und-PP-disaccharides, tri-saccha rides, hexa-saccharides, and heptasaccharides are prepared as previously described (30–32).
2. Then 0.13 nmol of Und-PP-sugar (0.15–0.20 μg depending on the sugar size) is resuspended in 40 μL of 9:1 EtOH/MeOH with vortexing and sonication (in a sonicating water bath), then diluted with 360 μL of lectin-binding buffer (*see* above).
3. Twenty microliters of SBA resin, pre-equilibrated with binding buffer, is added and the mixture is incubated for 2 h. Following incubation the resin is collected by centrifugation (500×*g*, 1 min). The supernatant is removed and the resin washed with another 400 μL of binding buffer.
4. The bound radiolabeled LLOs are eluted with lectin elution buffer (1 × 400 μL, 2 × 200 μL) after incubating for 15 min and are then combined.
5. The eluted material is separated by HPLC and radioactivity measured by scintillation counting as previously described (30) (**Table 13.1**). All reactions are done in duplicate with an associated error of 1–2% for each of the values. Results show that SBA-agarose is capable of recognizing LLOs as small as Und-PP-disaccharide.

Table 13.1
Assessment of binding affinities of SBA-agarose with radiolabeled und-PP-linked sugars

Substrate	Percentage of unbound	Percentage of bound
Und-PP-disaccharide	57.05447[a]	42.94553
Und-PP-trisaccharide	42.57978	57.42022
Und-PP-hexasaccharide	17.80852	82.19148
Und-PP-heptasaccharide	14.77273	85.22727

[a]Experiments performed in duplicate with an associated error of 1–2% for all samples.

3.1.4. Affinity-Capture of Bacterial LLOs

1. The concentrated lipid extract is dissolved in a minimum of 9:1 EtOH:MeOH by vortexing and sonication, then diluted 10-fold with lectin-binding buffer. The lipid extract should form a uniform suspension.
2. The lipid suspension is incubated with 100 μL SBA-agarose for 1 h at room temperature with mixing on a Clay Adams Nutator (*see* **Note 4**).
3. The resin is collected by centrifugation (500×*g*, 1 min, 4°C), and washed three times with binding buffer.

4. The bound LLOs are eluted from the SBA-agarose with lectin elution buffer as per the manufacturer's instructions and used without further purification (*see* **Note 5**).

3.1.5. CE–MS Analysis of Lipid-Linked Oligosaccharides

1. Mass spectra are acquired using a 4000 QTrap mass spectrometer (Applied Biosystems/Sciex, Concord, ON, Canada) equipped with a turbo spray ion source. Spectra are processed using the Analyst software package.
2. CE is performed using a Prince CE system (Prince Technologies, Netherlands). The CE separation is obtained on a 90-cm length of bare fused-silica capillary (365 μm OD×50 μm ID) with CE–MS coupling using a liquid sheath-flow interface and isopropanol:methanol (2:1) as the sheath liquid. An aqueous buffer consisting of 10 mM ammonium acetate is used for all experiments (*see* **Note 6**). A 10-s injection of sample (about 110 nL) was used in each analysis.
3. Precursor ion scanning for Und-PP ($[M\text{-}H\text{-}H_2O]^-$, *m/z* 907.6) is performed in negative-ion mode (**Fig. 13.2 A and C).** Parameters for the mass spectrometer are summarized in **Table 13.2** (*see* **Note 7**).
4. For MS/MS experiments in the positive ion mode, the declustering potential on the instrument is set at 400 V in order to produce oligosaccharide ions without the Und-PP linker, which limits detailed structural characterization. The instrument settings are summarized in **Table 13.3**. Sodiated molecular ions with masses corresponding to the sugar portion of the original Und-PP-linked oligosaccharides, observed previously in the negative ion mode, are selected for collision-induced dissociation in the positive ion mode to obtain detailed and complete sequence information (**Fig. 13.2D**).

3.2. Comment on Limitations and Future Applications

Results from this study provide new insight into the function and control of the *C. jejuni* N-glycan pathway, which will aid in future glycoengineering efforts in bacteria. This method can be applied to the characterization and validation of the synthesis of designer oligosaccharides at the lipid intermediate level in genetically engineered bacterial glycofactories. This is a significant improvement over traditional methods for analyzing glycan biosynthesis at the glycoprotein level, which relies on the assumption that the oligosaccharyltransferase recognizes all LLO species with the same affinity. This methodology has the potential to improve our understanding of glycoconjugate biosynthesis pathways in other bacteria, identify novel enzyme activities, and can be expanded to the investigation of eukaryotic systems. While

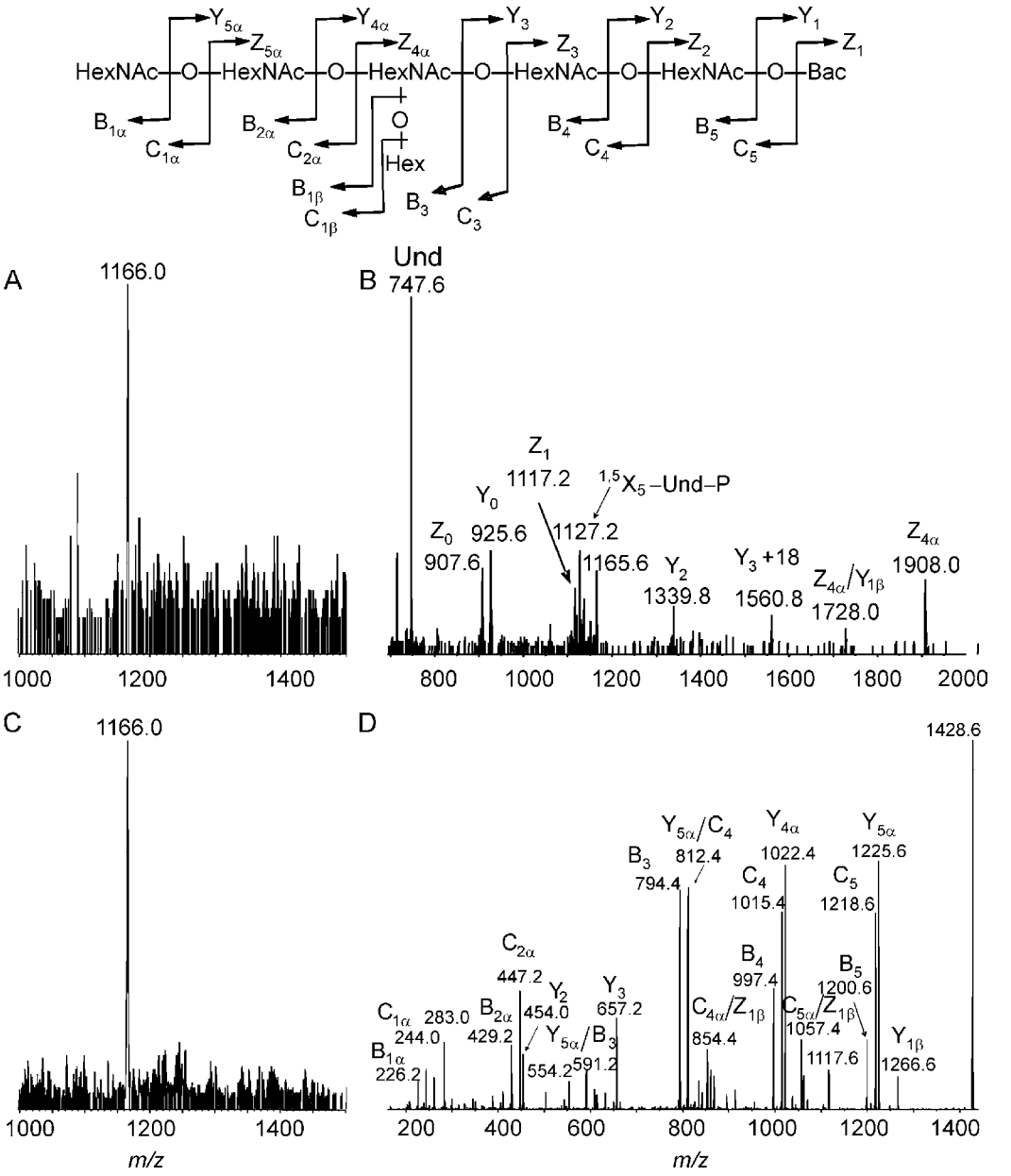


Fig. 13.2. CE–MS negative ion mode precursor ion scans for Und-PP (*m/z* 907.6) in glycolipid-enriched extracts of *C. jejuni pgl* mutants. Analysis of *pglB* (A) and *pglK* (C) mutants showing accumulation of Und-PP linked heptasaccharide (*m/z* 1166.0); negative ion mode MS/MS spectrum of *pglB* showing loss of Und-PP (B). MS/MS analysis of *m/z* 1166.0 in the *pglK* mutant in the positive ion mode with high orifice voltage to cleave off the lipid portion of the molecule, confirming the structure of the *N*-glycan (D). (Reproduced from reference (32) with permission from the American Chemical Society.)

this method is not as sensitive as radioisotope-based methods and requires the prior knowledge of the glycan under investigation, it provides a useful new tool for the study of glycan biosynthetic pathways.

Table 13.2
Mass spectrometer settings for precursor ion scanning for und-PP in the negative ion mode

Parameter	Setting
Scan type	Precursor ion
Polarity	Negative
Precursor (Und-PP)	907.5 amu
Resolution Q1	Low
Resolution Q2	Low
MR Pause	5.0070 ms
Step size	0.50 amu
Scan range	
Start	800 amu
Stop	1600 amu
Collision energy	
Start	–70
Stop	–77
Ion spray	
Cur	12.00
IS	–5300
TEM	60.00
GS1	2.00
GS2	0.00
DP (orifice voltage)	–120.00

Table 13.3
Mass spectrometer settings for MS/MS analysis of free oligosaccharides in the positive ion mode

Parameter	Setting
Scan type	Enhanced product ion (EPI)
Polarity	Positive
Ion source	Turbo spray
Resolution Q1	Unit
Scan rate	4000 amu/s
MR pause	5.0070 ms
LIT fill time	20.00 ms
Dynamic fill time	On

(continued)

Table 13.3 (continued)

Parameter	Setting
Step size	0.50 amu
Ion spray	
CUR	12.00
IS	5200.00
TEM	60.00
GS1	2.00
MSMS (positive)	
CAD	High
DP (orifice voltage)	370.00
Collision energy (CE)	125.00

4. Notes

1. For the growth of *C. jejuni* 11168 and *pgl* mutants, BHI broth is inoculated with cells that are passaged twice.
2. Total cell density of *C. jejuni* cells harvested is on average between 10^{12} and 10^{13} c.f.u.
3. All solvents for lipid extraction are made fresh and stored in dark glass bottles.
4. By switching the lectin employed, this method can be applied to the analysis of a wide variety of LLOs of bacterial and potentially eukaryotic origin.
5. High concentrations of galactose in the eluting buffer can cause signal suppression in the mass spectrometer.
6. Separation of structurally similar LLOs is not possible under these CE conditions. The CE is used predominantly to separate analyte from interfering salt ions.
7. MS/MS in the negative ion mode is useful for gaining information on the lipid carrier (**Fig. 13.2 B**), but does not provide adequate information about the attached oligosaccharide so then MS/MS analysis in the positive ion mode is necessary.

References

1. Abu-Qarn, M., Eichler, J., and Sharon, N. (2008) Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr. Opin. Struct. Biol.* **18**, 544–550.
2. Eichler, J. and Adams, M. W. W. (2005) Posttranslational Protein Modification in Archaea. *Microbiol. Mol. Biol. Rev.* **69**, 393–425.
3. Messner, P. (2004) Prokaryotic glycoproteins: unexplored but important. *J. Bacteriol.* **186**, 2517–2519.
4. Weerapana, E. and Imperiali, B. (2006) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **16**, 91R–R101.
5. Szymanski, C. M. and Wren, B. W. (2005) Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Microbiol.* **3**, 225–237.
6. Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J., and Guerry, P. (1999) Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* **32**, 1022–1030.
7. Banerjee, A. and Ghosh, S. K. (2003) The role of pilin glycan in neisserial pathogenesis. *Mol. Cell Biochem.* **253**, 179–190.
8. Schirm, M., Soo, E. C., Aubry, A. J., Austin, J., Thibault, P., and Logan, S. M. (2003) Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* **48**, 1579–1592.
9. Castric, P., Cassels, F. J., and Carlson, R. W. (2001) Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan. *J. Biol. Chem.* **276**, 26479–26485.
10. Spiro, R. G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R–R456.
11. Power, P. M., Seib, K. L., and Jennings, M. P. (2006) Pilin glycosylation in *Neisseria meningitidis* occurs by a similar pathway to wzy-dependent O-antigen biosynthesis in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **347**, 904–908.
12. Chaban, B., Voisin, S., Kelly, J., Logan, S. M., and Jarrell, K. F. (2006) Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked glycosylation pathways in Archaea. *Mol. Microbiol.* **61**, 259–268.
13. Faridmoayer, A., Fentabil, M. A., Mills, D. C., Klassen, J. S., and Feldman, M. F. (2007) Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. *J. Bacteriol.* **189**, 8088–8098.
14. Young, N. M., Brisson, J.-R., Kelly, J., Watson, D. C., Tessier, L., Lanthier, P. H., Jarrell, H. C., Cadotte, N., St Michael, F., Aberg, E., and Szymanski, C. M. (2002) Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J. Biol. Chem.* **277**, 42530–42539.
15. Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., Mills, D. C., Watson, D. C., Hernandez, M., Kelly, J. F., Wacker, M., and Aebi, M. (2006) Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25**, 1957–1966.
16. Spiro, R. G., Spiro, M. J., and Bhoyroo, V. D. (1976) Lipid-saccharide intermediates in glycoprotein biosynthesis. II. Studies on the structure of an oligosaccharide-lipid from thyroid. *J. Biol. Chem.* **251**, 6409–6419.
17. Kelleher, D. J., Karaoglu, D., and Gilmore, R. (2001) Large-scale isolation of dolichol-linked oligosaccharides with homogeneous oligosaccharide structures: determination of steady-state dolichol-linked oligosaccharide compositions. *Glycobiology* **11**, 321–333.
18. Gao, N. and Lehrman, M. A. (2002) Analyses of dolichol pyrophosphate-linked oligosaccharides in cell cultures and tissues by fluorophore-assisted carbohydrate electrophoresis. *Glycobiology* **12**, 353–360.
19. Moini, M. (2002) Capillary electrophoresis mass spectrometry and its application to the analysis of biological mixtures. *Anal. Bioanal. Chem.* **373**, 466–480.
20. Li, J., Wang, Z., and Altman, E. (2005) In-source fragmentation and analysis of polysaccharides by capillary electrophoresis–mass spectrometry. *Rapid Commun. Mass Spectrom.* **19**, 1305–1314.
21. Li, J. and Richards, J. C. (2007) Application of capillary electrophoresis mass spectrometry to the characterization of bacterial lipopolysaccharides. *Mass Spectrom. Rev.* **26**, 35–50.
22. Li, Y. L., Su, X., Stahl, P. D., and Gross, M. L. (2007) Quantification of diacylglycerol molecular species in biological samples by electrospray ionization mass spectrometry after one-step derivatization. *Anal. Chem.* **79**, 1569–1574.
23. Curatolo, W., Yau, A. O., Small, D. M., and Sears, B. (1978) Lectin-induced agglutination of phospholipid/glycolipid vesicles. *Biochemistry* **17**, 5740–5744.

24. Smith, D. F. (1983) Glycolipid-lectin interactions: detection by direct binding of 125I-lectins to thin layer chromatograms. *Biochem. Biophys. Res. Commun.* **115**, 360–367.
25. Smith, D. F. and Torres, B. V. (1989) Lectin affinity chromatography of glycolipids and glycolipid-derived oligosaccharides. *Methods Enzymol.* **179**, 30–45.
26. Torres, B. V., McCrumb, D. K., and Smith, D. F. (1988) Glycolipid-lectin interactions: reactivity of lectins from *Helix pomatia*, *Wisteria floribunda*, and *Dolichos biflorus* with glycolipids containing N-acetylgalactosamine. *Arch. Biochem. Biophys.* **262**, 1–11.
27. Linton, D., Allan, E., Karlyshev, A. V., Cronshaw, A. D., and Wren, B. W. (2002) Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. *Mol. Microbiol.* **43**, 497–508.
28. Kelly, J., Jarrell, H., Millar, L., Tessier, L., Fiori, L. M., Lau, P. C., Allan, B., and Szymanski, C. M. (2006) Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. *J. Bacteriol.* **188**, 2427–2434.
29. Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A. V., Moule, S., Pallen, M. J., Penn, C. W., Quail, M. A., Rajandream, M. A., Rutherford, K. M., van Vliet, A. H., Whitehead, S., and Barrell, B. G. (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**, 665–668.
30. Chen, M. M., Weerapana, E., Ciepichal, E., Stupak, J., Reid, C. W., Swiezewska, E., and Imperiali, B. (2007) Polyisoprenol Specificity in the *Campylobacter jejuni* N-Linked Glycosylation Pathway. *Biochemistry* **46**, 14342–14348.
31. Glover, K. J., Weerapana, E., and Imperiali, B. (2005) In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14255–14259.
32. Reid, C. W., Stupak, J., Chen, M. M., Imperiali, B., Li, J., and Szymanski, C. M. (2008) Affinity-capture tandem mass spectrometric characterization of polyprenyl-linked oligosaccharides: tool to study protein N-glycosylation pathways. *Anal. Chem.* **80**, 5468–5475.

Chapter 14

A Glycomics Approach to the Discovery of Potential Cancer Biomarkers

Hyun Joo An and Carlito B. Lebrilla

Abstract

Glycosylation is highly sensitive to the biochemical environment and plays a key role in development and disease manifestation. Moreover, glycan biosynthesis depends on several highly competitive processes; thus, variations in the concentration of specific glycosyltransferases produce different products. For this reason, monitoring changes in glycosylation may be a more specific and sensitive approach to biomarker discovery and possibly disease diagnosis. Glycans in serum are of particular interest as approximately half of all proteins are glycosylated. We have developed the methods for profiling the glycans in human serum to identify glycan biomarker. Global release methods were used including chemical and enzymatic to access O-linked and N-linked glycans, respectively. Glycans were released from the culture medium of various cancer cell lines, in control sera, and in cancer patients and isolated using solid phase extraction (SPE) with a porous graphitized carbon. The SPE fractions were analyzed by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry (MALDI FTICR MS). Glycan compositions were determined based on accurate masses and tandem mass spectrometry. Glycosylation changes between control and patient group were monitored. Several glycans were identified as potential markers for ovarian, breast, and prostate cancer. In short, direct glycan analysis of human serum without any protein identification represents a new and innovative approach to disease marker discovery.

Key words: Glycan, glycosylation, MALDI, mass spectrometry, biomarker, serum.

1. Introduction

Changes in glycosylation have been implicated in many diseases including cancer (1–3). Glycosylation, whether O-linked or N-linked, is highly sensitive to the biochemical environment and is one of the most common post-translational modification of proteins particularly on the cell surface and in the extracellular

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_14,

matrix (4, 5). The biosynthesis of glycan relies on a number of highly competitive processes involving glycosyltransferases, suggesting that the expression of the glycans products is highly variable. Thus, monitoring changes in glycosylation may prove to be an innovative technique for biomarker discovery and eventually disease diagnosis by offering more specificity and sensitivity.

Lectin-based glycan detection methods (6–10) such as lectin affinity, immuno-affinity electrophoresis, and lectin blotting methods have been used for studying the roles of glycans in diseases. However, these approaches select for specific glycans or classes of glycans such as those that are sialylated or fucosylated. More recently, global glycome approaches have been explored as a source for disease markers. These approaches have been hindered by the complexities of glycan structures and the limitations in analytical methods for elucidating structures. With the advent of mass spectrometry, which provides a rapid quantitative tool for component analysis, the limitations are diminishing swiftly.

Global glycan profiling of human serum with matrix-assisted laser desorption/ionization (MALDI) (11–17) and electrospray ionization (ESI) (18–20) mass spectrometry has recently been explored and has already led to several potentially promising markers for several diseases. These approaches rely on the notion that many diseases cause aberrant glycosylation. Monitoring and diagnosing a specific disease will rely primarily on the changes in glycosylation with little regard for the specific protein to which the glycans may be bound (21). The additional advantage of this approach is that patient sera can be examined without any protein purification, while glycans from glycosylated proteins originating from the tumor or from circulating cancer cells are both simultaneously observed.

Initial glycan profiling efforts focused on the O-linked glycans, which are glycans attached to the serine (Ser) or threonine (Thr) residue of the polypeptide backbone. In this approach glycans were globally released from glycoproteins by reductive β-elimination (11, 12) and analyzed by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance (MALDI FTICR) mass spectrometry. Mass profiles were used to examine changes between diseased and control patients. For ovarian cancer, 15 possible O-linked glycan markers were found in patient sera that were not found in normal sera (11, 22, 23). Several breast cancer tumor cell lines and patient sera were also analyzed for possible breast cancer "glycan signatures" (12). In these studies, many glycan masses were found to be derived from N-linked glycans, produced by the peeling reaction during the sodium hydroxide/sodium borohydrate treatment. β-Elimination therefore released both N-linked and O-linked glycans simultaneously. It was also determined that because of the relative abundances of the N-linked glycans, they contaminated

the O-linked glycan analysis. Nonetheless, there were distinct glycans that could be readily identified in the clinical setting that corresponded to the onset of the disease (22).

To complement the O-linked glycan release, the N-linked glycans were examined by selective release in ovarian, breast, and prostate cancer patient sera (14). N-Glycans have several advantages over O-linked glycans that make them an easier group to explore for disease markers (24). N-Linked glycans are significantly more abundant in human sera than the O-linked glycans. Their syntheses follow a set of well-prescribed rules, which make them perhaps an easier group to explore for disease marker. N-Linked glycans are released by enzymes under physiological conditions.

Gycans released by the enzyme PNGase F were profiled by MALDI FTICR MS providing high-mass accuracy and high resolution. The exact glycan compositions (hexoses, *N*-acetylhexosamines, sialic acids, and fucoses) were easily deduced based on accurate masses achieved by FTICR MS and by high-performance time-of-flight (TOF). Tandem mass spectrometry in the form of infrared multiphoton dissociation (IRMPD) and collision-induced dissociation (CID) were used to confirm composition and to obtain structural information of the potential biomarkers.

This chapter will illustrate the glycan profiling methods using global release of O-linked and N-linked glycans in human serum for the discovery of cancer biomarkers. These methods utilize chemical and enzymatic processes to release glycans from cell line supernatant and cancer patient sera.

2. Materials

2.1. Ovarian Cancer Cell Growth and Supernatant Acquisition

1. Caov-3, OVCAR-3, ES-2, and SK-OV3 ovarian cancer tumor cell lines were obtained from the American Type Culture Collection (ATCC).
2. Caov-3 and OVCAR-3 cancer cells were cultured in RPMI1640 cell medium supplemented with 10% fetal bovine serum, 100 units/mL penicillin/streptomycin, and 1% glutamine. ES-2 and SK-OV-3 cells were grown in McCoy's medium.
3. Conditioned medium (CM, supernatant) was removed from the cells during log (non-confluent) or death (confluent) cell growth, and frozen at –70°C. The CM was thawed, sterile filtered (0.2 μ filter), and concentrated using Vivacell 70 or Vivaspin 20 concentrators (VivaScience, Edgewood, NY).

2.2. Human Serum Samples

1. All human serum samples were acquired from the University of California-Davis Medical Center Clinical Laboratories using an Internal Review Board (IRB) approved protocol.
2. For ovarian cancer, 48 ovarian cancer patients and 24 healthy controls were examined. The ovarian cancer patients were at various stages of the disease ranging from post-operative, under treatment with chemotherapy, at recurrence, and under surveillance. CA125 testing was conducted by the UC Davis Medical Center Clinical Laboratory using the CA125 assay kit (AXSYM test for CA125, Abbott, Abbott Park, IL).
3. For prostate cancer, 20 serum samples from individuals with prostate cancer ($n = 10$, under active surveillance for prostate cancer, PSA = 5.4-27.0 ng/mL) and individuals with their prostate removed ($n = 10$, post-radical retropubic prostatectomy (post-RRP), PSA<0.1 ng/mL).PSA testing was conducted by the UC Davis Medical Center Clinical Laboratory using the ADVIA Centaur PSA assay test on a Bayer ADVIA Centaur (Bayer Diagnostics, Tarrytown, NY).

2.3. Release and Purification of Oligosaccharides

1. Release solution for O-linked glycan: 1.0 M $NaBH_4$ in 0.1 M NaOH, prepared freshly.
2. Digestion buffer for releasing N-linked glycans: 200 mM ammonium bicarbonate (NH_4HCO_3) and 10 mM dithiothreitol (DTT).
3. Dialysis (Slide-A-Lyzer 10 kDa MWCO Dialysis Cassettes, Thermo Fisher Scientific Inc., Rockford, IL).
4. Neutralizing solution for β-reductive elimination: 1.0 M HCl.
5. Solid phase extraction (SPE) graphitized carbon cartridges (150 mg, 4 mL) from Alltech Associates, Inc. (Deerfield, IL).
6. Activation solution: nanopure water and 80% acetonitrile (AcN) in 0.1% trifluoroacetic acid (TFA) (v/v).
7. Washing solution: nanopure water.
8. Elution solutions: 10 and 20% AcN in H_2O and 40% AcN with 0.05% TFA in H_2O.

2.4. MALDI FTICR Analyses of Oligosaccharides

1. Mass spectra were recorded on an external source HiResMALDI (IonSpec Corporation, Irvine, CA) equipped with a 7.0 Tesla magnet. The HiResMALDI is equipped with a pulsed Nd:YAG laser (355 nm).
2. Matrix for mass spectrometry analysis: 2,5-dihydroxybenzoic acid (DHB) (5 mg/100 μL in 50% AcN in H_2O).
3. Positive mode dopant: 0.1 M NaCl in 50:50 H_2O/AcN.

3. Methods

The procedure, which included the release and isolation of the glycans, is outlined in **Fig. 14.1**. These streamlined methods were developed to minimize sample handling and increase speed. The glycans were released, pre-concentrated, and fractionated into solvents of differing polarity. The glycan mixtures eluted from three SPE fractions were mass profiled using MALDI FTICR MS. In this way, extensive chromatography is avoided while possible suppression of glycan signals by other glycan components is minimized. Because high-performance mass spectrometry is used, the glycan composition can be determined solely based on the masses (*see* **Note 1**).

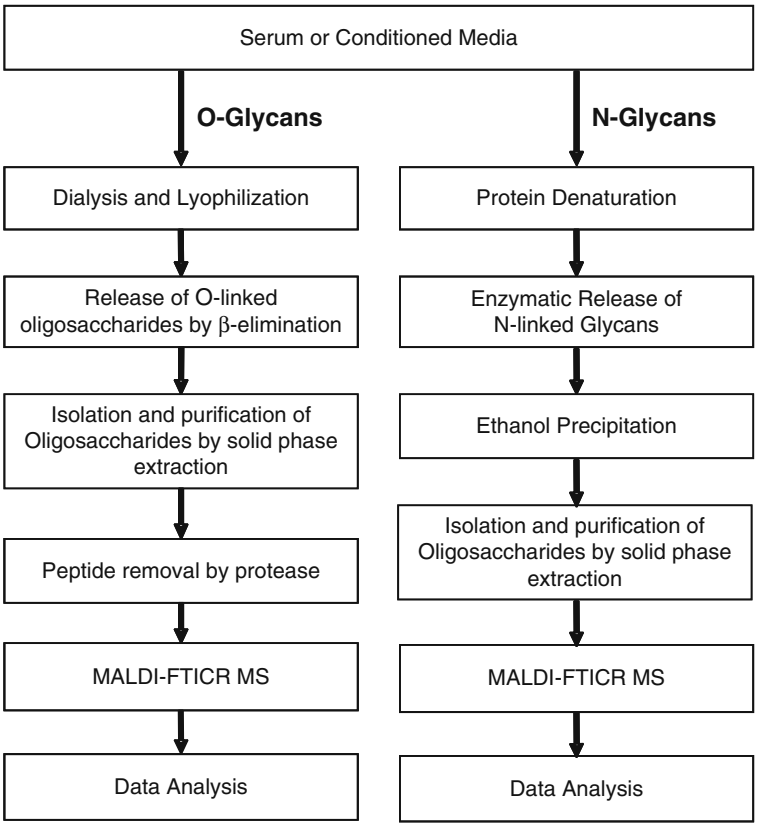


Fig. 14.1. The strategy for the release and isolation of glycans in the conditioned media of cell lines and human serum. Left pathway shows the release of O-linked glycan while right represents the N-linked glycan.

3.1. Release of O-Linked Glycans by β -Elimination

1. Serum samples (100 μL) were dialyzed against nanopure water at 4°C for 12–16 h and then lyophilized.

2. Alkaline borohydride solution (500 μL) was added to 2–3 mg of lyophilized serum samples.
3. The mixture was incubated at 42°C for 12 h in a water bath.
4. After the reaction, 1.0 M hydrochloric acid solution was slowly added to samples in an ice bath to stop the reaction and destroy the excess sodium borohydride.
5. Released glycans were desalted and fractionated prior to MS analysis.

3.2. Enzymatic Release of N-Linked Glycans

1. Human serum (100 μL) was added to 100 μL of digestion buffer.
2. The reaction mixture was heated to 100°C for 2 min to denature the protein.
3. After cooling at room temperature, 2.5 μL of Peptide *N*-glycosidase F (PNGase F) was added and the mixture (pH 7.5) was incubated at 37°C for 24 h.
4. Chilled 800 μL ethanol was added and the mixture was frozen for 1 h and then centrifuged. In the process proteins were collectively in a pellet and the glycans were in the supernatant.
5. The supernatant was completely dried down to remove the ethanol and purified by GCC-SPE.

3.3. Glycan Purification Using a Graphitized Carbon Cartridge-Solid Phase Extraction (GCC-SPE)

1. Prior to use, GCC was washed with deionized water and 80% AcN in 0.05% aqueous TFA (v/v).
2. Glycan solutions were applied to the GCC cartridge and subsequently washed with several cartridge volumes of deionized water at a flow rate of 1 mL/min for desalting.
3. Glycans were eluted with 10% AcN in H_2O (v/v), 20% AcN in H_2O (v/v), and 40% acetonitrile in 0.05% aqueous TFA (v/v).
4. Each fraction was collected and dried in a Centrivap apparatus. Fractions were reconstituted in nanopure water prior to MS analysis.

3.4. Mass Spectrometry Analysis

1. A saturated solution of NaCl in 50% AcN in H_2O was used as a cation dopant to increase signal sensitivity.
2. The glycan solution (1 μL) was applied to the MALDI probe followed by matrix solution (1 μL).
3. The sample was dried under vacuum prior to mass spectrometric analysis.

3.5. Data Analysis

3.5.1. Glycan Analysis of Ovarian Cancer Cell Lines

In this study, the conditioned medium of four ovarian cancer cell lines (Caov-3, OVCAR-3, ES-2, and SK-OV-3) in both log and death phases were extracted and examined. The glycans were released by β-elimination, i.e., in the presence of $NaBH_4$ and NaOH. This is the standard procedure for releasing O-linked glycans. Representative MALDI FTMS spectra of glycans found in the conditioned medium of ovarian cell lines in the positive mode are shown in **Fig. 14.2**. The four ovarian cancer cell lines had many glycans in common, but some (open circle) were unique to each cell line. Labeled peaks (solid circle) in the spectrum are prominent glycans present in all cell lines. Samples eluted with 40% AcN seem to provide the most abundant and largest glycans, using a positive mode on MS. They were identified to be neutral glycans based on accurate mass and tandem mass spectrometry.

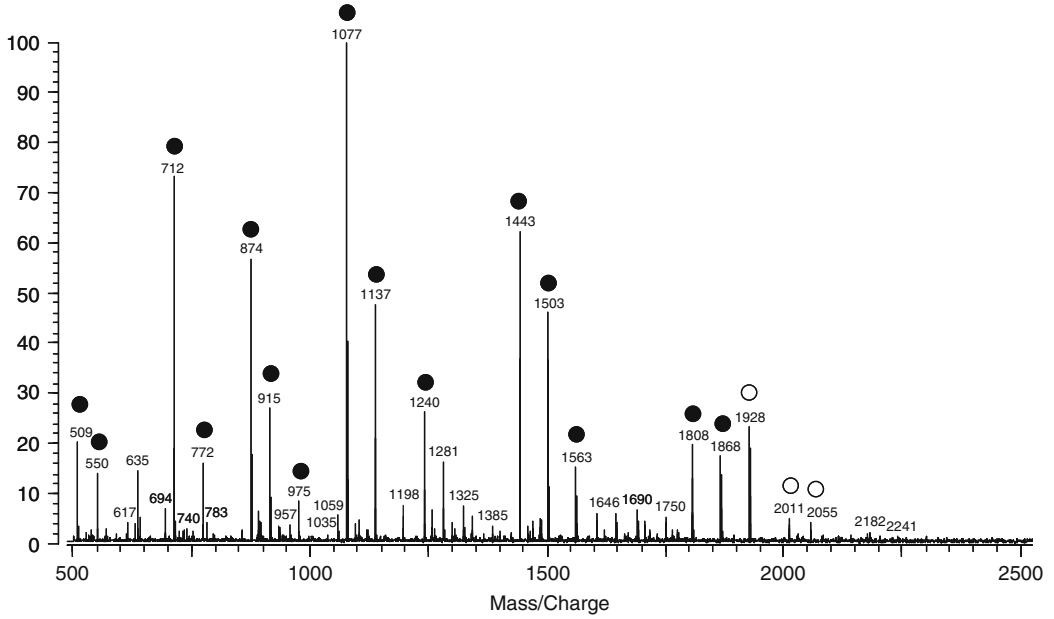


Fig. 14.2. Representative MALDI-FTMS spectra (positive mode) of glycans found in the conditioned media of ovarian cell lines released using alkaline sodium borohydrate. The majority of the peaks come from fragments of N-glycans due to peeling reaction. *Solid circle* represents common glycans in all cell lines while *open circle* represents cell-line specific glycans.

As controls, the complete medium containing 10% fetal bovine serum (FBS) as well as the defined medium without FBS was analyzed in a similar manner as the conditioned medium from the cell lines. From the complete medium containing 10% FBS, few glycans were observed, and from pure FBS only a small number of glycans were observed (less than five distinct masses). The abundances of these ions were significantly less than the cultured cell by as much as two orders of magnitude. The major ion peaks corresponded primarily to peptides whose masses were not observed in the cultured cell samples.

3.5.2. Glycan Profiling of Ovarian Cancer Patient Sera

In this small pilot study, human serum samples of five ovarian cancer patients and five healthy controls were examined to determine whether the same glycans observed in the conditioned medium of the ovarian cancer cell lines can be detected in ovarian cancer patient serum. Samples were treated with the same procedure as the cell lines. A typical mass spectrum of cancer patients and control group is shown in **Fig. 14.3**. The spectra from the patient sera contained many of the glycans found in the same fraction of the samples obtained from the ovarian cancer cell lines. In the normal sera, these signals are either totally absent or were present as only one or two peaks with very minor abundances. Tandem MS spectra were performed on the individual peaks and they were found to fragment and yield successive glycan losses.

To determine the feasibility of this approach on human serum, the glycan analysis was expanded to a larger set of serum samples consisting of 48 ovarian cancer patients and 24 healthy individuals. The ovarian cancer patient samples were chosen primarily based on their CA125 levels (23 samples for low CA125: <35 U/mL, 25 samples for high CA125: >150 U/mL). The ovarian cancer patients were high heterogeneous at various phases in their disease course. Forty-four of 48 patients had same glycans (at least 16 glycans) found in ovarian cancer cell lines and small set of serum samples while 23 of 24 control subjects had no detectable glycans.

For better comparison and quantification, the mass signal intensities of a selected group of glycans (labeled peaks in **Fig. 14.3a**) were summed for both sets of samples. In **Fig. 14.4 t**he samples on the left, with little or no intensities, are the normal samples while those on the right are samples from cancer patients. It is readily apparent that patient groups have high levels of "signature" glycans, while control groups have low or absent levels of the same glycans.

Receiver operating characteristic (ROC) curves were performed on a larger set of cancer patients (both with high and low CA125) using the glycan markers and compared to the CA125 serum levels. The ROC curve represents a trade-off between sensitivity and specificity. The greater the area under the curve indicates the better the test performance of the diagnosis. The area under the ROC curve (25) for the glycan markers is 0.95 (95% CI: 0.896, 1.0) indicating a "very good" to "excellent" test, while those for CA125 was 0.67 (95% CI: 0.57, 0.79) indicating a "fair" test for ovarian cancer (**Fig. 14.5**).

3.5.3. Glycan Profiling of N-Linked Glycans from Prostate Cancer Patient Sera

In this study a small set of serum samples consisting of 10 patients and 10 controls were initially analyzed. Globally released glycans were profiled by enzyme PNGase F without the need for protein identification, separation, and purification. **Figure 14.6**

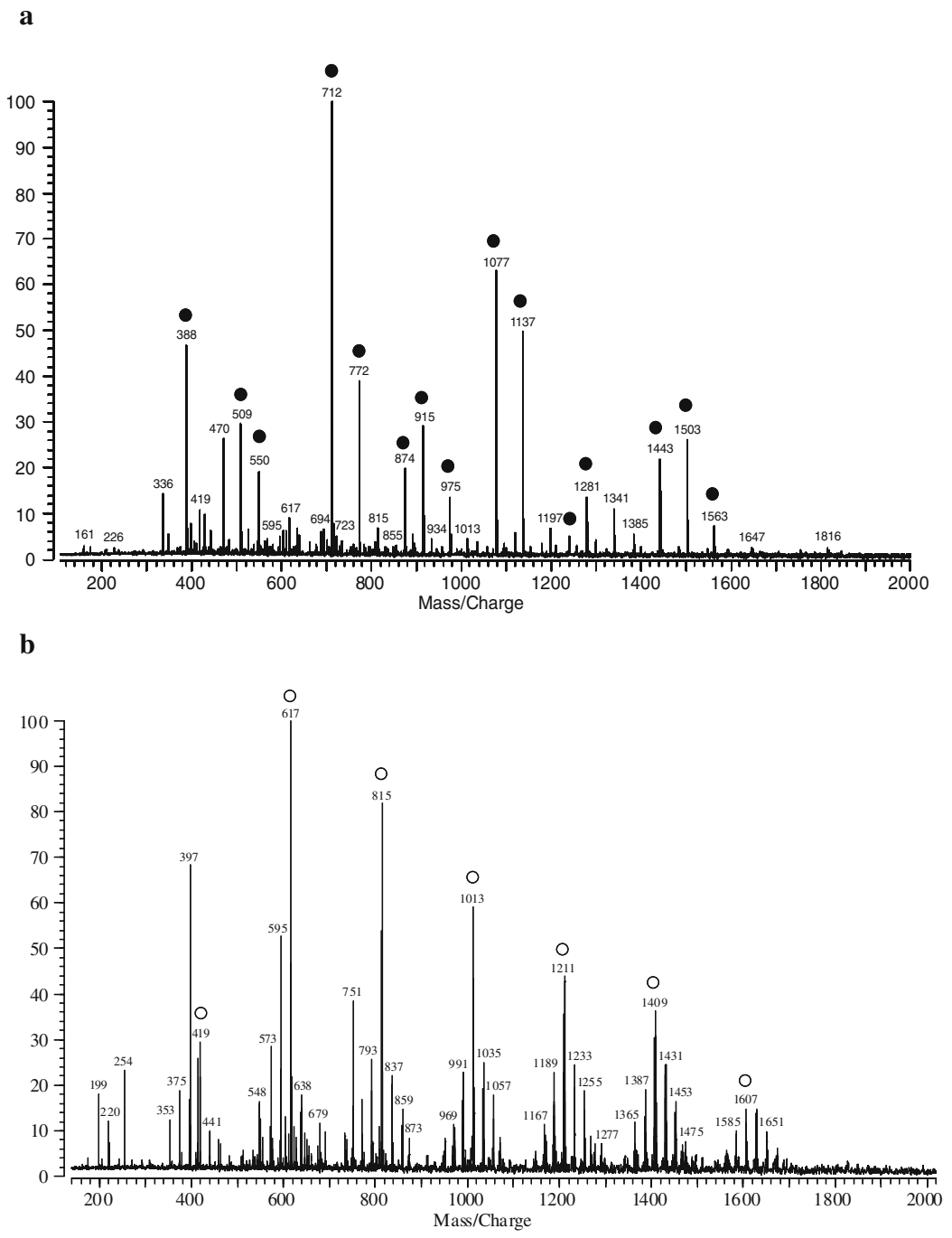


Fig. 14.3. Representative MALDI-FTMS spectra of (a) ovarian cancer patients and (b) normal individuals in the positive mode. *Solid circles* in (a) indicate glycans unique to ovarian cancer patients while *open circles* in (b) correspond to glycans found in both normal and cancer patients.

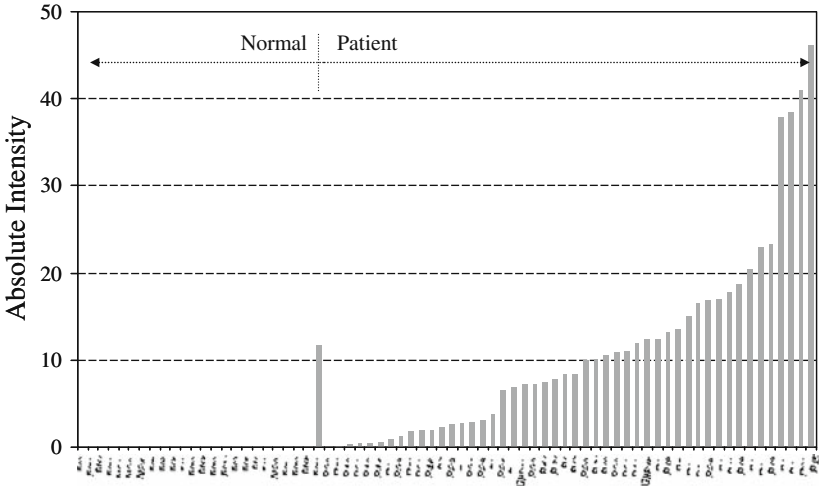


Fig. 14.4. Total intensities of glycans labeled in **Fig. 14.3a** for each individual (24 healthy controls and 48 patients with ovarian cancer). The levels of glycans were increased in ovarian cancer.

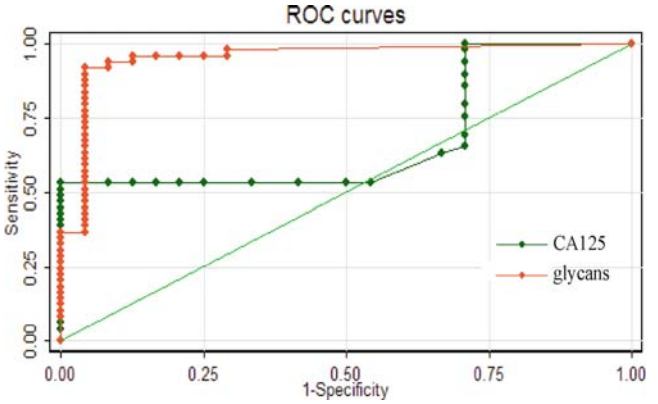


Fig. 14.5. Receiver operating characteristic (ROC) curves for all samples (48 ovarian cancer and 24 healthy control) using glycomics assay and CA125 value.

shows the representative spectra of the prostate cancer serum samples in 10, 20, and 40% fractions. N-Linked glycans are readily identified by trimannosyl core ($Man_3GlcNAc_2$) and accurate masses obtained from FTICR MS (*see* **Note 2**). The figure gives the structures of the abundant glycans only. Zooming in on the low-abundance regions resulted in exposure of more glycans. In the 10% AcN spectrum, the glycans are mostly the neutral ones in the low mass region and the scans are done in the positive mode. These are mostly the mannose and complex-type sugars. The acidic glycans in the higher mass region come out at 40% AcN and thus, the spectral scan is in the negative mode. The 20% AcN fraction contains the neutral and acidic glycans in the middle mass region and so the positive and negative scans were done on

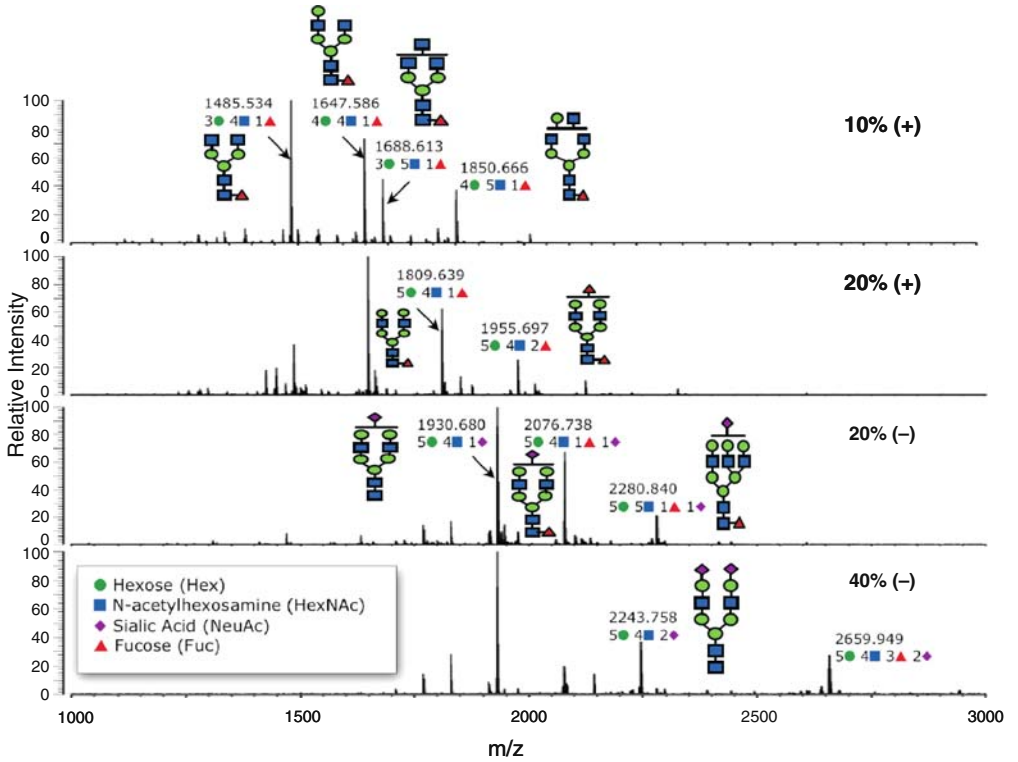


Fig. 14.6. Representative MALDI FT-ICR mass spectra of N-linked glycans in patient and control sera. 10, 20, and 40% represent glycans eluted with corresponding fraction of acetonitrile (AcN) in the positive (+) and negative (–) ion mode of the mass spectrometer. Structures are putative and are based only on accurate masses.

the said fraction. Thus, a progression from 10 to 40% AcN fractions shows not only an increase in glycan size but also in glycan polarity.

For the comparison of the peaks in patient and normal samples, statistical analysis using one-way analysis of variance (ANOVA) was performed. The absolute intensity of peaks was divided by the total ion intensity of the spectrum then multiplied by 1000. The normalized intensity of base peaks of each sample was in the range of 90–120. If the *p*-value was less than 0.05, the difference between the two groups was considered statistically significant. The calculations were carried out by the SYSTAT 11 (Systat Software Inc., San Jose, CA).

Seven potential N-linked glycan biomarkers (**Table 14.1**) were identified from the three SPE fractions. These were mostly high mannose type N-linked glycans. Four high-mannose (Man6–Man9) type, one neutral, and one acidic complex-type glycans are downregulated in the patient group while one acidic complex-type glycan is upregulated in the patient group with active disease (under active surveillance). These glycan changes are

Table 14.1
The seven possible N-linked glycan biomarkers that are statistically significant ($p < 0.05$). Glycan compositions are deduced from the accurate mass determined by MALDI-FTICR MS

m/z [M+Na]+	Composition	Type
1419.475	2HexNAc:6Hex	High mannose
1581.528	2HexNAc:7Hex	High mannose
1647.586	4HexNAc:4Hex:1Fuc	Complex/hybrid
1743.581	2HexNAc:8Hex	High mannose
1905.634	2HexNAc:9Hex	High mannose
1976.659	4HexNAc:5Hex:1NeuAc	Complex
2568.884	4HexNAc:7Hex:2Fuc:1NeuAc	Complex/hybrid

significantly different from individuals without active disease after having undergone RRP.

In many cases the overall profile of the abundant glycans especially for N-linked glycans is the same for the patient and control groups. A number of methods to determine the most reproducible and reliable glycan biomarkers in serum were developed in our laboratory. A glycan library was created to rapidly annotate composition in glycan mass profiles. Another method was used to analyze MALDI FTICR MS data. The MALDI FTICR MS data analysis process was performed in six steps: baseline correction, data transformation, peak location, peak identification, normalization, and statistical analysis (26). A shifted-log transformation was employed to stabilize the variance. Peak identification was performed on the shifted log scale using a quadratic curve fitting procedure. Although the m/z scale was used for instrument calibration, additional accuracy was obtained using a secondary calibration in which "strong" peaks – peaks that are large in all spectra – are made to correspond exactly in mass location. This procedure was done using a cubic interpolation spline (interpSpline in the splines package in the R software package) of the strong peaks. At this point, peaks in different spectra whose masses differed by at most 0.01 Da were considered as identical on the m/z scale. The spectra were additively normalized on the log scale using the average height of the strong peaks in the spectrum.

Statistical analysis to identify the glycan profiles for the cancer data was conducted using a one-way ANOVA model on each m/z set of peaks (corresponding to a compound). Since thousands of compounds were being generated for each glycan profile, multiple testing using the method of Benjamini and Hochberg

(27) was applied to control the false discovery rate (FDR) at 0.1. The Benjamini and Hochberg method can be highly sensitive to the set of starting *p*-values. Since biomarkers should be strongly present in at least one of the groups analyzed, those compounds that had a large peak in at least "K" spectra, where "K" was a number to be determined. For each data set, the set of significant differences was calculated for all possible values of K from one to the total number of spectra. As a check on the *p*-values obtained by these methods, a permutation test was performed on the peaks that were found to be significantly different to get a distribution-free estimate of the *p*-values. The results from the analysis for the prostate cancer samples show that the glycans were found to be statistically different between patients with prostate cancer and normal healthy controls.

4. Notes

1. N-Linked glycans are often released with an enzyme whereas O-linked glycans are released chemically. It is possible to release both types simultaneously but not with the desired completeness.
2. High mass accuracy and high resolution are advantages in profiling oligosaccharides. The capability of high-resolution mass analyzer such as Fourier transform ion cyclotron resonance (FTICR) to provide accurate mass has been crucial for obtaining basic glycan structure information. For example, an oligosaccharide with quasimolecular ion ($[M+Na]^+$) at m/z 2201.819 has three possible compositions within a tolerance of ± 0.1 mass units. Only with a tolerance of 0.01 mass units is the correct composition of two Fuc, four Hex, and six HexNAc obtained.

Acknowledgments

The authors gratefully acknowledge the financial support by the National Institutes of Health (R01 GM049077).

References

1. Brockhausen, I. (1999) Pathways of O-glycan biosynthesis in cancer cells. *Biochim. Biophys. Acta* **1473**, 67–95.
2. Dube, D. H. and Bertozzi, C. R. (2005) Glycans in cancer and inflammation. Potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **4**, 477–488.
3. Fuster, M. M. and Esko, J. D. (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat. Rev. Cancer* **5**, 526–542.
4. Apweiler, R., Hermjakob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of

the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8.
5. Dennis, J. W., Granovsky, M., and Warren, C. E. (1999) Protein glycosylation in development and disease. *Bioessays* **21**, 412–421.
6. Manning, J. C., Seyrek, K., Kaltner, H., Andre, S., Sinowatz, F., and Gabius, H. J. (2004) Glycomic profiling of developmental changes in bovine testis by lectin histochemistry and further analysis of the most prominent alteration on the level of the glycoproteome by lectin blotting and lectin affinity chromatography. *Histol. Histopathol.* **19**, 1043–1060.
7. Thaysen-Andersen, M., Thogersen, I. B., Lademann, U., Offenberg, H., Giessing, A. M. B., Enghild, J. J., Nielsen, H. J., Brunner, N., and Hojrup, P. (2008) Investigating the biomarker potential of glycoproteins using comparative glycoprofiling – application to tissue inhibitor of metalloproteinases-1. *Biochim. Biophys. Acta* **1784**, 455–463.
8. Uchiyama, N., Kuno, A., Tateno, H., Kubo, Y., Mizuno, M., Noguchi, M., and Hirabayashi, J. (2008) Optimization of evanescent-field fluorescence-assisted lectin microarray for high-sensitivity detection of monovalent oligosaccharides and glycoproteins. *Proteomics* **8**, 3042–3050.
9. Kuno, A., Uchiyama, N., Koseki-Kuno, S., Ebe, Y., Takashima, S., Yamada, M., and Hirabayashi, J. (2005) Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nat. Methods* **2**, 851–856.
10. Patwa, T. H., Zhao, J., Anderson, M. A., Simeone, D. M., and Lubman, D. M. (2006) Screening of glycosylation patterns in serum using natural glycoprotein microarrays and multi-lectin fluorescence detection. *Anal. Chem.* **78**, 6411–6421.
11. An, H. J., Miyamoto, S., Lancaster, K. S., Kirmiz, C., Li, B. S., Lam, K. S., Leiserowitz, G. S., and Lebrilla, C. B. (2006) Profiling of glycans in serum for the discovery of potential biomarkers for ovarian cancer. *J. Proteome Res.* **5**, 1626–1635.
12. Kirmiz, C., Li, B. S., An, H. J., Clowers, B. H., Chew, H. K., Lam, K. S., Ferrige, A., Alecio, R., Borowsky, A. D., Sulaimon, S., Lebrilla, C. B., and Miyamoto, S. (2007) A serum glycomics approach to breast cancer biomarkers. *Mol. Cell. Proteomics* **6**, 43–55.
13. Kyselova, Z., Mechref, Y., Al Bataineh, M. M., Dobrolecki, L. E., Hickey, R. J., Vinson, J., Sweeney, C. J., and Novotny, M. V. (2007) Alterations in the serum glycome due to metastatic prostate cancer. *J. Proteome Res.* **6**, 1822–1832.
14. de Leoz, M. L. A., An, H. J., Kronewitter, S., Kim, J., Beecroft, S., Vinall, R., Miyamoto, S., de Vere White, R., Lam, K. S., and Lebrilla, C. B. (2008) Glycomic approach for potential biomarkers on prostate cancer: profiling of N-linked glycans in human sera and pRNS cell lines. *Dis. Markers: Glycobiol.* **25**, 243–258.
15. Kita, Y., Miura, Y., Furukawa, J., Nakano, M., Shinohara, Y., Ohno, M., Takimoto, A., and Nishimura, S. (2007) Quantitative glycomics of human whole serum glycoproteins based on the standardized protocol for liberating N-glycans. *Mol. Cell. Proteomics* **6**, 1437–1445.
16. Morelle, W., Flahaut, C., Michalski, J. C., Louvet, A., Mathurin, P., and Klein, A. (2006) Mass spectrometric approach for screening modifications of total serum N-glycome in human diseases: application to cirrhosis. *Glycobiology* **16**, 281–293.
17. Kam, R. K. T., Poon, T. C. W., Chan, H. L. Y., Wong, N., Hui, A. Y., and Sung, J. J. Y. (2007) High-throughput quantitative profiling of serum N-glycome by MALDI-TOF mass spectrometry and N-glycomic fingerprint of liver fibrosis. *Clin. Chem.* **53**, 1254–1263.
18. Harvey, D. J., Royle, L., Radcliffe, C. M., Rudd, P. M., and Dwek, R. A. (2008) Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Anal. Biochem.* **376**, 44–60.
19. Isailovic, D., Kurulugama, R. T., Plasencia, M. D., Stokes, S. T., Kyselova, Z., Goldman, R., Mechref, Y., Novotny, M. V., and Clemmer, D. E. (2008) Profiling of human serum glycans associated with liver cancer and cirrhosis by IMS-MS. *J. Proteome Res.* **7**, 1109–1117.
20. Zhao, J., Qiu, W. L., Simeone, D. M., and Lubman, D. M. (2007) N-linked glycosylation profiling of pancreatic cancer serum using capillary liquid phase separation coupled with mass spectrometric analysis. *J. Proteome Res.* **6**, 1126–1138.
21. Lebrilla, C. B. and An, H. J. (2009) The prospects of glycan biomarkers for the diagnosis of diseases. *Mol. Biosyst.* **5**, 17–20.
22. Leiserowitz, G. S., Lebrilla, C., Miyamoto, S., An, H. J., Duong, H., Kirmiz, C., Ll, B., Liu, H., and Lam, K. S. (2008) Glycomics analysis of serum: a potential new biomarker

for ovarian cancer? *Int. J. Gynecol. Cancer* **18**, 470–475.
23. Li, B. S., An, H. J., Kirmiz, C., Lebrilla, C. B., Lam, K. S., and Miyamoto, S. (2008) Glycoproteomic analyses of ovarian cancer cell lines and sera from ovarian cancer patients show distinct glycosylation changes in individual proteins. *J. Proteome Res.* **7**, 3776–3788.
24. Lebrilla, C. B. and An, H. J. (2008) The prospects of glycan biomarkers for the diagnosis of diseases. *Mol. Biosyst.* **5**, 17–20.
25. Greiner, M., Pfeiffer, D., and Smith, R. D. (2000) Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev. Vet. Med.* **45**, 23–41.
26. Barkauskas, D. A., An, H. J., Kronewitter, S. R., de Leoz, L. M., Chew, H. K., de Vere White, R. W., Leiserowitz, G. S., Miyamoto, S., Lebrilla, C. B., and Rocke, D. M. (2009) Detecting glycan cancer biomarkers in serum samples using MALDI FT-ICR mass spectrometry data. *Bioinformatics*, **25**, 251–257.
27. Hopkins, A. M., Miller, C. J., Connolly, A. J., Genovese, C., Nichol, R. C., and Wasserman, L. (2002) A new source detection algorithm using the false-discovery rate. *Astron. J.* **123**, 1086–1094.

Chapter 15

Extraction of Chondroitin/Dermatan Sulfate Glycosaminoglycans from Connective Tissue for Mass Spectrometric Analysis

Alicia M. Bielik and Joseph Zaia

Abstract

Chondroitin/dermatan sulfate (CS/DS) glycosaminoglycans (GAGs) are present in high levels in connective tissue where they play roles as structural molecules and in protein-binding interactions. Recent developments in the techniques for analysis of CS/DS using capillary electrophoresis (CE) have enabled progress in the understanding of changes in CS/DS structure that accompany connective tissue diseases including osteoarthritis. Key to these developments is the ability to extract CS/DS GAGs from small quantities of connective tissue. This chapter describes a method for connective tissue GAG extraction, derivatization, and workup for subsequent capillary electrophoretic and/or mass spectrometric analysis.

Key words: Chondroitin sulfate, dermatan sulfate, glycosaminoglycan, connective tissue, cartilage, glycomics, mass spectrometry.

1. Introduction

Chondroitin sulfate (CS) is a glycosaminoglycan (GAG) present in extracellular matrices (1, 2) and on cell surfaces. It is present in high levels in connective tissue, bound to large proteoglycans including aggrecan and versican and small leucine rich repeat proteoglycans including decorin and biglycan. Dermatan sulfate (DS) is a form of CS that contains a high level of iduronic acid (IdoA) residues. As shown in **Fig. 15.1A**, CS/DS chains consist of repeats of GlcA/IdoA-GalNAc with variable sulfation and uronic acid epimerization. They interact with heparan sulfate-binding proteins including fibroblastic growth factors. In addition to structural roles, biological activities of CS/DS chains include

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_15,

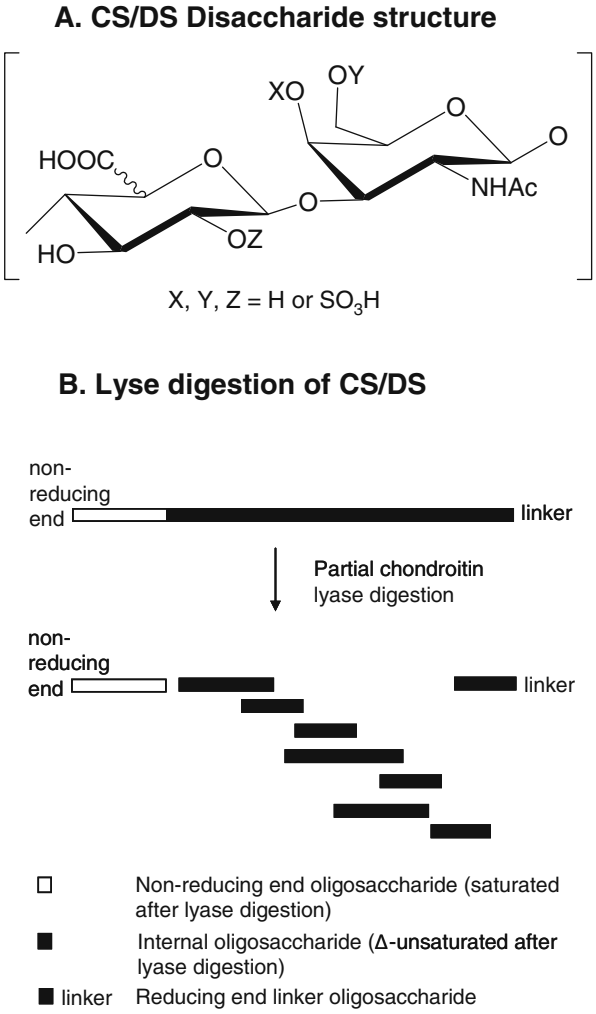


Fig. 15.1. (A) Repeating disaccharide structure of chondroitin/dermatan sulfate. (B) Polysaccharide lyase digestion of chondroitin/dermatan sulfate.

regulation of cell growth during central nervous system development (3). The structures of CS/DS chains in cartilage vary during development, specifically in the ratios of 4- to 6-sulfated GalNAc residues. In juveniles, both 4- and 6-sulfated GalNAc repeats are abundant, while in adults, 6-sulfated GalNAc repeats predominate (4). The expression of specific CS/DS glycoforms has been correlated with disease states including osteoarthritis (5), in which the sulfation pattern resembles that of a late juvenile individual.

CS/DS chains are synthesized in the Golgi apparatus as repeating units of [4GlcAβ1-3GalNAcβ1-] polymerized on a protein bound tetrasaccharide linker GlcAβ1-3Galβ1-3Galβ1-4Xylβ1-O-Ser (6, 7). Following elongation, the chains are acted upon by glucuronosyl epimerases and sulfo-transferases to

create mature chains with cell-type specific domain structures. CS sequences are commonly sulfated on the GalNAc 4-O-position or 6-O-position and less commonly at the GlcA-2-O-position. DS sequences are usually sulfated at the GalNAc 4-O-position and less commonly at the IdoA 2-O-position (8). While a given proteoglycan core protein may have the same structure in several different tissues, the structures of attached glycosaminoglycans vary. This was shown using a mass spectrometric method for the CS/DS proteoglycan decorin (9, 10). It is believed that alteration of glycosaminoglycan structures is a mechanism whereby cells respond dynamically to varying growth conditions (11).

Liquid chromatography–mass spectrometry (LC–MS) is emerging as a particularly effective technique for analysis of GAGs (12). Chromatography modes for GAGs that are compatible with MS detection include reversed phase (13), reversed phase ion pairing (14, 15), graphitized carbon (16, 17), size exclusion (10, 18, 19), and hydrophilic interaction (20–22). As shown in **Fig. 15.1B**, the GAGs are typically depolymerized using polysaccharide lyase enzymes to produce a distribution of oligosaccharide sizes. Mass measurement serves to define the composition of the CS/DS oligosaccharides with respect to number of HexA, HexN, sulfate, and acetate groups. Oligosaccharides cleaved by lyases have a 4,5-unsaturated uronic acid residue at the non-reducing end, termed non-reducing end. Those deriving from the non-reducing end of the parent chain terminate in saturated HexA. Those oligosaccharides containing the tetrasaccharide linker are differentiated based on their mass. Tandem MS of CS/DS oligosaccharides serves to determine the glycoform mixture percentage (9, 10).

This chapter describes a capillary electrophoresis (CE) (23) and MS (10, 18, 21) compatible method for extraction and derivatization of CS/DS from connective tissue. The method has been optimized to separate GAGs from other anionic molecules that interfere with MS detection. Included are methods for chondroitin lyase digestion, reductive amination, and workup of the samples for CE and/or MS analysis.

2. Materials

2.1. Extraction of CS/DS from Connective Tissue

1. Eppendorf ™ uncoated polypropylene microcentrifuge tubes.
2. Phosphate buffered saline: Dilute 10X phosphate buffered saline (PBS, calcium and magnesium free, Sigma-Aldrich) to 1X using water.

3. Papain solution: Dissolve papain (Sigma-Aldrich) in 1X PBS at 1 mg/mL.
4. Papain digestion buffer: 1X PBS, 10 mM EDTA, and 5 mM cysteine.

2.2. Ethanol Precipitation

1. Absolute ethanol (Sigma-Aldrich) stored at –20°C.

2.3. Alkaline Borohydride Release of GAGs

1. Sodium borohydride solution: 1 M in water, freshly made.
2. Sodium hydroxide solution: 0.5 M.
3. Acetic acid, glacial, HPLC grade.

2.4. Reversed Phase Solid Phase Extraction

1. PepClean C18 spin columns (Thermo Scientific Pierce Protein Research Products).
2. Spin column solution: 0.1% acetic acid.

2.5. Strong Anion Exchange (SAX) Solid Phase Extraction

1. SAX microcolumn, strong anion exchange Microspin columns (Harvard Apparatus).
2. Strong buffer: 300 mM Na_2HPO_4.
3. Weak buffer: 50 mM phosphate pH 3.5.
4. Sodium chloride solution: 1 M.

2.6. Chondroitinase Digestion

1. Chondroitinase digestion buffer: 50 mM Tris HCl, pH 7.0, 5 mM ammonium acetate.
2. Chondroitinase ABC: (Northstar Bioproducts) 4 mU/μL in water.
3. Chondroitinase ACI: (Northstar Bioproducts) 2 mU/μL in water.
4. Chondroitinase B: (Northstar Bioproducts) 0.5 mU/μL in water.

2.7. Reductive Amination

1. Anhydrous dimethylsulfoxide (DMSO).
2. 2-aminoacridone (AMAC) aliquots. Pipet a 4-μM solution of AMAC (Sigma-Aldrich) in 99% methanol/0.1% ammonium hydroxide into aliquots of 5×10^{-7} mol per microcentrifuge tube. Dry in vacuo and store at –20°C in the dark.
3. AMAC reductive amination solution: 3:17 (v/v) acetic acid:DMSO.
4. Sodium cyanoborohydride solution: 1 M in water, freshly made.
5. 2-anthranilic acid (2AA, Sigma-Aldrich) in 6 mg aliquots.
6. Sodium cyanoborohydride in 6 mg aliquots.
7. 2AA reaction solution, 3:7 (v/v) acetic acid:DMSO.

2.8. Cellulose Microcolumn Solid Phase Extraction of Reductive Amination Reactions

1. Cellulose microspin columns (Harvard Apparatus).
2. 30% acetic acid solution.
3. Acetonitrile, HPLC grade.
4. 96% Acetonitrile.

2.9. High-Performance Size Exclusion Chromatography (SEC) of CS/DS Oligosaccharides

1. SEC column, Superdex Peptide PC 3.2/30 (GE Healthcare Life Sciences).
2. SEC mobile phase: 50 mM ammonium formate, 10% acetonitrile.
3. CS 4-sulfated disaccharide standard (ΔDi4S, Northstar Bioproducts): 100 pmol/μL.

3. Methods

The methods described may be used to prepare GAGs for disaccharide analysis and/or mass spectral analysis. Proteolysis is used to digest the connective tissue, followed by ethanol precipitation and alkaline borohydride release. The released GAGs are purified using solid phase extraction and then digested with chondroitinases. Exhaustive chondroitinase digestion is used to prepare samples for disaccharide analysis using capillary electrophoresis (23). Limited chondroitinase digestion is used to prepare samples for mass spectral analysis (10, 18, 21).

3.1. Extraction of CS/DS from Connective Tissue

1. Dry connective tissue (cartilage, meniscus, muscle, tendon, synovium) in vacuo and then determine weight.
2. Add 9 μL papain digestion buffer plus 1 μL papain solution per 10 μg dried tissue. Add additional solution as necessary to completely immerse the tissue sample.
3. Incubate for 24 h at 60°C with end over end mixing.
4. Add papain solution, using the same volume as above, and incubate for an additional 24 h with end-over-end mixing.
5. Pipet supernatant into a new microcentrifuge tube and dry in vacuo.
6. Dissolve the sample in 50 μL water and vortex.

3.2. Ethanol Precipitation

1. Add 450 μL chilled ethanol and vortex. Store at –20°C for 18 h or –80°C for 4 h.
2. Centrifuge at 12,000×*g* for 10 min using a fixed angle rotor. Mark the microcentrifuge tube to indicate the position of the pellet.

3. Pipet off the ethanol supernatant taking care not to disturb the pellet. Discard the ethanol supernatant. The pellet may not be visible, so use the mark as a guide.
4. Repeat the ethanol precipitation two more times.

3.3. Alkaline Borohydride Release of GAGs

1. Combine 450 μL of sodium borohydride solution and 50 μL of sodium hydroxide solution and vortex.
2. Add 25 μL to the dried GAG sample and incubate at 45°C for 18 h.
3. Add 2.5 μL glacial acetic acid or more as necessary to achieve a pH of 5.

3.4. Reversed Phase Solid Phase Extraction

1. Place a C18 microspin column in a 1.5-mL microcentrifuge tube and wash three times with 30 μL spin column wash solution. Wash by centrifugation at 1500×*g* for 15 s. Discard wash solution and 1.5 mL microcentrifuge tube. Place C18 microspin column into a new 0.5 mL Eppendorf tube.
2. Pipet the acidified borohydride release solution onto the column and let sit at room temperature for 5 min.
3. Add 30 μL spin column wash solution and centrifuge at 1500×*g* for 15 s. Repeat two more times, combining all eluants.
4. Dry the combined eluant in vacuo.
5. Precipitate the GAGs in ethanol as in **Sect. 3.2**, one time.

3.5. Strong Anion Exchange Solid Phase Extraction

1. Place an SAX microspin column into a 1.5-mL microcentrifuge tube and apply 200 μL water to the column. Let the SAX column sit at room temperature for 10 min.
2. Centrifuge at 1500×*g* for 15 s. Repeat with two more 200 μL volumes of water. Discard eluants.
3. Wash the column with 200 μL volumes of strong buffer and centrifuge, discarding eluant. Repeat for a total of four washes.
4. Add 200 μL strong buffer, cover tubes, and let sit at room temperature for 1 h.
5. Centrifuge at 1500×*g* for 15 s, discard eluant.
6. Wash column with 100 μL of weak buffer, centrifuge at 1500×*g* for 15 s and discard eluant. Repeat for a total of three washes.
7. Dissolve the GAG pellet in 50 μL weak buffer and pipet it onto the spin column.
8. Centrifuge at 500×*g* for 15 s. Collect the eluant and pipet onto the top of the spin column. Repeat two more times.
9. Apply the eluant to the column again and let sit for 5 min at room temperature.

10. Wash the column with 100 μL weak buffer and centrifuge at 1500×*g* for 15 s, discarding eluant. Repeat for a total of three washes. Discard the 1.5-mL microcentrifuge tube and place the spin column in a new 0.5 mL Eppendorf tube.
11. Apply 75 μL sodium chloride solution and centrifuge at 1500×*g* for 15 s. Repeat one time, combining eluants.
12. Dry combined eluants in vacuo.

3.6. Ethanol Precipitation

1. Ethanol precipitate the sample one time as in **Sect. 3.2** and dry *in vacuo*.

3.7. Exhaustive Chondroitinase Digestion for Subsequent Disaccharide Analysis

1. Dissolve samples in water, remove 10% for exhaustive digestion and re-dry.
2. Add 92 μL of chondroitinase digestion buffer to the dried GAG sample.
3. Add 3 μL chondroitinase ABC solution, 3 μL chondroitinase ACI solution, and 2 μL chondroitinase B solution and vortex.
4. Incubate at 37°C for 2 h.
5. Add 3 μL chondroitinase ABC solution.
6. Incubate at 37°C for 1 h.
7. Boil the sample for 1 min.
8. Measure absorbance of the digestion solution at 232 nm relative to buffer blank.
9. Dry sample in vacuo.

3.8. Reductive Amination with 2-Aminoacridone

1. To a dried AMAC aliquot add 10 μL methanol, vortex, and briefly centrifuge to force solution to the bottom of the tube.
2. To the dried GAG sample add 10 μL water.
3. Add the AMAC solution to the GAG solution and vortex.
4. Incubate at 45°C for 10 min in the dark.
5. Dry in vacuo.
6. Add 5 μl AMAC reductive amination solution plus 5 μL sodium cyanoborohydride solution and vortex.
7. Briefly centrifuge to bring the solution to the bottom of the microcentrifuge tube.
8. Incubate at 45°C for 4 h in the dark.
9. Dilute the sample with 10 μL DMSO.

3.9. Reductive Amination Cleanup

1. Place a cellulose microspin column into a 1.5-mL microcentrifuge tube and add 200 μL water to the column and let sit at room temperature for 10 min.

2. Solvent washes: add solvent, centrifuge at 1500×*g* for 15 s. Discard eluant.
3. Wash three times with 200 μL water.
4. Wash four times with 200 μL 30% acetic acid.
5. Wash four times with 200 μL acetonitrile. Discard the 1.5 mL microcentrifuge tube and replace with a new 1.5 mL microcentrifuge tube.
6. Immediately apply GAG sample from **Sect. 3.8** in Step 9 and centrifuge at 500×*g* for 15 s.
7. Collect eluant, re-apply it to the top of the column, and centrifuge at 500×*g*.
8. Collect eluant, re-apply it to the top of the column, and let sit for 10 min at room temperature.
9. Wash three times with 200 μL acetonitrile, discard eluant.
10. Wash two times with 200 μL 96% acetonitrile, discard eluant. Discard the 1.5 mL microcentrifuge tube and place microspin column in a new 0.5 mL Eppendorf tube.
11. Add 100 μL water, centrifuge at 1500×*g* for 15 s and collect eluant. Repeat with a second volume water and combine eluants.
12. Dry the eluants in vacuo.

3.10. Analysis Disaccharide

1. Determine disaccharide composition using capillary electrophoresis as described (23).

3.11. Limited Chondroitinase Digestion

1. To the remaining sample from **Sect. 3.7** in Step 1 add 92 μL chondroitinase digestion buffer and chondroitinase enzyme solutions (3 μL chondroitinase ABC, 3 μL chondroitinase ACI, and 2 μL chondroitinase B).
2. Monitor digestion using absorbance at 232 nm. Stop the reaction at 30% of the absorbance relative to that measured in **Sect. 3.7** in Step 8 by boiling for 1 min.
3. Dry the sample in vacuo.

3.12. Reductive Amination Using 2-Anthranilic Acid

1. Add 100 μL 2AA reaction solution to 6 mg dry 2AA.
2. Pipet this solution into a tube containing 6 mg dry sodium cyanoborohydride.
3. Add 10 μL of this solution to a tube containing the partially digested GAG sample from **Sect. 3.11** in Step 3 and vortex.
4. Incubate at 65°C for 3 h.
5. Dilute with 10 μL DMSO.
6. Cleanup as described in **Sect. 3.9**.

3.13. High-Performance Size Exclusion Chromatography

1. Equilibrate a Superdex peptide PC 3.2/30 column with SEC mobile phase at 40 μL/min for 3 h.
2. Inject 10 μL CS disaccharide standard to determine elution time by monitoring at an absorbance of 232 nm.
3. Dissolve the derivatized GAG sample in 10 μL SEC mobile phase and inject onto column. Collect fractions from 20 min through the time of elution of the disaccharide standard.
4. Dry fraction in vacuo.

3.14. Mass Spectrometry

1. Analyze using LC–MS as described (10, 18, 21)

4. Notes

1. Eppendorf ™ brand uncoated polypropylene microcentrifuge tubes produce good results. Other brands of plasticware may also be acceptable. Siliconized tubes are not recommended.
2. Unless stated otherwise all water used should have a resistivity of 19.2 MΩ-cm and a total organic content of <5 parts per billion.
3. Make stock solutions of chondroitinase enzymes, aliquot into tubes, and store at –80°C. Remove aliquots and discard after use. Do not re-freeze aliquots.
4. Samples should be dried in a centrifugal evaporator, if available.
5. The Superdex size exclusion column has a pressure limit of approximately 300 psi. It is advisable to use a backpressure regulator in line with the column to prevent accidental overpressure. It is advisable to use an in-line filter after the injector to prevent damage to the column by particulates. These devices are available from Upchurch Scientific.
6. Ethanol precipitation should be repeated three times following papain digestion to completely remove artifacts. Ethanol precipitation following reversed phase and strong anion exchange solid phase extraction steps must only be done once, in order to minimize sample losses.
7. To achieve complete binding of the GAG sample to the microspin columns, it is imperative to pass the GAG sample through the columns three times using 500×*g*. If the sample is just left to sit on top of the column, complete binding will not occur and sample loss is possible.

8. To achieve complete sample recovery, it is imperative that sample eluants from all microspin columns are collected into a 0.5-mL Eppendorf tube and NOT a 1.5-mL microcentrifuge tube.
9. A UV spectrophotometer must be used to achieve an accurate partial chondroitin lyase digestion. The absorbance at 232 nm of the exhaustively digested GAG sample is the 100% value. Calculate the 30% OD value from this number. Stop all 30% partial digestions at the calculated OD by watching the sample in a UV spectrophotometer at 232 nm.
10. A 30% digestion of 10 μg of tissue will take less than 1 min. Tissues with high contents of CSB will tend to digest at a slower rate.
11. It is imperative to note that the AMAC reductive amination solution is 3:17 (AcOH:DMSO), while the anthranilic acid reductive amination solution is 3:7 (AcOH:DMSO).

Acknowledgments

The authors were supported by NIH grant P41RR10888 and NIH contract NO1HV28178.

References

1. Heinegard, D., Bjorne-Persson, A., Coster, L., Franzen, A., Gardell, S., Malmstrom, A., Paulsson, M., Sandfalk, R., and Vogel, K. (1985) The core proteins of large and small interstitial proteoglycans from various connective tissues form distinct subgroups. *Biochem. J.* **230**, 181–194.
2. Ruoslahti, E. (1988) Structure and biology of proteoglycans. *Annu. Rev. Cell Biol.* **4**, 229–255.
3 Sugahara, K., Mikami, T., Uyama, T., Mizuguchi, S., Nomura, K., and Kitagawa, H. (2003) Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. *Curr. Opin. Struct. Biol.* **13**, 612–620.
4. Roughley, P. J. and White, R. J. (1980) Age-related changes in the structure of the proteoglycan subunits from human articular cartilage. *J. Biol. Chem.* **255**, 217–224.
5. Sharif, M., Osborne, D. J., Meadows, K., Woodhouse, S. M., Colvin, E. M., Shepstone, L., and Dieppe, P. A. (1996) The relevance of chondroitin and keratan sulphate markers in normal and arthritic synovial fluid. *Br. J. Rheumatol.* **35**, 951–957.
6. Coster, L., Hernnas, J., and Malmstrom, A. (1991) Biosynthesis of dermatan sulphate proteoglycans. The effect of beta-D-xyloside addition on the polymer-modification process in fibroblast cultures. *Biochem. J.* **276** (Pt 2), 533–539.
7. Fransson, L. A., Belting, M., Jonsson, M., Mani, K., Moses, J., and Oldberg, A. (2000) Biosynthesis of decorin and glypican. *Matrix. Biol.* **19**, 367–376.
8. Cheng, F., Heinegard, D., Malmstrom, A., Schmidtchen, A., Yoshida, K., and Fransson, L. A. (1994) Patterns of uronosyl epimerization and 4-/6-O-sulphation in chondroitin/dermatan sulphate from decorin and biglycan of various bovine tissues. *Glycobiology* **4**, 685–696.
9. Miller, M. J. C., Costello, C. E., Malmström, A., and Zaia, J. (2006) A tandem mass spectrometric approach to determination of chondroitin/dermatan sulfate oligosaccharide glycoforms. *Glycobiology* **16**, 502–513.
10. Hitchcock, A. M., Costello, C. E., and Zaia, J. (2006) Glycoform quantification of chondroitin/dermatan sulfate using an

LC/MS/MS platform. *Biochemistry* **45**, 2350–2361.
11. Turnbull, J., Powell, A., and Guimond, S. (2001) Heparan sulfate: decoding a dynamic multifunctional cell regulator. *Trends Cell Biol.* **11**, 75–82.
12. Zaia, J. (2009) On-line separations combined with MS for analysis of glycosaminoglycans. *Mass Spectrom. Rev.* **28**, 254–272.
13. Mason, K. E., Meikle, P. J., Hopwood, J. J., and Fuller, M. (2006) Characterization of sulfated oligosaccharides in mucopolysaccharidosis type IIIA by electrospray ionization mass spectrometry. *Anal. Chem.* **78**, 4534–4542.
14. Kuberan, B., Lech, M., Zhang, L., Wu, Z. L., Beeler, D. L., and Rosenberg, R. (2002) Analysis of heparan sulfate oligosaccharides with ion pair-reverse phase capillary high performance liquid chromatography-microelectrospray ionization time-of-flight mass spectrometry. *J. Am. Chem. Soc.* **124**, 8707–8718.
15. Thanawiroon, C., Rice, K. G., Toida, T., and Linhardt, R. J. (2004) Liquid chromatography/mass spectrometry sequencing approach for highly sulfated heparin-derived oligosaccharides. *J. Biol. Chem.* **279**, 2608–2615.
16. Estrella, R. P., Whitelock, J. M., Packer, N. H., and Karlsson, N. G. (2007) Graphitized carbon LC-MS characterization of the chondroitin sulfate oligosaccharides of aggrecan. *Anal. Chem.* **79**, 3597–3606.
17. Karlsson, N. G., Schulz, B. L., Packer, N. H., and Whitelock, J. M. (2005) Use of graphitised carbon negative ion LC-MS to analyse enzymatically digested glycosaminoglycans. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **824**, 139–147.
18. Hitchcock, A. M., Yates, K. E., Shortkroff, S., Costello, C. E., and Zaia, J. (2006) Optimized extraction of glycosaminoglycans from normal and osteoarthritic cartilage for glycomics profiling. *Glycobiology* **17**, 25–35.
19. Zaia, J. and Costello, C. E. (2001) Compositional analysis of glycosaminoglycans by electrospray mass spectrometry. *Anal. Chem.* **73**, 233–239.
20. Naimy, H., Leymarie, N., Bowman, M., and Zaia, J. (2008) Characterization of heparin oligosaccharides binding specifically to antithrombin III using mass spectrometry. *Biochemistry* **47**, 3155–3161.
21. Hitchcock, A., Yates, K. E., Costello, C., and Zaia, J. (2008) Comparative glycomics of connective tissue glycosaminoglycans. *Proteomics* **8**, 1384–1397.
22. Staples, G., Bowman, M., Costello, C. E., Hitchcock, A., Lau, J., Leymarie, N., Miller, C., Naimy, H., Shi, X., and Zaia, J. (2009) A chip-based amide-HILIC LC–MS platform for glycosaminoglycan glycomics. *Proteomics* **9**, 686–695. 9/10/08.
23. Hitchcock, A., Bowman, M., Staples, G., and Zaia, J. (2008) Improved workup for glycosaminoglycan disaccharide analysis using capillary electrophoresis with laser-induced fluorescence detection. *Electrophoresis* **29**, 4538–4548.

Chapter 16

N-Linked Protein Glycosylation in a Bacterial System

Harald Nothaft, Xin Liu, David J. McNally, and Christine M. Szymanski

Abstract

N-Linked protein glycosylation is conserved throughout the three domains of life and influences protein function, stability, and protein complex formation. *N*-Linked glycosylation is an essential process in Eukaryotes; however, although *N*-glycosylation affects multiple cellular processes in Archaea and Bacteria, it is not needed for cell survival. Methods for the analyses of *N*-glycosylation in eukaryotes are well established, but comparable techniques for the analyses of the pathways in Bacteria and Archaea are needed. In this chapter we describe new methods for the detection and analyses of *N*-linked, and the recently discovered free oligosaccharides (fOS), from whole cell lysates of *Campylobacter jejuni* using non-specific pronase E digestion and permethylation followed by mass spectrometry. We also describe the expression and immunodetection of the model *N*-glycoprotein, AcrA, fused to a hexa-histidine tag to follow protein glycosylation in *C. jejuni*. This chapter concludes with the recent demonstration that high-resolution magic angle spinning NMR of intact bacterial cells provides a rapid, non-invasive method for analyzing fOS in *C. jejuni* in vivo. This combination of techniques provides a powerful tool for the exploration, quantification, and structural analyses of *N*-linked and free oligosaccharides in the bacterial system.

Key words: *Campylobacter jejuni*, free oligosaccharide (fOS), *N*-linked protein glycosylation, glycomics, oligosaccharyltransferase, permethylation, pronase E, mass spectrometry, glycobiology, high-resolution magic angle spinning NMR spectroscopy (HR-MAS NMR).

1. Introduction

1.1. Asparagine-Linked (N-Linked) Protein Glycosylation Is Present in All Three Domains of Life: Bacteria, Archaea, and Eukaryotes

N-Linked protein glycosylation in Eukaryotes is an essential process that influences multiple protein functions including protein sorting, targeting, localization, and stability. *N*-Linked glycosylation also plays a role in the quality control of protein synthesis and turnover with STT3 being the essential catalytic subunit of the multi-protein oligosaccharyl-transferase complex (1). In Archaea, *N*-linked protein glycosylation is not essential for cell

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_16,

survival although deletion of the pathway results in reduced ability to grow at elevated salt concentrations, defects in motility and unstable S-layers (2, 3). Transfer of oligosaccharides to proteins is mediated by AlgB, a single subunit STT3 homolog (4). *N*-Linked protein glycosylation in Bacteria was first identified in the human foodborne pathogen *Campylobacter jejuni*. The STT3 homolog, PglB (5), is the central enzyme involved in *N*-glycosylation and is also found in a few epsilon (i.e., *Wolinella succinogenes*) and delta proteobacteria (i.e., *Desulfovibrio desulfuricans* (6, 7)), indicating that this pathway is more widespread among Bacteria than originally thought.

1.2. The *Campylobacter* N-*Linked* *Glycosylation* *Pathway*

For *C. jejuni*, disruption of the *N*-glycan pathway alters many biological processes. It has been shown that mutations in the *N*-glycan biosynthetic genes lead to changes in protein antigenicity, loss of type IV protein complex assembly and DNA uptake, reduction in adherence and invasion in vitro, and loss of colonization in mice and chickens in vivo (6, 8–12). Presently, approximately 40 proteins have been experimentally proven to be *N*-glycoproteins (13, 14). A recent survey of the theoretical *C. jejuni* proteome of proteins with predicted signal peptides suggests that over 150 may carry this post-translational modification (Nothaft H., personal communication). In addition, there appears to be a high degree of variation not only in the number of glycosylation sites per protein, but also in the glycosylation state at each site. For example, while the ratio of glycosylated versus non-glycosylated PEB3 is 1:1 in vivo, the AcrA protein is predominately fully glycosylated at both sites of modification (5, 14, 15). Recently, it was shown that the *C. jejuni* PglB is also involved in the release of the lipid-linked sugars into the periplasmic space resulting in the accumulation of large amounts of free oligosaccharides (fOS) (16).

1.3. Differences Between N-Glycan Pathways

Lipid-linked oligosaccharides (LLOs) carrying preassembled oligosaccharides are the donors for the OTase-dependent formation of *N*-linked glycoproteins in all organisms. While the oligosaccharide carrier in Eukaryotes is dolichyl-pyrophosphate (Dol-PP), dolichyl-phosphate (Dol-P) is also used in Archaea (17), whereas the glycan in *C. jejuni* is linked to an undecaprenyl-pyrophosphate carrier (18, 19) (*see* Chapter 13 by Reid et al. for more details).

1.3.1. Lipid Carrier

1.3.2. Sugar Composition

The first sugar added to the lipid carrier in Eukaryotes is almost invariably *N*-acetylglucosamine (GlcNAc). The *C. jejuni* pathway uses di-*N*-acetyl bacillosamine (2,4-diacetamido-2,4,6-trideoxyglucose) as the linking saccharide at the reducing end (5, 14), whereas in the archaeal species *Haloferax volcanii* and

Methanococcus voltae the first sugar was identified as a hexose or a GlcNAc, respectively (17). Since the discovery of an *N*-glycosylation pathway in *C. jejuni* and the subsequent identification of pathway homologs in other bacteria, methods for studying bacterial *N*-glycans are needed. The well-established methods for studying *N*-linked protein glycosylation in eukaryotes are not applicable to bacteria as these techniques are predominately based on the release of the reducing-end, GlcNAc, from protein by a sugar-specific PNGase enzyme (reviewed by Merry et al. (20)).

1.3.3. Methods to Detect Protein N-Glycosylation in the Bacterial System

Currently, the detection and analyses of protein *N*-glycosylation in *C. jejuni* has relied mainly on *N*-glycan structure-specific lectins. Following purification with this lectin (i.e., agarose-conjugated), glycoproteins are then subjected to proteolytic digestion prior to mass-spectrometric analyses (21, 22). Because lectins are only useful when the oligosaccharide structure is known, this approach has limited use for studying novel glycan structures in Bacteria. To further characterize the *C. jejuni* *N*-glycan pathway and recently discovered fOSs as well as look at homologous pathways in other proteobacteria, a universally applicable methodology that will allow the identification, analysis, and quantification of novel bacterial glycans is required.

2. Materials

2.1. Bacterial Cell Cultures and Lysis

1. MH broth (Difco): dissolve powder in double distilled H_2O and autoclave in a baffled Erlenmeyer flask for 20 min. To prepare MH plates, add 1.2% (w/v) agar.
2. 30 mg/mL trimethoprim (Tmp) in DMSO (1000-fold), always prepare fresh and store at –20°C.
3. 20 mg/mL chloramphenicol (Cm) in 70% ethanol (1000-fold), store at –20°C.
4. Cell washing buffer: 10 mM Trizma Base in double distilled H_2O, adjust the pH to 7.2 with 1 M HCl, autoclave, and store at 4°C.
5. Trigas incubator that can be adjusted to simulate microaerophilic growth conditions (5% O_2, 10% CO_2, 85% N_2) (*see* **Note 1**).
6. Sonicator (we use a Branson sonicator equipped with a microtip).
7. Spectrophotometer (we use a Nanodrop ND-1000 spectrophotometer).

2.2. Cloning of His_6-Tagged AcrA Protein

1. Obtain plasmid pCE111-28 (9) and plasmid pET24 (AcrA-His_6) (5).
2. For DNA-gel elution, plasmid preparation, DNA-restriction analyses, and DNA-ligation use buffers and conditions as recommended by the manufacturer of enzymes and DNA-purification kits.
3. Transform *E. coli* DH5 alpha, use pre-made chemical competent cells, and perform the reactions according to the manufacturer's instructions.

2.3. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

1. Separating buffer (4-fold): 1.5 M Trizma base in double distilled H_2O (adjust pH to 8.7 with 1 M HCl), 0.4% SDS. Store at room temperature.
2. Stacking buffer (4-fold): 0.5 M Trizma base in double distilled H_2O (adjust pH to 6.8 with 1 M HCl), 0.4% SDS. Store at room temperature.
3. 30% acrylamide/bis solution (29:1) (Bioshop, Canada, ON), store at 4°C.
4. *N*,*N*,*N*′,*N*′-Tetramethyl-ethylenediamine (TEMED); ammonium persulfate (APS): prepare 10% stock solution in deionized H_2O and store as single use (200 μL) aliquots at –20°C.
5. Isopropanol (2-propanol), store at room temperature.
6. SDS-PAGE running buffer (10-fold): 0.25 M Trizma base, 1.92 M glycine, 1% (w/v) SDS, in double distilled H_2O, store at room temperature.
7. SDS-PAGE sample buffer (6-fold): 0.05 M Trizma base in deionized H_2O (adjust pH to 6.8 with 1 M HCl), 25% (v/v) glycerol, 10% (w/v) DTT, 0.05% (w/v) bromophenol blue, and 0.4% (w/v) SDS.
8. Pre-stained molecular weight markers.

2.4. Western Blotting for AcrA-His_6 Detection

1. Twentyfold phosphate buffered saline (PBS): 1.3 M NaCl, 0.07 M Na_2HPO_4, 0.03 M NaH_2PO_4 in deionized H_2O, adjust pH to 7.2 with 10 M NaOH (store at room temperature).
2. PBS-T: 1-fold PBS (dilute 20-fold in double distilled water), 0.05% Tween 20 (add Tween 20 slowly while stirring the solution).
3. Transfer buffer (1-fold): 20% (v/v) 10-fold SDS-PAGE running buffer, 40% (v/v) methanol, in deionized H_2O, store at 4°C.
4. Blocking buffer: 2% (w/v) Skim Milk (Difco) in PBS–T (always prepare fresh).

5. Antibody solution: 30% (v/v) blocking buffer; 70% (v/v) PBS–T.
6. Monoclonal anti-His antibodies (mAb-His), alkaline phosphatase conjugated anti-mouse antisera (Santa Cruz Biotechnology).
7. Developing solution: 100 mM Trizma Base, 100 mM NaCl, 50 mM $MgCl_2$, in double distilled H_2O, adjust the pH with 1 M NaOH to 9.5.
8. NBT–BCIP stock solution (Roche).
9. Polyvinylidene fluoride (PVDF) membrane (Millipore); blotting paper (Whatman, 3MM).
10. Instruments to perform SDS-PAGE and Western Blotting (in our laboratory we use the BIORAD tank system).

2.5. Pronase E Digest

1. Incubating buffer (10-fold): 1 M Trizma Base (adjust to pH 7.5 with 1 M HCl), dilute 10-fold in double distilled H_2O prior to use.
2. Pronase E solution (Sigma-Aldrich), 10 mg/mL, prepare in incubation buffer, store at –20°C.
3. Non-porous graphitic carbon (PGC) cartridge (Alltech Associates, Inc., Deerfield, IL).
4. Acetonitrile (ACN) and trifluoroacetic acid (TFA).

2.6. Analyses of Permethylated Glycans Using Capillary Electrophoresis and Mass Spectrometry (CE–MS)

1. Sheath solution: isopropanol/methanol (2:1).
2. Running solution: 75% methanol, 25% distilled H_2O.
3. Bare fused-silica capillaries (o.d. 360 μm; i.d. 75 μm, Polymicro Technologies, Phoenix AZ).
4. CE–MS system (in our laboratory, a Prince CE system (Prince Technologies, The Netherlands) is coupled to a 4000 Q-Trap mass spectrometer (Applied Biosystems/MDS Sciex, Canada)).

2.7. HR-MAS-NMR

1. Bacterial cell "killing" buffer: 10 mM PBS buffer, 10% (w/v) sodium azide (NaN_3) in 98% D_2O (Cambridge Isotopes Laboratories Inc., Andover, USA).
2. Cell washing buffer: 10 mM PBS (without NaN_3) in 98% D_2O.
3. Internal standard solution (δ_H 0 ppm): 1 g of TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt) in 100 mL of 98% D_2O.
4. NMR spectrometer (we used a Varian Inova 500 MHz spectrometer equipped with a Varian 4 mm indirect detection gradient nano-NMR probe with a broadband decoupling coil), VNMRJ software (Varian Inc., Palo Alto, USA).

5. Gel loading tips.
6. Nano NMR tube (Varian Inc., Palo Alto, USA).

3. Methods

3.1. Analyses and Quantification of Bacterial N-Linked Glycosylation and Free Oligosaccharides (fOS) Using Nonspecific Proteolytic Digestion and Permethylation

Currently, permethylation analysis of glycoproteins is restricted to examining glycans that can be released from glycoproteins using PNGase glycosidases. However, due to the variability of the reducing end sugars of bacterial glycoproteins, this method will not release *N*-linked glycans.

The following section describes a novel, universally applicable method for characterizing and quantifying novel glycans produced within the bacterial system. As an example, we will discuss how this method can be used to study the *N*-glycan produced by *C. jejuni*. Because this method is based on the nonspecific proteolytic digestion and permethylation (*see* **Note 8**) of bacterial glycans, it is rapid, reproducible, and cost-effective (*see* **Note 16**).

3.1.1. Preparation of Samples

Campylobacter whole cell extracts:

1. Thaw (or restreak) *Campylobacter* cells from a frozen glycerol stock (–80°C) (using a sterile loop) onto MH agar and incubate under microaerobic conditions for 36 h at 37°C (*see* **Note 1**).
2. Restreak cells onto fresh MH agar plates (full plate) and grow under the same conditions for 18 h.
3. Harvest cells from one plate with 2 mL sterile MH.
4. Inoculate 100 mL MH broth supplemented with Trimethoprim for 18 h at 37°C under the same conditions with shaking (100 rpm) (*see* **Note 2**).
5. Cool the bacterial culture on ice (approximately 10 min) and harvest cells by centrifugation (10 min, 3500×*g*, at 4°C).
6. Wash cells twice in 1 culture volume cell washing buffer and centrifuge as described above (*see* **Note 3**).
7. Resuspend the cell pellet in 1/100 of culture volume using the same buffer and transfer to a sterile 50 mL Falcon tube.
8. Prepare whole cell lysates by sonication with an amplitude setting of 15 for 3 × 1 min, 4 s pulse, 1 s off. Cool down cells between the cycles for 5 min in an ice-water bath.
9. Centrifuge lysed cells for 45 min, 10,500×*g*, 4°C (we divided lysed cells into 1.5 mL sterile Eppendorf tubes, enabling the use of a refrigerated tabletop centrifuge).

10. Transfer the supernatants (discard the pellet) to a fresh sterile 15 mL Falcon tube and determine the protein concentrations spectrophotometrically. Use cell lysis buffer as the blank.
11. Adjust protein concentrations to 10 mg/mL using cell washing buffer. At this point, cell lysates can be processed directly or can be stored at 4°C (short term for up to 24 h) or at –20°C (long term).

3.1.2. Pronase E Digestion

1. Add 1 mg pronase E (i.e., 100 μL of stock solution) to 100 μL (containing 1 mg) *Campylobacter* whole cell extracts, mix by inverting the tube.
2. Incubate the reaction at 37°C for 48–72 h.
3. Heat the reaction to 100°C for 5 min to deactivate the pronase E.

3.1.3. Preparation of Free and N-Linked Glycans

1. Wash the PGC cartridge with 3.0 mL of 80% (v/v) ACN, containing 0.1% TFA, followed by 3.0 mL water.
2. Load the pronase E digest or untreated protein extract on the PGC cartridge.
3. Wash the cartridge with water (3.0 mL) three times to remove salts.
4. Elute sample with 3.0 mL of 50% ACN in 0.1% TFA.
5. Lyophilize the fraction collected from the 50% ACN elution step.

3.1.4. Permethylation (See Chapter 18 by Liu et al. for further details)

1. Dissolve the lyophilized sample in 50 μL 75% methanol.

3.1.5. CE–MS Analysis of Permethylated Glycans

2. Setup the CE–MS system: deliver the sheath solution at a flow rate of 1.0 μL/min. Separations are obtained on about 90 cm length bare fused-silica capillary using 75% methanol. Set the electrospray ionization voltage to 5 kV for positive ion mode detection.
3. Inject the sample at 500 mbar for 0.2 min.
4. Acquired the MS data with dwell times of 2.0 ms per step of 0.1 m/z unit in full mass scan mode. In the MS/MS (enhanced product ion scan or EPI) experiments, set the scan speed to 4000 Da/s, with Q_0 trapping. In MS/MS experiments, set the trap fill time as "dynamic" and the resolution of Q_1 as "unit".

3.2. Immunodetection of the Model N-Linked Glycoprotein, AcrA (Expressed as His Tagged Fusion Protein in Campylobacter jejuni)

This following section describes the expression of the model *C. jejuni N*-glycoprotein, AcrA, expressed as a hexa-histidine fusion protein in the native host (*see* **Note 4**) followed by immunodetection with commercially available anti-His mAbs thus eliminating the requirement for glycoprotein-specific antisera and/or sugar-specific lectins. We used AcrA because it is one of the best characterized *C. jejuni N*-glycoproteins so far possessing two glycosylation sequons (N123, N273; (13)). However, we suggest that other oligosaccharide acceptor peptides can be used to perform similar experiments (*see* **Note 17**).

3.2.1. Cloning and Expression of C. jejuni His-Tagged AcrA

1. Digest 1 μg of pET24b-*acrA* plasmid DNA (5) with *Bgl*II-*Sma*I.
2. Purify the 2.6 kb *acrA*-hexahis-tag fragment from a 0.75% agarose gel using the DNA-gel elution kit.
3. Digest plasmid pCE111/28 (9) with *Bam*HI-*Eco*RV and purify the DNA with the PCR purification kit.
4. Determine the DNA concentration spectrophotometrically, use the kit elution buffer as the blank.
5. Ligate the purified DNA fragments (vector:insert ratio 1:3, using 20–30 ng vector DNA) overnight at 16°C (23).
6. Transform *E. coli* DH5 alpha with $^1/_2$ of the ligation reaction (23) (store $^1/_2$ of the ligation reaction at –20°C), plate onto LB (Cm) plates, and incubate for at least 18 h at 37°C.
7. Select at least six single colonies from the plate, inoculate into 4 mL LB (Cm) broth, and grow culture for 18 h at 37°C.
8. Prepare plasmids of candidate colonies, digest prepared plasmid DNA with *Xho*I, use 100 ng undigested DNA of one of the *acrA*-His_6-positive plasmids (that shows DNA fragments of 1100 bp and 7800 bp on a 1% agarose gel) to transform *E. coli* RK212.2 chemically competent cells (23).
9. Conjugate *Campylobacter* cells following the protocol of Labigne-Rousell and co-workers (24). Incubate plates for up to 72 h under microaerophilic conditions. Single colonies should be visible after 36 h of incubation. Extended incubation times (longer than 72 h) might result in the growth of false-positive candidate colonies that do not contain the plasmid.
10. Select at least three Cm-resistant *Campylobacter* colonies (*acrA*-His_6 positive), restreak colonies at least twice on selective MH plates (Cm, Tmp) and grow for at least 18 h under microaerophilic conditions, restreak cells to two MH plates and grow for 18 h under the same conditions.

Harvest cells from each plate with 2 mL sterile MH and transfer cells into a sterile 1.5 mL Eppendorf tube.

11. Centrifuge cells for 2 min at 8900×*g* at room temperature, discard the supernatant, resuspend one cell pellet of each clone in MH supplemented with 20% (v/v) glycerol and store at –80°C, resuspend the other cell pellet in cell washing buffer and proceed as described in **Sect.** 3.1 to prepare whole cell extracts.

3.2.2. Immunodetection of AcrA-His$_6$ in Campylobacter *Whole Cell Lysates (SDS-PAGE and Western Blotting)*

1. Mix 20 μg of cell extracts with 6× SDS-loading buffer and apply to denaturing 10% SDS-polyacrylamide gel electrophoresis (SDS-PAGE).
2. Disassemble SDS-PAGE unit, remove stacking gel, and submerge separation gel for 30 min in transfer buffer on a rocking platform.
3. Cut PVDF membrane to the size of the separating gel, submerge membrane in methanol for 10 s, then submerge in transfer buffer, and place on the rocking platform for 30 min.
4. Assemble the transfer unit as follows (submerge all components in transfer buffer prior to use): one sheet of foam, two sheets of blotting paper (*see* **Note 5**), PVDF membrane, separating gel, two sheets blotting paper, and one sheet of foam.
5. Close cassette and place into the blotting tray and fill tank with transfer buffer.
6. Transfer for 30 min at 50 V at room temperature.
7. Disassemble the transfer unit and place membrane (protein side up) in a tray containing 30 mL blocking buffer. Incubate for 1 h at room temperature on a rocking platform.
8. Discard blocking solution and incubate membrane for 45 min in 10 mL of the primary antibody solution (1:2000 dilution of anti-His-mAb in antibody solution).
9. Remove primary antibody solution (*see* **Note 6**) and rinse the membrane five times for 20 s with 50 mL PBS–T.
10. Apply secondary antibody (prepare fresh for each experiment as a 1:2000 dilution of alkaline phosphatase conjugated anti-mouse antibody in antibody solution) and incubate for 45 min at room temperature on a rocking platform.
11. Discard the secondary antibody solution and rinse the membrane 10 times for 30 s with PBS–T.

12. Prepare developing solution: combine 100 μL of the NBT–BCIP stock solution with 5 mL of developing buffer in a 12-mL Falcon tube and mix well by inverting.
13. Pour the NBT–BCIP solution on the membrane and incubate for up to 10 min. Specific signals will appear as dark-red stains at the molecular weight of the AcrA-His_6 protein (*see* **Fig.** 16.1) (*see* **Note 9**).

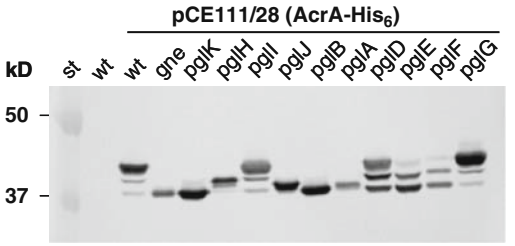


Fig. 16.1. Western Blot from a 12% SDS-PAGE gel showing the immunoreactive signals of *C. jejuni* 11168 wild-type (wt, negative control) and *C. jejuni* pCE111/28(AcrA-His_6) derivatives (as indicated) after probing whole-cell lysates with His_6-specific antiserum. The masses (in kDa) of the prestained protein standard (st) are indicated on the left. AcrA-His_6 expressed in the wild-type (wt) is predominantly modified with two *N*-linked glycans. The three signals correspond to di-glycosylated, mono-glycosylated, and unglycosylated AcrA-His_6 protein (with glyo-AcrA-His_6 proteins running at higher molecular weights). When we probe the isogenic *C. jejuni* protein glycosylation (*pgl*) mutants, an almost identical signal pattern to wt is observed with the *pglI* and *pglG* mutants. The *pglD* mutant shows lower levels of AcrA-His_6 *N*-glycosylation (as described previously (21)). AcrA-His_6-specific signals in the *pglA*, *pglH*, and *pglJ* mutants migrate at a higher molecular weight compared to unglycosylated AcrA-His_6 in the wild-type. This strongly suggests that reduced amounts of incomplete *N*-glycan are transferred. Only unglycosylated AcrA-His_6 is present in *gne*, *pglB*, and *pglK* mutants. The residual *N*-glycosylation activity in the *pglE* and *pglF* backgrounds is most probably due to partial complementation by enzymes of the flagellin glycosylation pathway as shown previously for *C. jejuni* 81-176 (34). We suggest that this method can be applied to examine and quantify the glycosylation state of any *N*-linked glycoprotein in *C. jejuni*.

14. Dry membrane on a piece of blotting paper for 3–4 h at room temperature. Take a picture of the membrane or scan the membrane at a high resolution (i.e., 600 dpi) to preserve the obtained results.

3.3. High-Resolution Magic Angle Spinning NMR Spectroscopy (HR-MAS NMR) of Campylobacter jejuni *cells*

There are several advantages to using high-resolution magic angle spinning NMR (HR-MAS NMR) to study *N*-linked protein glycosylation in *C. jejuni*. Unlike conventional NMR which requires isolation and purification of the *N*-linked glycan, HR-MAS NMR can be used to examine protein glycosylation from intact, whole bacterial cells (14, 25, 26). HR-MAS NMR provides more biologically relevant information since artifacts that may occur during the isolation step are

avoided; only a small amount of cells are needed (growth from one plate is usually enough) and cells do not require isotope labeling. Due to advances in NMR software and hardware, HR-MAS NMR is a straightforward technique that can now be routinely used in the laboratory (*see* **Notes 18–21**).

The following is a practical methodology that was developed to study *N*-linked protein glycosylation in *C. jejuni* (25, 26).

1. Harvest *C. jejuni* cells from one MH agar plate (overnight growth) as described above (**Sect. 3.1.1**) and place cells in 1 mL of cell killing buffer for 1 h at room temperature (*see* **Note 10**).
2. Pellet cells by centrifugation (8900×*g* for 2 min at room temperature). Discard the medium supernatant and wash cells by resuspending (*see* **Note 11**) the pellet in cell washing buffer, centrifuge cells under the same conditions and resuspend the resulting pellet in internal standard solution.
3. Load 40 μl of the suspension into a nano NMR tube: Use a long tipped pipette pre-cut diagonally 1 cm from the end (*see* **Note 7**).
4. Fill the nano NMR tube starting from the bottom. It is important to completely fill the nanotube with cells and to avoid introducing any air bubbles.
5. Perform HR-MAS NMR experiments. Spin cell samples at 3 kHz and record spectra in ambient temperature (23°C) or at 10°C to shift the HOD signal.
6. Perform NMR experiments with suppression of the HOD signal: use Presat to reduce the intensity of the water signal. Acquire proton NMR (^{1}H) spectra using the Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence (90-(τ-180-τ)$_n$-acquisition) to remove unwanted signals from lipids and solid-like materials on the cell surface (27). Set the total duration of the CPMG (n*2 τ) to 10 ms with τ set to (1/MAS spin rate).
7. Obtain proton HR-MAS spectra (^{1}H) by using 256–1024 scans (15 min to 1 h) (*see* **Note 12**).
8. Use selective HR-MAS 1D-NOESY and 1D-TOCSY experiments to irradiate signals of interest to establish connectivites as well as to explore intermolecular and intramolecular correlations (26). Modify 1D-TOCSY experiments to replace the DIPSI-2 pulses sequences (28, 29) with adiabatic WURST-2 pulses (26) (*see* **Notes 13–15**).
9. Use the standard VNMRJ software to process raw NMR data.

4. Notes

4.1. General Notes

1. The microaerophilic atmosphere can also be achieved through the use of sealed plastic bags after introducing a trigas mix (5% O_2, 10% CO_2, 85% N_2 gas) or by the use of an anaerobic jar in combination with the Campy*Gen*™ atmosphere generation system (Oxoid).
2. Trimethoprim is added to MH broth and MH agar to inhibit growth of contaminants; *Campylobacter* strains are naturally resistant to this antibiotic.
3. Briefly resuspend cell pallets by inverting and "tipping" the tube. Do not resuspend cells using a pipette tip as this might lead to cell lysis and release of glycans prior to cell disruption by sonication.
4. Standard protocols and techniques throughout **Sect.** 3.2.1 were performed according to *Molecular Cloning: A Laboratory Manual (Third Edition)* (23).
5. No air bubbles should be introduced between the different layers of the Western blot setup as this will result in an incomplete protein transfer.
6. The primary antibody solution can either be discarded or stored at 4°C for several days and can be re-used at least twice under the same conditions.
7. Cutting the pipette tip increases the diameter of the opening thereby preventing blockage and allowing the loading of a thick bacterial suspension.

4.2. Limitations

8. Pronase E digestion and permethylation are difficult to perform on a small scale due to the permethylation reaction (*see* Chapter 18 by Liu et al. for further details). No *N*-linked or free glycans were detected when the method was applied to a *C. jejuni pglH* or *pglJ* mutant although His-tagged AcrA protein detection by Western Blotting (as described in **Sect. 3.2.2** and **Fig. 16.1**) does show residual protein modification.
9. Although immunodetection of His-tagged proteins in the native host is more sensitive than MS or NMR, Western Blotting does not give information of the sugar composition of the oligosaccharide.
10. Placing cells in 10 mM potassium-buffered 98% D_2O containing 10% sodium azide (w/v) for 1 h at room temperature was determined to be sufficient to kill *C. jejuni* cells through analysis of different concentrations of sodium azide, but it is important to optimize killing conditions

individually for all organisms to avoid contaminating NMR equipment with harmful bacteria.

11. When preparing samples for HR-MAS NMR, it is important not to damage the bacterial cells. When cells rupture, signals originating from the internal contents such as nucleic acids, proteins, free sugars, and small peptides become visible and could potentially overlap or mask signals from the *N*-linked glycan or other glycans of interest. Because HR-MAS NMR requires cells to be spun at a high rate of speed (3 kHz), it is unavoidable that cells will get damaged during the analysis. We have found that *C. jejuni* cells can be spun for 24 h and still give high-quality spectra. After 24 h of spinning, however, proton signals from internal cellular components become visible indicating deterioration of membrane integrity.
12. For some strains, NMR proton signals from the *N*-glycan overlap with those originating from the capsule. For these strains, capsule-deficient mutants, such as the *kpsM* background, need to be constructed before HR-MAS NMR analysis of the *N*-glycan can be done.
13. *N*-Linked glycan resonances are often minor components on bacterial cells, therefore the time needed to obtain TOCSY and NOESY spectra varies from 1 to 8 h.
14. Due to the physics involved in NMR spectroscopy, glycans must have some degree of mobility in order to be visible with HR-MAS NMR. For this reason, only certain surface glycans are ideally suited for study using HR-MAS NMR. Capsular polysaccharides are often well suited to study by HR-MAS NMR because they are long chain surface polysaccharides that extend away from the cell surface. At the same time, sub-cellular glycans that are relatively non-restricted in their movement, such as the recently discovered free oligosaccharides derived from the *N*-glycan pathway, may also be visible with HR-MAS NMR. Since the mobility of cell-associated glycans vary and are difficult to predict, HR-MAS NMR spectroscopy of whole cells can require trial and error of each system in the initial stages.
15. Spectra obtained using HR-MAS NMR of cell-associated glycans are usually of lower resolution compared to those obtained for the purified glycan using conventional liquid NMR. The restricted motion of cell-associated glycans can cause considerable broadening of signals. Furthermore, with HR-MAS NMR of non-purified whole bacterial cells, it is not uncommon to detect signals from molecules within buffer, or broad signals from cell surface components such as proteins, peptides, and lipids. These signals may prove

problematic if they overlap with or mask signals of interest. At the same time, it may be worth investigating the origins of these signals since they could potentially represent novel glycans or cell components.

4.3. Further Applications

16. Pronase E digestion could also be used for bacterial protein O-glycosylation analyses and for the analyses of pathways that form lipid-linked oligosaccharide intermediates (i.e., *Pseudomonas* and *Neisseria* species). The pronase E methodology is a less harsh for labile glycans compared to beta-elimination and is an ideal method for isolating intact sugars for NMR structural analysis or for conjugation of the amino acid-linked sugars to supports such as those used for surface plasmon resonance (SPR).
17. Immunodetection of His_6-tagged proteins with multiple glycosylation sites might give an extended read-out on glycosylation efficiency and can be used in combination with available MS^n technologies for oligosaccharide structure determination.
18. When examining *N*-linked protein glycosylation in *C. jejuni* by HR-MAS-NMR, cells do not require labeling with NMR-active isotopes which can be costly and laborious. However, artificially enriching cells with NMR-active isotopes can be advantageous and can expedite the study of low natural abundance nuclei (i.e., ^{13}C, ^{15}N). In a recent example, *C. jejuni* cells that were grown on $^{15}NH_4Cl$-enriched media were shown to incorporate ^{15}N into a nitrogen-containing glycan structure on capsular polysaccharide which facilitated its structural elucidation using HR-MAS NMR (30).
19. HR-MAS NMR is a versatile technique and there is a wide variety of one-dimensional (1D) and two-dimensional (2D) experiments available for HR-MAS applications. For HR-MAS proton NMR (^1H) experiments, reasonable quality proton spectra can be obtained in a matter of minutes. This makes HR-MAS NMR particularly attractive as a high-throughput method of screening large numbers of cells. In a recent work, we used HR-MAS as a rapid, high-throughput means to examine multiple isolates of *C. jejuni* and to screen a library of capsule mutants for the presence of an unusual *O*-methyl phosphoramidate (MeO*P*N) modification (31).
20. We have used this HR-MAS NMR protocol in a number of studies to examine *N*-linked glycans and capsular polysaccharide production in *C. jejuni* from whole cells (25, 26, 30–33). We have found it useful to begin with a proton

(^{1}H) experiment. The proton experiment is arguably the most powerful experiment because it is rapid, straightforward and will provide insight into the types of molecules that are detectable on the cell surface as well as contaminants that are present in the sample. From there, other HR-MAS NMR experiments can be strategically selected as needed such as 1D TOCSY or NOESY experiments (26, 33). Two-dimensional TOCSY and NOESY experiments (33) or 2D heteronuclear experiments such as ^{1}H–^{31}P HSQC (30–33).

21. Once working conditions have been optimized, the HR-MAS NMR methodology can be automated through the use of macros or shortcuts written within the NMR software. HR-MAS NMR is a versatile technique and can be used to examine the *N*-glycan pathway in multiple *C. jejuni* strains and other bacterial species.

Acknowledgments

We would like to dedicate this chapter to our fantastic mentors and colleagues at the National Research Council – Institute for Biological Sciences who are always pushing forward the boundaries in glycomics. In particular, for these studies, we thank Jean-Robert Brisson, Harold Jarrell and Jianjun Li. This work was funded by the NRC Genomics and Health Initiative.

References

1. Helenius, A. and Aebi, M. (2001) Intracellular functions of *N*-linked glycans. *Science* **291**, 2364–2369.
2. Abu-Qarn, M., Yurist-Doutsch, S., Giordano, A., Trauner, A., Morris, H. R., Hitchen, P., Medalia, O., Dell, A., and Eichler, J. (2007) *Haloferax volcanii* AglB and AglD are involved in *N*-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer. *J. Mol. Biol.* **374**, 1224–1236.
3. Chaban, B., Voisin, S., Kelly, J., Logan, S. M., and Jarrell, K. F. (2006) Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae N*-linked glycans: insight into *N*-linked glycosylation pathways in Archaea. *Mol. Microbiol.* **61**, 259–268.
4. Yurist-Doutsch, S., Chaban, B., VanDyke, D. J., Jarrell, K. F., and Eichler, J. (2008) Sweet to the extreme: protein glycosylation in Archaea. *Mol. Microbiol.* **68**, 1079–1084.
5. Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W., and Aebi, M. (2002) *N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**, 1790–1793.
6. Szymanski, C. M., and Wren, B. W. (2005) Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Microbiol.* **3**, 225–237.
7. Nothaft, H., Amber, S., Aebi, M., and Szymanski, C. M. (2008) *N*-Linked Protein Glycosylation in *Campylobacter*. In: *Campylobacter* (3rd edn.) Nachamkin, I., Szymanski, C. M., and Blaser, M. J. (eds). Washington, DC: ASM Press, pp. 447–469.
8. Hendrixson, D. R. and DiRita, V. J. (2004) Identification of *Campylobacter jejuni* genes

involved in commensal colonization of the chick gastrointestinal tract. *Mol. Microbiol.* **52**, 471–484.
9. Larsen, J. C., Szymanski, C., and Guerry, P. (2004) *N*-linked protein glycosylation is required for full competence in *Campylobacter jejuni* 81-176. *J. Bacteriol.* **186**, 6508–6514.
10. Szymanski, C. M., Goon, S., Allan, B., and Guerry, P. (2005) Protein glycosylation in *Campylobacter*. In: *Campylobacter*: Molecular and Cellular Biology. Ketley, J. and Konkel, M. (eds). Norfolk: Horizon Bioscience, pp. 259–273.
11. Szymanski, C. M., Logan, S. M., Linton, D., and Wren, B. W. (2003) Campylobacter – a tale of two protein glycosylation systems. *Trends Microbiol.* **11**, 233–238.
12. Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J., and Guerry, P. (1999) Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* **32**, 1022–1030.
13. Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., Mills, D. C., Watson, D. C., Hernandez, M., Kelly, J. F., Wacker, M., and Aebi, M. (2006) Definition of the bacterial *N*-glycosylation site consensus sequence. *EMBO J.* **25**, 1957–1966.
14. Young, N. M., Brisson, J. -R., Kelly, J., Watson, D. C., Tessier, L., Lanthier, P. H., Jarrell, H. C., Cadotte, N., St Michael, F., Aberg, E., and Szymanski, C. M. (2002) Structure of the *N*-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J. Biol. Chem.* **277**, 42530–42539.
15. Alaimo, C., Catrein, I., Morf, L., Marolda, C. L., Callewaert, N., Valvano, M. A., Feldman, M. F., and Aebi, M. (2006) Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J.* **25**, 967–976.
16. Liu, X., McNally, D. J., Nothaft, H., Szymanski, C. M., Brisson, J. -R., and Li, J. (2006) Mass spectrometry-based glycomics strategy for exploring *N*-linked glycosylation in eukaryotes and bacteria. *Anal. Chem.* **78**, 6081–6087.
17. Abu-Qarn, M., Eichler, J., and Sharon, N. (2008) Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr. Opin. Struct. Biol.* **18**, 544–550.
18. Glover, K. J., Weerapana, E., and Imperiali, B. (2005) *In vitro* assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic *N*-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14255–14259.
19. Weerapana, E., Glover, K. J., Chen, M. M., and Imperiali, B. (2005) Investigating bacterial *N*-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **127**, 13766–13767.
20. Merry, T., Taverna, M., Tran, T., and Harvey, D. (2002) Analyses of glycans of recombinant glycoproteins. In: Cell Engineering. Al-Rubeai, M. (ed). Dordrecht: Springer, pp. 1–60.
21. Kelly, J., Jarrell, H., Millar, L., Tessier, L., Fiori, L. M., Lau, P. C., Allan, B., and Szymanski, C. M. (2006) Biosynthesis of the *N*-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. *J. Bacteriol.* **188**, 2427–2434.
22. Linton, D., Allan, E., Karlyshev, A. V., Cronshaw, A. D., and Wren, B. W. (2002) Identification of *N*-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. *Mol. Microbiol.* **43**, 497–508.
23. Sambrook, J. and Russell, D. W. (2001) Molecular Cloning: A Laboratory Manual (3rd edn). Cold Spring Harbor Laboratory, New York.
24. Labigne-Roussel, A., Harel, J., and Tompkins, L. (1987) Gene transfer from *Escherichia coli* to *Campylobacter* species: development of shuttle vectors for genetic analysis of *Campylobacter jejuni*. *J. Bacteriol.* **169**, 5320–5323.
25. St Michael, F., Szymanski, C. M., Li, J., Chan, K. H., Khieu, N. H., Larocque, S., Wakarchuk, W. W., Brisson, J. -R., and Monteiro, M. A. (2002) The structures of the lipooligosaccharide and capsule polysaccharide of *Campylobacter jejuni* genome sequenced strain NCTC 11168. *Eur. J. Biochem.* **269**, 5119–5136.
26. Szymanski, C. M., Michael, F. S., Jarrell, H. C., Li, J., Gilbert, M., Larocque, S., Vinogradov, E., and Brisson, J. -R. (2003) Detection of conserved *N*-linked glycans and phase-variable lipooligosaccharides and capsules from campylobacter cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy. *J. Biol. Chem.* **278**, 24509–24520.
27. Meiboom, S. and Gill, D. (1958) Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.* **29**, 688–691.
28. Brisson, J. -R., Sue, S. C., Wu, W. G., McManus, G., Nghia, P. T., and Uhrin, D. (2002) NMR of carbohydrates: 1D homonuclear selective methods. In: NMR

Spectroscopy of Glycoconjugates. Jimenez-Barbero, J. and Peters, T. (eds). Weinhem, Germany: Wiley-VCH, pp. 59–93.
29. Uhrín, D. and Brisson, J. -R. (2000) Structure determination of microbial polysaccharides by high resolution NMR spectroscopy. In: NMR in Microbiology: Theory and Applications. Barbotin, J. N. and Portais, J. C. (eds). Wymondham, UK: Horizon Scientific Press, pp. 165–210.
30. McNally, D. J., Lamoureux, M., Li, J., Kelly, J., Brisson, J. -R., Szymanski, C. M., and Jarrell, H. C. (2006) HR-MAS NMR studies of ^{15}N labeled cells confirm the structure of the O-methyl phosphoramidate CPS modification in *Campylobacter jejuni* and provide insight into its biosynthesis. *Can. J. Chem.* **84**, 676–684.
31. McNally, D. J., Lamoureux, M. P., Karlyshev, A. V., Fiori, L. M., Li, J., Thacker, G., Coleman, R. A., Khieu, N. H., Wren, B. W., Brisson, J. -R., Jarrell, H. C., and Szymanski, C. M. (2007) Commonality and biosynthesis of the O-methyl phosphoramidate capsule modification in *Campylobacter jejuni*. *J. Biol. Chem.* **282**, 28566–28576.
32. McNally, D. J., Jarrell, H. C., Khieu, N. H., Li, J., Vinogradov, E., Whitfield, D. M., Szymanski, C. M., and Brisson, J. -R. (2006) The HS:19 serostrain of *Campylobacter jejuni* has a hyaluronic acid-type capsular polysaccharide with a nonstoichiometric sorbose branch and O-methyl phosphoramidate group. *Febs J.* **273**, 3975–3989.
33. McNally, D. J., Jarrell, H. C., Li, J., Khieu, N. H., Vinogradov, E., Szymanski, C. M., and Brisson, J. -R. (2005) The HS:1 serostrain of *Campylobacter jejuni* has a complex teichoic acid-like capsular polysaccharide with nonstoichiometric fructofuranose branches and O-methyl phosphoramidate groups. *Febs J.* **272**, 4407–4422.
34. Guerry, P., Ewing, C. P., Schoenhofen, I. C., and Logan, S. M. (2007) Protein glycosylation in *Campylobacter jejuni*: partial suppression of *pglF* by mutation of *pseC*. *J. Bacteriol.* **189**, 6731–6733.

Chapter 17

Characterization of Polysaccharides Using Mass Spectrometry for Bacterial Serotyping

Eleonora Altman and Jianjun Li

Abstract

Mass spectrometry provides a rapid and reliable method for characterization of bacterial polysaccharides. Application of the in-source fragmentation technique to promote the formation of structurally relevant repeating units of heterogeneous capsular polysaccharides and O-chain polysaccharides has proven to be particularly useful for detection of non-carbohydrate functionalities and subtle differences arising across bacterial serotypes. Here we discuss application of these methods to the direct analysis of bacterial cells allowing for rapid analysis of cell surface polysaccharide antigens and providing a basis for serological typing and epidemiological surveillance studies of human and animal pathogens.

Key words: Lipopolysaccharide, mass spectrometry, preconcentration, serotyping, classification.

1. Introduction

Lipopolysaccharide (LPS) is the major cell surface component of *Aeromonas salmonicida*, a Gram-negative bacteria implicated in furunculosis of salmonid fish (1). The LPS molecule consists of three distinct regions: an O-chain polysaccharide, a core oligosaccharide, and a lipid A moiety (2). The differences between the composition and structure of the LPS O-chain polysaccharide components provide the basis for serological classification of bacterial isolates. We have analyzed both typical and atypical isolates of *A. salmonicida* according to the structures of their respective LPS O-chain polysaccharides using a microanalytical

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_17,

capillary electrophoresis–mass spectrometry (CE–MS) and have shown that *A. salmonicida* strains examined fall into three distinct structural groups, types A, B, and C (3). A relatively simple and robust in-source collision-induced dissociation (IS-CID) technique was developed that allowed a direct analysis on as little as 10^8 to 10^{10} bacterial cells (3, 4). This method provided a new MS-based approach to LPS-based classification of *A. salmonicida*. Furthermore, application of magnetic beads (Dynabeads) coated with *A. salmonicida* LPS-specific polyclonal antisera to MS-based characterization of *A. salmonicida* LPS resulted in at least a 1000-fold improvement of the detection sensitivity affording direct CE–MS analysis of in vivo-cultured cell samples

2. Materials

2.1. Bacterial Culture and Isolation of LPS

1. Difco™ Tryptic Soy Agar or Difco™ Tryptic Soy Broth without glucose (Becton, Dickinson and Company, Sparks, MD).
2. 0.01 M Dulbecco's phosphate buffered saline, pH 7.4 (10×, Sigma, St. Louis, MO).
3. 1% (w/v) phenol (BDH Inc., Toronto, Ontario, Canada).
4. 2.5% (w/v) sodium chloride (saline) (ACP Chemicals Inc., Montreal, Quebec).
5. 95% phenol (BDH Inc., Toronto, Ontario, Canada).
6. Enzymes: lysozyme, deoxyribonuclease I (DNase), ribonuclease A (RNase), and trypsin (Sigma, St. Louis, MO).
7. Water bath.
8. Mechanical stirrer (EUROSTAR power-b, IKA Labortechnik, Germany).
9. Stainless steel beakers, 1 L and 2 L.

2.2. Purification of LPS and Preparation of the O-Chain Polysaccharide

1. 2% (w/v) acetic acid (EM Science, Darmstadt, Germany).
2. Bio-Gel P-2 gel, fine 45–90 μm (Bio-Rad, Hercules, CA).
3. Bio-Gel P-10 gel, fine 45–90 μm (Bio-Rad, Hercules, CA).
4. C column 16/100 (1.6 × 100 cm) (Pharmacia, Uppsala, Sweden).

2.3. Hydrolysis of LPS and Preparation of Alditol Acetates

1. 2 M trifluoroacetic acid (Fisher Scientific, Fair Lawn, NJ).
2. Sodium borohydride (Sigma, St. Louis, MO).
3. Methanol (EMD Chemicals Inc., Gibbstown, NJ).
4. 10% acetic acid (ACP Chemicals Inc., Montreal, Quebec).

5. Pyridine (ACP Chemicals Inc., Montreal, Quebec).
6. Acetic anhydride (ACP Chemicals Inc., Montreal, Quebec).
7. Reacti-Therm heating/stirring module (Pierce, Rockford, IL).
8. Reacti-Block C-1 (for 3.5 mL screw cap septum vials) (Pierce, Rockford, IL).
9. Screw cap septum vials (3.5 mL) (Pierce, Rockford, IL).
10. Screw caps, open top (Pierce, Rockford, IL).
11. Teflon silicone discs (12 mm, white) (Pierce, Rockford, IL).

2.4. Methylation Analysis

1. Methyl iodide (iodomethane) (Sigma, St. Louis, MO).
2. Dimethylsulfoxide (EM Science, Darmstadt, Germany).
3. Sodium hydroxide (ACP Chemicals Inc., Montreal, Quebec).
4. Sodium borodeuteride (Sigma, St. Louis, MO).
5. Dichloromethane (J.T. Baker Inc., Phillipsburg, NJ).
6. Anhydrous sodium sulfate (BDH Inc., Toronto, Ontario, Canada).
7. Reacti-Therm heating/stirring module (Pierce, Rockford, IL).
8. Reacti-Block C-1 (for 3.5 mL screw cap septum vials) (Pierce, Rockford, IL).
9. Screw cap septum vials (3.5 mL) (Pierce, Rockford, IL).
10. Screw caps, open top (Pierce, Rockford, IL).
11. Teflon silicone discs (12 mm, white) (Pierce, Rockford, IL).

2.5. Preparation of In Vitro-Cultured or In Vivo-Cultured Cell Samples

1. 10 mM Dulbecco's phosphate buffered saline, pH 7.4 (10×, Sigma, St. Louis, MO).
2. 20 mM ammonium acetate, pH 7.5 (Fisher Scientific, Fair Lawn, NJ).
3. Enzymes: deoxyribonuclease I (DNase), ribonuclease A (RNase), and proteinase K (Sigma, St. Louis, MO).
4. 2% acetic acid (ACP Chemicals Inc., Montreal, Quebec).

2.6. Immunomagnetic Binding and Elution of Target Antigens

1. Superparamagnetic polystyrene beads precoated with sheep anti-rabbit IgG antibodies (Dynabeads M-280; Dynal AS, Oslo, Norway).
2. A Dynal Sample Mixer (Dynal Biotech Sample Mixer, Dynal AS, Oslo, Norway).

3. A magnetic particle concentrator (MPC-s, Dynal AS, Oslo, Norway).
4. *A. salmonicida* LPS-specific polyclonal rabbit antiserum (*see* **Note 1**).
5. Solution of 10 mM Dulbecco's phosphate buffered saline, pH 7.4, containing 0.1% BSA (bovine serum albumin).
6. Bovine serum albumin (BSA) (Sigma, St. Louis, MO).
7. 0.2 M glycine–HCl buffer, pH 2.2 (ACP Chemicals Inc., Montreal, Quebec).
8. 0.1 N NaOH (ACP Chemicals Inc., Montreal, Quebec).

2.7. Capillary Electrophoresis System

1. Fused-silica capillaries with 185 μm o.d. × 50 μm i.d. (Polymicro Technologies, Phoenix, AZ).
2. Methanol and isopropanol (EM Science, Gibbstown, NJ).
3. Anhydrous hydrazine (Fisher Scientific, Fair Lawn, NJ).
4. Ammonium acetate (Fisher Scientific, Fair Lawn, NJ).
5. Formic acid (BDH Inc., Toronto, Ontario, Canada).

3. Methods

Capillary electrophoresis–mass spectrometry (CE–MS) has proven to be a powerful and versatile tool for rapid analysis and structural characterization of bacterial lipopolysaccharides. We have employed an in-source collision-induced dissociation (IS-CID) technique for CE–MS-based analysis of typical and atypical isolates of *A. salmonicida* performed directly on bacterial cells (3, 4). The results of this study have led to the identification of three structural LPS variants among *A. salmonicida* isolates examined, types A, B, and C, according to the structure of their corresponding O-chain polysaccharides (**Fig. 17.1**). Each group displayed a characteristic MS fragmentation pattern which could be used for fingerprinting of a particular structural type. The majority of typical *A. salmonicida* isolates belonged to structural type A and displayed a complete O-chain polysaccharide structure composed of the trisaccharide repeating unit containing L-rhamnose (L-Rha), 2-acetamido-2-deoxy-D-mannose (D-ManNAc) and D-glucose (D-Glc) and non-carbohydrate O-acetyl substituent (OAc) (5). Identical fragmentation patterns in the CE–MS spectra were obtained for all *A. salmonicida* isolates belonging to this structural group represented by *A. salmonicida* strain A449. The second group, structural type B, was represented by the O-chain polysaccharide of *A. salmonicida* strain

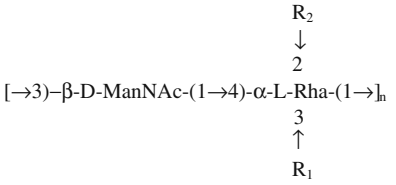


Fig. 17.1. Structural variants of the O-chain polysaccharide of *A. salmonicida* LPS: (A) type A, representative strain A449 [R_1= α-D-Glc-(1→); R_2=OAc (75%)]; and (B) type B, representative strain 33659 (R_1=H; R_2 = H); (C) type C, representative strain N4705 [R_1 = H; R_2 = OAc (30%)].

33659 and consisted of disaccharide repeating units containing L-Rha and D-ManNAc only. Finally, the third structural group, type C, was represented by the O-chain polysaccharide of atypical strain N4705 consisting of disaccharide repeating units containing L-Rha and D-ManNAc, and non-carbohydrate O-acetyl component. The composition and substitution pattern of constituent sugars in the purified O-chain polysaccharides of *A. salmonicida* strains 33659 and N4705 were confirmed by compositional and linkage analysis and the results compared with structural data obtained for the O-chain polysaccharide of *A. salmonicida* strain A449 (3).

Application of commercially available magnetic beads (Dynabeads) coated with *A. salmonicida* LPS-specific polyclonal antisera to CE–MS-based characterization of bacterial LPS resulted in the improvement of the concentration limit of detection. The developed affinity-based preconcentration procedure was applied to CE–MS analysis of the LPS O-chain polysaccharide from in vivo-grown *A. salmonicida* cells co-cultured with trout macrophages in intraperitoneal chambers (6). Bacterial cell suspensions from three representative *A. salmonicida* strains, A449 (type A), 33659 (type B), and N4705 (type C), grown in vivo under co-culture conditions were analyzed. In this procedure (*see* below), the availability and specificity of antibodies was a key factor for the successful isolation and characterization of the O-chain polysaccharide LPS component resulting in the removal of extraneous impurities responsible for the signal suppression of the target LPS molecule and allowing for at least 1000-fold improvement in detection sensitivity. This general approach is applicable to the characterization of bacterial antigens from any animal or human origin for which a specific antibody/capture agent is available (*see* data figures later).

3.1. Bacterial Culture and Isolation of LPS by Hot Phenol–Water Extraction Method

1. The bacteria, *A. salmonicida* strains A449, N4705, and 33659, are cultured on Tryptic Soy Agar or Tryptic Soy Broth without glucose at 18°C or 25°C for 48–72 h (7).
2. Bacterial cells are killed with 1% (w/v) phenol solution (22°C, 4 h).

3. Bacterial cells are washed with 0.01 M phosphate buffered saline, pH 7.4, and harvested by low-speed centrifugation (3000 rpm, 4°C, 25 min).
4. Bacterial cells, 100 g wet mass, are suspended in 2.5% (w/v) aqueous sodium chloride solution (500 mL) and gently stirred until fully suspended (1 h, 22°C).
5. The cells are spun down at 7000 rpm for 20 min at 4°C.
6. The cell pellet is re-suspended in 400 mL of distilled water in a stainless steel container until a homogeneous cell suspension is formed (*see* **Note 2**).
7. The cell suspension is treated with lysozyme (1 mg/100 g cells, wet weight) and incubated at 37°C for 1 h with constant stirring.
8. To 100 mL of the cell suspension 0.203 g of $MgCl_2 \times 6 H_2O$ is added and dissolved (to make 10 mM solution), followed by 2 mg each, of DNase and RNase, and the cell suspension is incubated at 37°C for 1–2 h with constant stirring.
9. The cell suspension is treated with trypsin (10 mg/100 g cells, wet weight) and incubated at 37°C for 1 h with constant stirring.
10. Distilled water (100 mL) is added and the temperature brought to 70°C.
11. Equal volume of phenol solution (95% w/v), 500 mL, is heated to 75°C in a boiling water bath (or a microwave oven) and added to a hot cell suspension. Caution has to be exercised; a lab coat and protective goggles must be worn.
12. Extraction with phenol is carried out in a fume hood in a stainless steel container. Caution has to be exercised; a lab coat and protective goggles must be worn. The mixture is stirred vigorously with a mechanical stirrer at 70°C for 20 min.
13. The cell suspension is cooled to a room temperature and transferred to 500 mL phenol-resistant centrifuge bottles and centrifuged overnight (3000 rpm, 4°C).
14. Phenol and aqueous layers are collected separately, transferred to a dialysis tubing, molecular mass cutoff 10,000–14,000 Da, and dialyzed against running tap water for 4–5 days or until phenol free (no phenol smell) (*see* **Notes 3 and 4**).
15. The dialysis tubing contents are filtered through a separatory funnel filled with a glass wool and lyophilized to yield a crude LPS (*see* **Note 5**).

3.2. Purification of LPS and Preparation of the O-Chain Polysaccharide

1. Crude LPS is dissolved in 1% saline solution (200 mL) and centrifuged (3000 rpm, 4°C, 30 min) and the supernatant is subjected to ultracentrifugation (105,000×*g*, 4°C, 16 h).
2. The pellet is dissolved in a minimal amount of distilled water, transferred to ultracentrifuge tubes, and subjected to another round of ultracentrifugation (105,000×*g*, 4°C, 16 h).
3. The final LPS pellet is lyophilized.
4. Purified aqueous phase LPS (60 mg) is hydrolyzed with 0.2 M acetic acid (100°C, 2 h).
5. The reaction mixture is cooled down on ice.
6. The insoluble lipid A is removed by low-speed centrifugation.
7. The supernatant is lyophilized and fractionated by gel chromatography on a Bio-Gel P-2 column using pyridinium acetate as an eluant (0.5 M, pH 5.4).
8. Fractions corresponding to the O-chain polysaccharide are pooled and lyophilized.
9. The fraction containing O-chain polysaccharide is further purified on a Bio-Gel P-10 column (Bio-Rad) (*see* **Note 6**).

3.3. Preparation of Alditol Acetates

1. The O-chain polysaccharide sample (0.5 mg) is hydrolyzed with 2 M trifluoroacetic acid (100°C, 16 h) (*see* **Note 7**) (8).
2. The sample is concentrated to dryness, dissolved in 0.5 mL distilled water (0.5 mg), and reduced with sodium borohydride (10 mg, 22°C, 1 h). It is neutralized with 10% (w/v) aqueous solution of acetic acid, concentrated, and distilled with methanol to remove borate (4×0.5 mL).
3. The reduction products are acetylated with acetic anhydride/pyridine (1:1, v/v) at 100°C for 1 h.
4. Distilled water, 0.5 mL, is added and the alditol acetate derivatives are extracted thrice with dichloromethane (0.5 mL) and the organic layers combined.
5. Combined organic phase is dried by the addition of a few crystals of anhydrous sodium sulfate, filtered through a glass wool, and concentrated to dryness (*see* **Note 8**).
6. The residue is dissolved in dichloromethane and alditol acetates are analyzed by GLC using a Hewlett–Packard chromatograph equipped with a 30-m DB-17 capillary column [180°C (2 min), 3.5°C/min to 260°C (23 min)] and by GC–MS in the electron impact (EI) mode recorded using a Varian Saturn 2000 mass spectrometer. An example of results produced is shown in **Table 17.1**.

Table 17.1
Compositional analysis of the O-chain polysaccharide of *A. salmonicida* strains A449, 33659 and N4705[a]

	Mol %		
Strain	**Glucose**	**Rhamnose**	**Mannosamine**
A449	42.2[b]	31.2	26.6
33659	–	51.0	49.0
N4705	–	47.2	52.8

[a]Relative molar ratios.
[b]Based on the detector response (total ion count) (Reproduced from (3) with permission from Elsevier Science).

3.4. Methylation Analysis

1. Polysaccharide sample (3–5 mg) is dissolved in dry DMSO (0.5 mL) and stirred overnight at 22°C (9).
2. Freshly finely powdered sodium hydroxide (25 mg) is added and the solution is stirred for 2 h at 22°C (*see* **Note 9**).
3. Methyl iodide (1 mL) is added and the mixture is stirred for 4 h at 22°C. The reaction is terminated by addition of 0.5 mL distilled water.
4. The product is extracted with chloroform (3×0.5 mL), the organic layers combined and dried by addition of a few crystals of anhydrous sodium sulfate and concentrated to dryness.
5. The permethylated polysaccharide is subjected to hydrolysis with 2 M trifluoroacetic acid (100°C, 18 h) or using a procedure of Stellner et al. for amino sugar-containing polysaccharides (10) and converted to alditol acetate derivatives as described above. Please note that in this case the reduction step is carried out with sodium borodeuteride (18 h, 22°C).
6. Partially methylated alditol acetates are analyzed by GC–MS (11) as previously described for alditol acetates. An example of results produced is shown in **Table 17.2**.

3.5. Preparation of In Vitro-Cultured or In Vivo-Cultured Cell Samples

1. The cell pellet is suspended in 360 μL 20 mM ammonium acetate buffer, pH 7.5. RNase and DNase stock solutions are added (final concentration 10 μg/mL each) and the mixture is incubated at 37°C for 2 h.
2. The cell pellet is recovered by low-speed centrifugation, suspended in 350 μL 10 mM phosphate buffered saline (PBS),

Table 17.2
Methylation analysis of the O-chain polysaccharide of *A. salmonicida* strains A449, 33659, and N4705[a]

Sugar linkage	Mol %		
	Strain A449	Strain 33659	Strain N4705
4-Substituted Rha	3.1[b]	45.9	45.1
Terminal Glc	39.1	–	–
3,4-Substituted Rha	33.6	–	–
3-Substituted ManNAc	24.2	54.1	54.9

[a]Relative molar ratios.
[b]Based on detector response (total ion count) (Reproduced from (3) with permission from Elsevier Science).

pH 7.4, and proteinase K stock solution is added (final concentration 200 μg/mL).

3. The mixture is incubated at 60°C for 1 h, heated at 100°C for 5 min, then centrifuged by low-speed centrifugation and lyophilized.
4. Lyophilized sample is dissolved in 0.5 mL 2% acetic acid, heated at 100°C for 2 h, lyophilized, and analyzed by CE–MS. Examples of CE–MS spectra of the O-chain polysaccharide of *A. salmonicida* from in vitro-cultured bacterial cell samples are shown in **Fig. 17.2**.

3.6. Immunomagnetic Binding and Elution of Target Antigens

*3.6.1. Coating of Dynabeads M-280 with Rabbit Anti-*A. salmonicida *LPS Polyclonal Sera*

1. Magnetic beads (0.5 mL, approx. 6×10^8 beads/mL) and corresponding primary rabbit polyclonal antiserum (1:20 dilution) (specific for *A. salmonicida* strains A449, N4705 or 33659 or their mixture) are mixed well by tilting and rotating in 10 mM PBS, pH 7.4, containing 0.1% BSA (0.5 mL) at 22°C for 1.5 h, using a Dynal sample mixer.
2. Magnetic beads are washed with 10 mM PBS, pH 7.4, containing 0.1% BSA (3×0.5 mL).
3. Magnetic beads are collected using a Dynal magnetic particle concentrator.

3.6.2. Immunomagnetic Binding and Elution of Carbohydrate Antigens of A. salmonicida

1. *A. salmonicida* LPS antisera-coated magnetic beads are incubated in 10 mM PBS, pH 7.4, containing 0.1% BSA (0.5 mL) at 22°C for 3 h with either O-chain polysaccharide samples (0.5 mg/mL), or enzyme-pretreated and delipidated bacterial cell samples (8×10^8 cells/0.5 mL) (5), followed by three wash cycles with the same buffer.

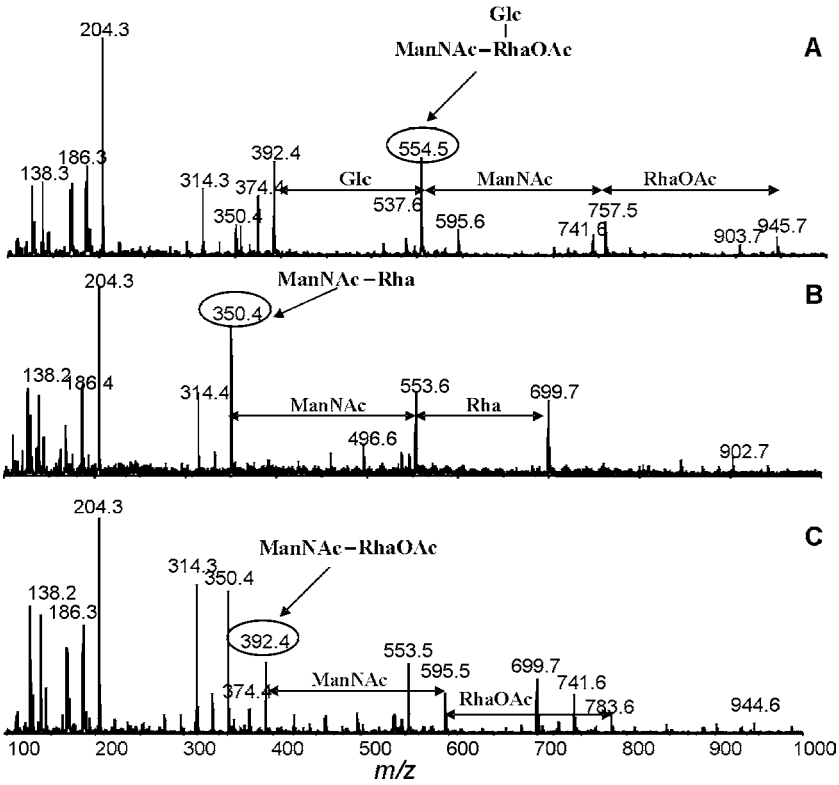


Fig. 17.2. CE–MS analysis of in vitro-cultured whole cell lysates of *A. salmonicida* in the positive ion mode: (A) type A, representative strain A449; (B) type B, representative strain 33659; (C) type C, representative strain N4705 (reproduced from (3) with permission from Elsevier Science).

2. Carbohydrate antigens of *A. salmonicida* are eluted from the antigen–bead complex with 0.5 mL 0.2 M glycine–HCl buffer (pH 2.2) and the mixture is incubated with gentle rotation at 22°C for 30 min. The elution procedure is repeated once and the supernatant is collected (*see* **Note 10**).
3. The supernatant is neutralized with 0.1 N NaOH and desalted using a centrifugal filter device (Micron-3 K, MW cutoff 3000, Pall Corporation, Novato, CA, USA).
4. The recovered carbohydrate antigen solution is sonicated for 3 h at 40°C and analyzed directly by a Crystal Model 310 CE instrument coupled to an API 3000 mass spectrometer. Examples of CE–MS analysis of the O-chain polysaccharide from *A. salmonicida* LPS and from in vivo-cultured bacterial cells following preconcentration procedure with sonication step are shown in **Fig. 17.3**.

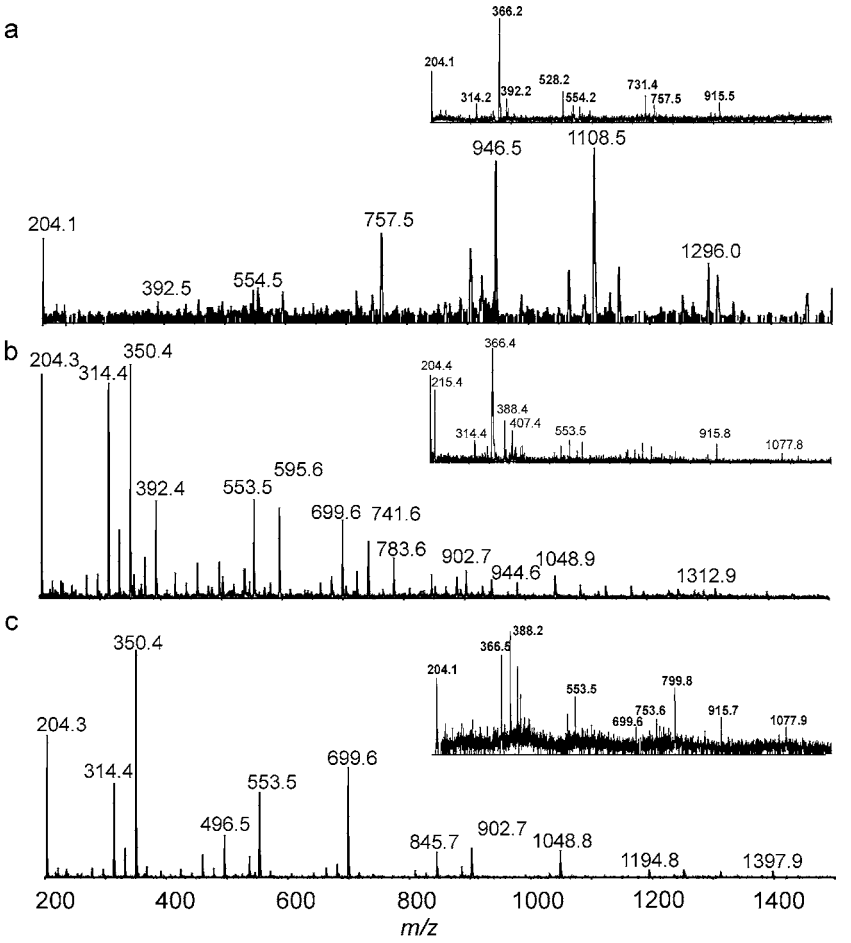


Fig. 17.3. CE–MS analysis of *A. salmonicida* cells from co-culture in vivo implants with preconcentration procedure: (a) strain A449; (b) strain N4705; and (c) strain 33659. The insets correspond to the CE–MS analysis of each sample without the preconcentration procedure. The detected ions and their corresponding fragment compositions: *m/z* 204.5, ManNAc; *m/z* 350.6, ManNAc Rha; *m/z* 392.6, ManNAc RhaOAc; *m/z* 553.7, $ManNAc_2$ Rha; *m/z* 595.5, $ManNAc_2$ RhaOAc; *m/z* 699.7, $ManNAc_2$ Rha_2; *m/z* 741.6, $ManNAc_2$ Rha RhaOAc; *m/z* 757.7, $ManNAc_2$ RhaOAc Glc; *m/z* 783.6, $ManNAc_2$ $RhaOAc_2$; *m/z* 902.5, $ManNAc_3$ Rha_2; *m/z* 944.6, $ManNAc_3$ Rha RhaOAc; *m/z* 945.7, $ManNAc_2$ $RhaOAc_2$ Glc. Average mass units were used for calculation of molecular mass based on proposed composition as follows: Glc, 162.1; ManNAc, 203.2; Rha, 146.1; OAc (acetate), 42.0.

3.7. Capillary Electrophoresis–Mass Spectrometry

1. Capillary electrophoresis is performed using a Prince CE system (Prince Technologies, The Netherlands).
2. The CE system is coupled to an API 3000 mass spectrometer via a microspray interface (AB/MDS Sciex, Concordia, Canada).
3. A sheath solution (isopropanol–methanol, 2:1) is delivered at a flow rate of 1 μL/min to a low dead volume tee (250 μm i.d.).

4. Separations are obtained on approximately 90 cm length of bare fused-silica capillary using 15 mM ammonium acetate in deionized water, pH 7.0.
5. Microscale CE–MS analysis on bacterial cells in a positive mode is performed using an in-source collision-induced dissociation (IS-CID) technique, i.e., increasing the orifice voltage from +30 V to +200 V to induce the formation of structurally relevant fragment ions.

4. Notes

1. LPS-specific polyclonal antiserum was raised in rabbits following immunizations with purified LPS (6). When unavailable, polyclonal antisera could be replaced with monoclonal antibodies.
2. Phenol solution is acidic. When extracting LPS containing acid-labile groups or sugar residues distilled water could be replaced with 10 mM PBS buffer, pH 7.4.
3. A simple device, consisting of the suction Erlenmeyer flask with Teflon tubing attached to a Pasteur pipette or a large syringe could be used to collect aqueous and phenol layers.
4. Usually LPS is present in the aqueous layer but exceptions are known (particularly when O-chain polysaccharide consists of amino sugars). Both aqueous and phenol layers should be collected, dialyzed, lyophilized, and checked for the presence of sugars.
5. Separatory funnel filled with a glass wool (to avoid carbohydrate contamination) is used for filtration of the dialysate and elimination of the remaining bacterial debris.
6. Gel permeation chromatography could be carried out using a number of different size exclusion gels, such as Sephadex G-50, Sephadex G-100, and Bio-Gel P-10. Bio-Gel P-2 column is frequently used for desalting and separation of the low-molecular-mass products and impurities.
7. Trifluoroacetic (TFA) acid is volatile and is therefore convenient for hydrolysis, especially on a small scale. Hydrolysis conditions (concentration, time, and temperature) could be optimized but usually 2 M TFA (100°C, 16 h) or 4 M TFA (125°C, 4 h) gives satisfactory results.
8. A disposable Pasteur pipette filled with a glass wool could be used for filtration of the dichloromethane solution.

9. Powdered sodium hydroxide is prepared by grinding a few NaOH pellets in a mortar just before the reaction.
10. Optimal elution conditions should be determined experimentally. The manufacturer's protocol suggests the use of 0.1 M glycine–HCl buffer (pH 2.5) for elution of the antigen. We found that these conditions did not result in disruption of the antibody–antigen complex and instead incubated the mixture in 0.2 M glycine–HCl buffer (pH 2.2) at 22°C for 30 min.

Acknowledgments

The authors would like to thank Dr. Laura L. Brown for encouragement and support of these studies. This work was supported by the National Research Council's Genomics and Health Initiative (GHI).

References

1. Bullock, G. L., Griffin, B. R., and Stuckey, H. M. (1980) Detection of *Corynebacterium salmoninus* by direct fluorescent antibody test. *Can. J. Fish. Aquat. Sci.* **37**, 719–721.
2. Raetz, C. R. H. and Whitfield, C. (2002) Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**, 635–700.
3. Wang, Z., Liu, X., Dacanay, A., Harrison, B. A., Fast, M., Colquhoun, D. J., Lund, V., Brown, L. L., Li, J., and Altman, E. (2007) Carbohydrate analysis and serological classification of typical and atypical isolates of *Aeromonas salmonicida*: A rationale for the lipopolysaccharide-based classification of *A. salmonicida*. *Fish Shellfish Immunol.* **23**, 1095–1106.
4. Li, J., Wang, Z., and Altman, E. (2005) In-source fragmentation and analysis of polysaccharides by capillary electrophoresis–mass spectrometry. *Rapid Commun. Mass Spectrom.* **19**, 1305–1314.
5. Wang, Z., Vinogradov, E., Larocque, S., Harrison, B. A., Li, J., and Altman, E. 2005. Structural and serological characterization of O-chain polysaccharide of *Aeromonas salmonicida* strains A449, 80204 and 80204-1. *Carbohydr. Res.* **340**, 693–700.
6. Garduño, R. A., Thornton, J. C., and Kay, W. W. (1993) *Aeromonas salmonicida* grown in vivo. *Infect. Immun.* **61**, 3854–3862.
7. Westphal, O. and Jann, K. (1965) Bacterial polysaccharides. Extraction with phenol-water and further applications of the procedure. *Methods Carbohydr. Chem.* **5**, 83–91.
8. Sawardeker, J. H., Sloneker, J. H., and Jeannes, A. (1967) Quantitative determination of monosaccharides as their alditol acetates by gas liquid chromatography. *Anal. Chem.* **39**, 1602–1604.
9. Ciucanu, I. and Kerek, F. (1984) A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **131**, 209–217.
10. Stellner, K., Saito, H., and Hakomori, S.-I. (1973) Determination of aminosugar linkages in glycolipids by methylation. Aminosugar linkages of ceramide pentasaccharides of rabbit erythrocytes and of Forssman antigen. *Arch. Biochem. Biophys.* **155**, 464–472.
11. Björndal, H., Lindberg, B. and Svensson, S. (1967) Gas-liquid chromatography of partially methylated alditols as their acetates. *Acta Chem. Scand.* **21**, 1801–1804.

Chapter 18

Is Permethylation Strategy Always Applicable to Protein N-Glycosylation Study?: A Case Study on the O-Acetylation of Sialic Acid in Fish Serum Glycans

Xin Liu and Luis Afonso

Abstract

O-Acetylation is one of the major modifications of sialic acids that significantly alters biological properties of the parent molecule. These O-acetylated forms are components of the cellular membrane and can affect physiological and pathological responses. Understanding the role of N-glycans in physiology is of increasing relevance to cellular biologists in various disciplines who study glycoproteomics yet lack information regarding the function of the attached glycans. However, permethylation, the most common mass spectrometric analytical means, leads to the loss of O-linked acetyl groups in sialic acids. In this chapter, we demonstrated that O-acetylation of sialic acid in Atlantic salmon serum N-glycan can be well investigated by capillary electrophoresis–mass spectrometry.

Key words: Atlantic salmon, serum, glycosylation, O-acetylation, sialic acid, permethylation.

1. Introduction

Protein glycosylation represents a major class of post-translational modifications that dramatically enhance the functional diversity of proteins (1). Sialic acids are a family of nine-carbon sugars which exist as two major types of sialic acids, N-acetylneuraminic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc), with various substitutions. They are typically found at the terminal of oligosaccharides on different glycoconjugates, e.g., glycoproteins and glycolipids. Up to now, approximately 40 sialic acid derivatives have been identified in nature. The most common modification of hydroxyl groups of sialic acid is O-acetylation, which often

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_18,

occurs at 4, 7, 8 and/or 9 positions (2). It has been realized that O-acetylated sialic acids might be involved in a variety of biological phenomena, such as virus binding, lectin recognition, tumor antigenicity, tissue morphogenesis, and cell/cell interaction (3, 4).

Modifications of sialic acids were found to developmentally regulate gene expression in a variety of systems and also to be tissue-specific (5–14). Due to a low abundance and instability, quantitative analysis of O-acetylated sialic acid has been a challenging task. In the past decades, several analytical procedures were employed in the study of sialic acid modification, in which sialic acids were released by a mild acid hydrolysis. The merit of this treatment is the complete cleavage of sialic acid with no loss of O-acetyl group. The released sialic acids were then derivatized and analyzed using a high-performance liquid chromatography with fluorescence (15) or mass spectrometry (MS) detection (16, 17). For example, by labeling with 1,2-diamino-4,5-methylenedioxybenzene (DMB), fmol-level of detection limit for sialic acids was achieved when fluorescence detection was used (18). The DMB derivatization strategy was applied to the study of sialic acids in tissues of mice and rats, by a combination of reverse-phase HPLC with MALDI-TOF MS (17) or coupled with electrospray ionization mass spectrometry (ESI-MS) (16). Although these methods for analysis of sialic acid modifications are well established and provide information on the liberated sialic acids, they do not establish their origin, e.g., glycolipids or glycoproteins. In other words, the linkage information between sialic acids and glycan chains, which is important for understanding of the biological functional role of sialic acid modifications, is lost due to their release from glycoconjugates.

So far, only several reports have focused on the MS analysis of native glycan containing O-acetylated sialic acids (19). The loss of sialic acid moiety has often occurred during the conventional MALDI procedure, which is less than ideal for analysis of sialylated oligosaccharides (20). Therefore, derivatization of carbohydrates prior to the mass spectrometric analysis is widely employed, such as permethylation (21) and amination (22). However, due to the base susceptibility of O-linked acetyl groups in sialic acids the abovementioned derivatization methods could not be applied to the analysis of O-acetyl groups. The loss of sialic acid moiety at native sialylated oligosaccharides can be avoided by the introduction of so-called cool matrix, e.g., 6-aza-2-thiothymine (ATT), or 2,4,6-trihydroxyacetophenone (THAP) in a negative ion detection mode (23). Using this approach, O-acetylation of sialic acids has been detected in two cysteine proteinase inhibitors isolated from a skin of Atlantic salmon (*Salmo salar* L.), in which di-O-acetylated sialic acids containing N-glycans were found to be the major sialylated glycoforms (24). Similar results were obtained for

kininogens, proteins isolated from spotted wolfish and Atlantic cod (25).

In this chapter, we report different approaches to investigate the profiling of N-glycan in fish serum. The O-acetylation of sialic acid of intact N-glycans from fish serum was successfully identified by capillary electrophoresis–mass spectrometry (26). Compared to MALDI-TOF, ESI-MS shows better performance toward the analysis of sialic acids without extra sample preparation and avoid the loss of sialic acid in the analytical procedure.

2. Materials

2.1. Release and Purification of N-Glycans from Fish Serum

1. Fish serum from Juvenile Atlantic salmon (*see* **Note 1**).
2. Peptide-N-glycosidase F (PNGase F) (Sigma-Aldrich, St. Louis, MO).
3. Non-porous graphitic carbon (PGC) cartridge (Alltech Associates, Inc., Deerfield, IL).
4. 20 mM sodium phosphate buffer, pH 7.5, containing 0.2% SDS and 0.1 M dithiothreitol (DTT).
5. 10% NP-40.
6. Acetonitrile (ACN).

2.2. Permethylation

1. Dimethyl sulfoxide (DMSO).
2. Sodium hydroxide (NaOH).
3. Methyl iodide.
4. Chloroform.
5. Methanol.
6. Trifluoroacetic acid (TFA).
7. C_{18} Sep-Pak cartridges (Alltech Associates, Inc., Deerfield, IL).

2.3. Mass Spectrometry Analysis

1. Matrix A: 10 mg/mL of 2,5-Dihydroxybenzoic acid (DHB) in 50% ACN containing 0.1% TFA.
2. Matrix B: 1 mg of 6-Azo-2-thiothymine (ATT) in 1 mL mixture of ethanol/20 mM aqueous ammonium citrate (1/1, v/v).
3. Matrix C: 3 mg/mL of 2,4,6-Trihydroxyacetophenone (THAP) in 50% ACN containing 20 mM aqueous citrate.
4. Diammonium citrate (DAC).
5. Morpholine.

3. Methods

3.1. The Release of N-Glycan in Fish Serum

1. Serum samples (10 μL) were dissolved in 90 μL sodium phosphate (20 mM, pH 7.5) containing 0.2% SDS and 0.1 M DTT and denatured at 100°C for 10 min.
2. Add 10% NP-40 (12 μL) in the solution after cooling.
3. Incubate with PNGase F (5 units) for 24 h at 37°C.
4. Quench the digestion reaction by boiling the solution for 5 min.
5. Wash PGC cartridge with 3.0 mL of 80% ACN, containing 0.1% TFA, followed by 3.0 mL of water.
6. Load the sample solution on PGC cartridge.
7. Wash the cartridge with water (3.0 mL) three times to remove salts.
8. Elute sample by 25% ACN in 0.1% TFA (*see* **Note 2**).
9. Lyophilize the sample.

3.2. Permethylation of N-Glycan

1. Dissolve the dried glycans in a glass tube using 50 μL of DMSO and 100 μL of DMSO–NaOH slurry.
2. Add 50 μL of methyl iodide.
3. Cap the tube.
4. Shake the reaction mixture for 10 min at room temperature.
5. Quench by the addition of water (2 mL).
6. Add chloroform (1 mL) by vortex mixing.
7. Wash the lower organic layer containing the permethylated glycans with water (2 mL ×3).
8. Dry down the organic phase under a stream of nitrogen at room temperature.

3.3. Purification of Permethylated Samples

1. Dissolve the dried permethylated products in 100 μL of 75% methanol.
2. Condition the C_{18} Sep-Pak cartridge sequentially by methanol (2 mL), water (2 mL), ACN (2 mL), and water (2 mL ×2).
3. Load sample on the cartridges.
4. Wash cartridge with water (2 mL ×2).
5. Elute sample sequentially by 15% ACN and 75% CAN.
6. Collect the 75% ACN fraction.
7. Freeze-dry the sample.

3.4. MALDI-TOF Analysis of Permethylation Glycans

1. Dissolve the purified permethylated glycans in 50 μL of 75% methanol.
2. Mix 1 μL of the above solution with 1 μL of freshly made DHB matrix solution in a 100 μL PCR tube.
3. Take 1 μL of sample–matrix mixture directly on the MALDI target.
4. Allow the mixture to dry at room temperature.
5. Acquire the mass spectra by a Voyager-DE STR mass spectrometer in the positive reflectron mode (*see* **Note 1**).

3.5. MALDI-TOF Analysis of Native Glycans

1. Dissolve the native glycans in 50 μL water.
2. Mix 1 μL of the above solution with 1 μL of freshly made ATT or THAP matrix solution in a 100 μL PCR tube.
3. Take 1 μL of sample–matrix mixture directly on the MALDI target.
4. Allow the mixture to dry at room temperature.
5. Acquire the mass spectra by a Voyager-DE STR mass spectrometer in the negative linear mode (*see* **Note 4**).

3.6. CE-ESI-MS Analysis of Native Glycans

1. Dissolve the native glycans in 50 μL water.
2. Setup CE–MS system (*see* **Note 5**).
3. Inject the sample at 500 mbar for 0.2 min.
4. Acquire CE–MS spectra: The MS data were acquired with dwell times of 2.0 ms per step of 0.1 *m/z* unit in full mass scan mode (*see* **Note 6**). In the MS/MS (enhanced product ion scan or EPI) experiments, the scan speed was set to 4000 Da/S, with Q_0 trapping. In MS/MS experiments, the trap fill time was set as "dynamic" and the resolution of Q_1 was set as "unit" (*see* **Note 7**).

4. Notes

1. Fish from each tank were quickly captured and immediately placed in a bucket containing a lethal dose (400 g/L) of tricaine methanesulfonate (Syndel Laboratories Ltd., Vancouver, BC, Canada). Blood samples, collected from the caudal vasculature using a 5-mL heparinized syringe (23 gauge needle), were obtained immediately after the fish were anesthetized. The blood was transferred to a 4-mL Vacutainer tube (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) containing 60 international units of heparin and stored on ice prior to centrifugation. Plasma was

separated by centrifugation (3000×*g* for 15 min) at 4°C, removed and aliquoted into 1.5 mL microcentrifuge tubes. Samples were stored at –80°C until subsequent analysis.

2. High concentration of ACN in elution buffer results in the elution of SDS.
3. Although permethylation is not suitable for characterizing O-linked modification at sialic acids, it can still provide a highly sensitive technique to analyze non-O-modified oligosaccharides, which might not be detected in their native forms by MADLI-MS or ESI-MS. **Figure 18.1** illustrates the MALDI-TOF MS spectrum of the permethylated *N*-glycans from pooled fish plasma samples, revealing the presence of two major ions at *m/z* 2793.8 and 3604.4 corresponding to di-sialylated and tri-sialylated oligosaccharides with chemical compositions of $Hex_5HexNAc_4Neu5Ac_2$ and $Hex_5HexNAc_4Neu5Ac_3$, respectively. Two mono-sialylated oligosaccharides were observed at *m/z* 1982.8 and 2431.5 with chemical compositions of $Hex_4HexNAc_3Neu5Ac_1$ and $Hex_5HexNAc_5Neu5Ac_1$, respectively. The ion at *m/z* 3965.5 corresponds to a tetra-sialylated oligosaccharide with a chemical composition of $Hex_6HexNAc_4Neu5Ac_4$. Ions at *m/z* 2214.9, 2663.4, 3040.9, and 3246.0, corresponding to fucosylated oligosaccharides, were neutral oligosaccharides. Their chemical compositions are $Hex_4HexNAc_4Fuc_2$, $Hex_5HexNAc_5Fuc_2$, $Hex_6HexNAc_5Fuc_3$, and $Hex_7HexNAc_5Fuc_3$, respectively. Two high mannose structures, i.e., $Hex_5HexNAc_2$ (*m/z* 1580.9) and $Hex_6HexNAc_2$ (*m/z* 1784.7), were also observed. These

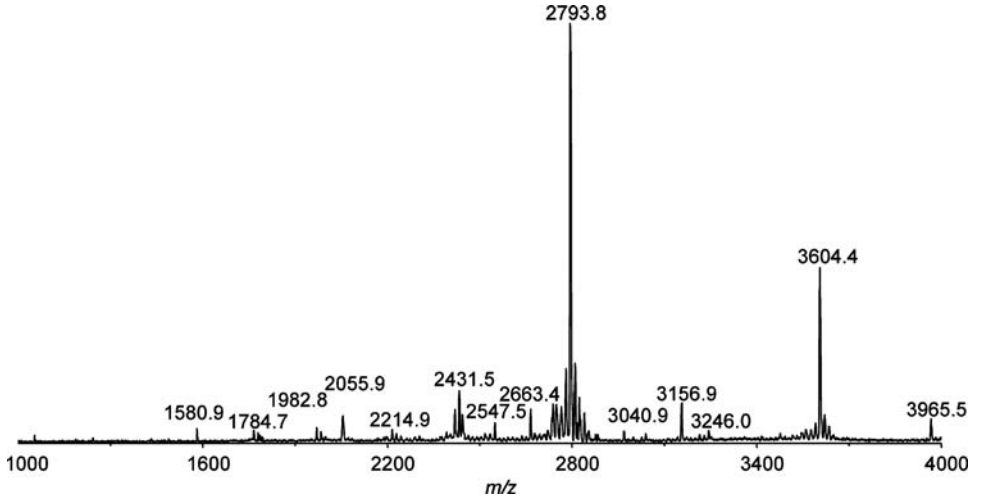


Fig. 18.1. MALDI-TOF analysis of permethylated N-glycans from salmon sera. (Reproduced from Liu et al. (26) with permission from John Wiley & Sons, Ltd., Copyright 2008.)

results demonstrated that fish serum has similar *N*-glycan profile to that of human serum, in which sialylated oligosaccharides are the major components (27), and O-acetylation of sialic acids, which have been found in the several glycoproteins from fish, was not taken into account.

4. As shown in **Fig. 18.2**, a series of de-protonated ions were detected at *m/z* 2221.8, 2263.4, and 2305.8. The corresponding sodium adduct ions were observed at *m/z* 2327.9, 2244.0, and 2286.0, respectively. These ions correspond to bi-antennary oligosaccharides with the addition of different amounts of acetyl groups. In addition, mono-

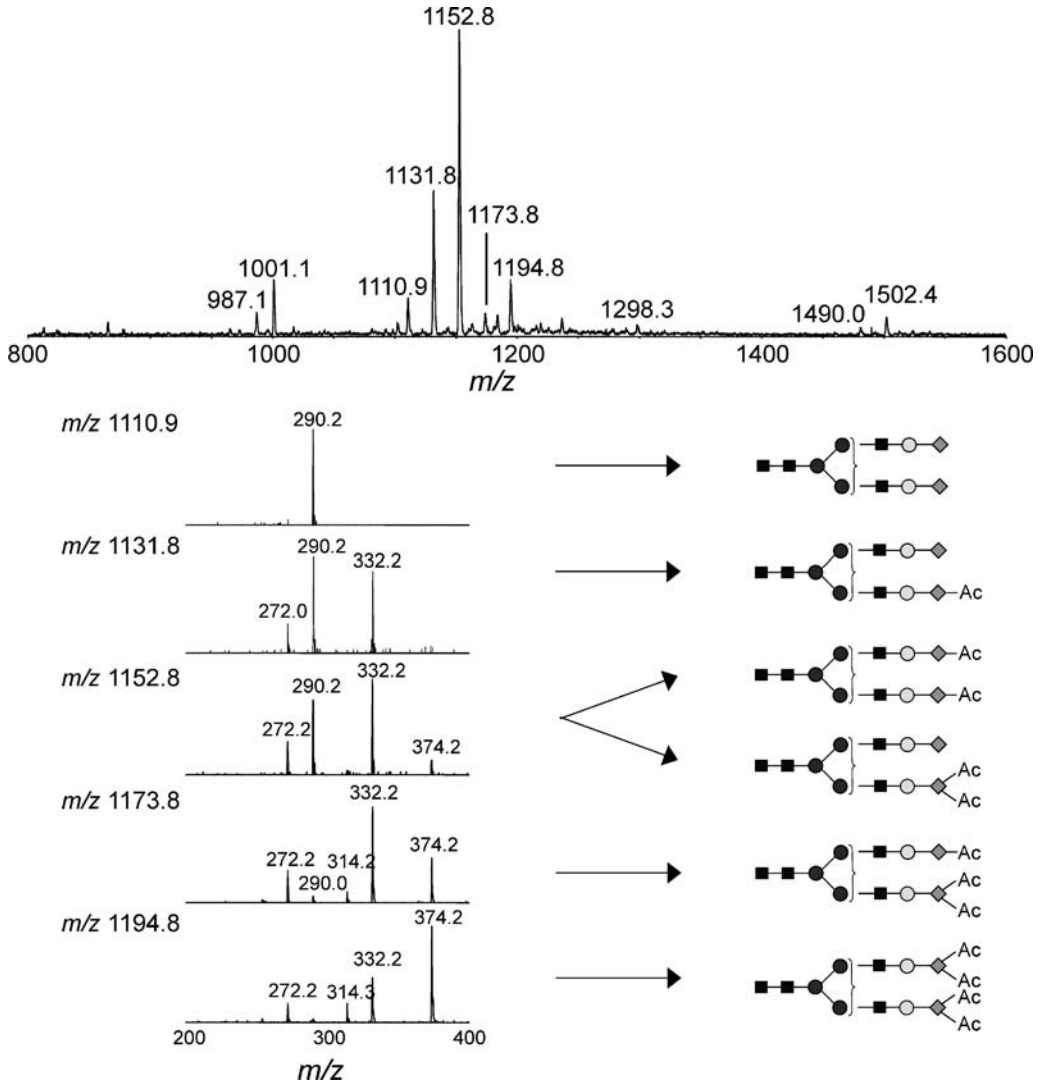


Fig. 18.2. CE–MS and MS/MS analysis of native N-glycans from salmon sera (Reproduced from Liu et al. (26) with permission from John Wiley & Sons, Ltd., Copyright 2008.)

antennary (*m/z* 1931.3, 1972.6) and tri-antennary oligosaccharides (*m/z* 2542) can also be observed.

5. A Prince CE system (Prince Technologies, The Netherlands) is coupled to a 4000 Q-Trap mass spectrometer (Applied Biosystems/MDS Sciex, Canada). A sheath solution (isopropanol/methanol, 2:1) is delivered at a flow rate of 1.0 μL/min. Separations were obtained on about 90 cm length bare fused-silica capillary using 30 mM morpholine in deionized water, pH 9.0. The electrospray ionization voltage (–5 kV) is used for negative ion mode detection.
6. As shown in **Fig. 18.3**, a series of doubly charged ions belonging to bi-antennary oligosaccharides were detected at *m/z* 1110.9, 1131.8, 1152.8, 1173.8, and 1194.8. In addition, tri-antennary structures were observed at *m/z* 987.1 and 1001.1 (triply charged ions), together with corresponding doubly charged ions at *m/z* 1490.0 and 1502.4, respectively. The major ion at *m/z* 1152.8 represents a common bi-antennary complex type oligosaccharide with additional two O-acetyl groups at its terminal sialic acid residues. This observation is not in agreement with previous publication on the skin glycoproteins, where the major glycan is a bi-antennary complex type oligosaccharide with four O-linked acetyl moieties (24, 25).

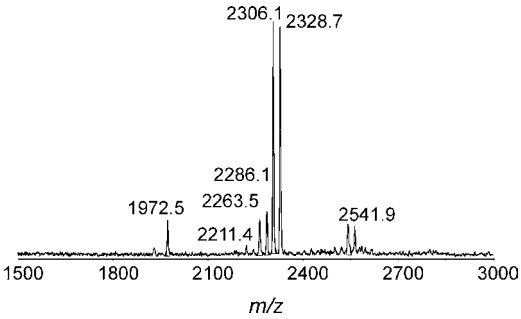


Fig. 18.3. MALDI-TOF analysis of native N-glycans from salmon sera.

7. The expanded MS/MS spectra for each ion of interest are shown in **Fig. 18.3**. The MS/MS spectrum of ion at *m/z* 1110.9 only produced an underivatized Neu5Ac fragment ion at *m/z* 290.2, which indicated the absence of O-acetylated sialic acid in this glycoform. The fragment ion at *m/z* 332.2, corresponding to the Neu5Ac with a mono-O-acetyl group was detected in the MS/MS spectrum of precursor ion at *m/z* 1131.8. The fragment ion at *m/z* 272.0 is probably resulted from the secondary

fragmentation of ion at *m/z* 332.2 with a loss of 60 Da, corresponding to CH_3COOH. A weak peak corresponding to di-O-acetyl Neu5Ac (*m/z* 374.2) was observed in the product ion spectrum of ion at *m/z* 1152.8, indicating that the ion might originated from two isomers. One isomer contains two mono-O-acetylated Neu5Ac residues, while the other has a Neu5Ac residue and a di-O-acetylated Neu5Ac residue. The MS/MS spectra of ions at *m/z* 1173.8 and 1194.8 demonstrated that the di-O-acetylated Neu5Ac was the major sialic acid form in their structures. Furthermore, no Neu5Ac with tri-O-acetyl groups (*m/z* 416.3) were observed in any tandem spectra obtained from this study. In a previous publication, three isomeric mono-O-acetylated Neu5Ac species in salmon glycoproteins were identified as Neu5,9Ac2, Neu5,7Ac2, and Neu5,8Ac2, and the di-O-acetylated Neu5Ac species carrying O-acetyl group at 7-OH, 8-OH, and 9-OH of Neu5Ac as Neu5,8,9Ac3, Neu5,7,9Ac3, and Neu5,7,8Ac3, were detected (24). It is noteworthy that in the current study, the locations of each O-acetyl group on sialic acids were not determined. Nevertheless, tandem mass spectrometry was still able to provide the direct proof of the modification of sialic acids in the specific sialylated oligosaccharides.

Acknowledgments

We would like to thank Drs. Jianjun Li, Eleonora Altman, Stewart Johnson, and Laura Brown for their encouragement and support of these studies. This work was funded by the NRC Genomics and Health Initiative.

References

1. Turnbull, J. E. and Field, R. A., (2007) Emerging glycomics technologies. *Nat. Chem. Biol.* **3**, 74–77.
2. Klein, A. and Roussel, P., (1998) O-acetylation of sialic acids. *Biochimie* **80**, 49–57.
3. Schauer, R. (2000) Achievements and challenges of sialic acid research. *Glycoconj. J.* **17**, 485–499.
4. Varki, A. (1992) Diversity in the sialic acids. *Glycobiology* **3**, 25–40.
5. Campbell, F., Appleton, M. A., Fuller, C. E., Greeff, M. P., Hallgrimsson, J., Katoh, R., Ng, O. L., Satir, A., Williams, G. T., and Williams, E. D. (1994) Racial variation in the O-acetylation phenotype of human colonic mucosa. *J. Pathol.* **174**, 169–174.
6. Corfield, A. P., Wagner, S. A., Paraskeva, C., Clamp, J. R., Durdey, P., Reuter, G., and Schauer, R., (1992) Loss of sialic acid O-acetylation in human colorectal cancer cells. *Biochem. Soc. Trans.* **20**, 94S.
7. Corfield, A. P., Myerscough, N., Warren, B. F., Durdey, P., Paraskeva, C., and Schauer, R. (1999) Reduction of sialic acid O-acetylation in human colonic mucins in the adenoma-carcinoma sequence. *Glycoconj. J.* **16**, 307–317.
8. Ghosh, S., Bandyopadhyay, S., Mallick, A., Pal, S., Vlasak, R., Bhattacharya, D.

K., and Mandal, C. (2005) Interferon gamma promotes survival of lymphoblasts overexpressing 9-O-acetylated sialoglycoconjugates in childhood acute lymphoblastic leukaemia (ALL). *J. Cell Biochem.* **95**, 206–216.
9. Ghosh, S., Bandyopadhyay, S., Mukherjee, K., Mallick, A., Pal, S., Mandal, C., Bhattacharya, D. K., and Mandal, C. (2007) O-acetylation of sialic acids is required for the survival of lymphoblasts in childhood acute lymphoblastic leukemia (ALL). *Glycoconj. J.* **24**, 17–24.
10. Shen, Y., Tiralongo, J., Iwersen, M., Sipos, B., Kalthoff, H., and Schauer, R. (2002) Characterization of the sialate-7(9)-O-acetyltransferase from the microsomes of human colonic mucosa. *Biol. Chem.* **383**, 307–317.
11. Shen, Y., Tiralongo, J., Kohla, G., and Schauer, R. (2004) Regulation of sialic acid O-acetylation in human colon mucosa. *Biol. Chem.* **385**, 145–152.
12. Shen, Y., Kohla, G., Lrhorfi, A. L., Sipos, B., Kalthoff, H., Gerwig, G. J., Kamerling, J. P., Schauer, R., and Tiralongo, J., (2004) O-acetylation and de-O-acetylation of sialic acids in human colorectal carcinoma. *Eur. J. Biochem.* **271**, 281–290.
13. Shi, W. X., Chammas, R., and Varki, A. (1996) Regulation of sialic acid 9-O-acetylation during the growth and differentiation of murine erythroleukemia cells. *J. Biol. Chem.* **271**, 31517–31525.
14. Sjoberg, E. R., Powell, L. D., Klein, A., and Varki, A. (1994) Natural ligands of the B cell adhesion molecule CD22 beta can be masked by 9-O-acetylation of sialic acids. *J. Cell Biol.* **126**, 549–562.
15. Butor, C., Diaz, S., and Varki, A. (1993) High level O-acetylation of sialic acids on N-linked oligosaccharides of rat liver membranes. Differential subcellular distribution of 7- and 9-O-acetyl groups and of enzymes involved in their regulation. *J. Biol. Chem.* **268**, 10197–10206.
16. Morimoto, N., Nakano, M., Kinoshita, M., Kawabata, A., Morita, M., Oda, Y., Kuroda, R., and Kakehi, K., (2001) Specific distribution of sialic acids in animal tissues as examined by LC-ESI-MS after derivatization with 1,2-diamino-4,5-methylenedioxybenzene. *Anal. Chem.* **73**, 5422–5428.
17. Stehling, P., Gohlke, M., Fitzner, R., and Reutter, W. (1998) Rapid analysis of O-acetylated neuraminic acids by matrix assisted laser desorption/ionization time-of-flight mass spectrometry. *Glycoconj. J.* **15**, 339–344.
18. Hara, S., Yamaguchi, M., Takemori, Y., Furuhata, K., Ogura, H., and Nakamura, M. (1989) Determination of mono-O-acetylated N-acetylneuraminic acids in human and rat sera by fluorometric high-performance liquid chromatography. *Anal. Biochem.* **179**, 162–166.
19. Higa, H. H., Manzi, A. and Varki, A., (1989) O-acetylation and de-O-acetylation of sialic acids. Purification, characterization, and properties of a glycosylated rat liver esterase specific for 9-O-acetylated sialic acids. *J. Biol. Chem.* **264**, 19435–19442.
20. Harvey, D. J., (2006) Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update covering the period 1999-2000. *Mass Spectrom. Rev.* **25**, 595–662.
21. Liu, X., Li, X., Chan, K., Zou, W., Pribil, P., Li, X. F., Sawyer, M. B., and Li, J. (2007) "One-pot" methylation in glycomics application: esterification of sialic acids and permanent charge construction. *Anal. Chem.* **79**, 3894–3900.
22. Sekiya, S., Wada, Y. and Tanaka, K., (2005) Derivatization for stabilizing sialic acids in MALDI-MS. *Anal. Chem.* **77**, 4962–4968.
23. Sagi, D., Kienz, P., Denecke, J., Marquardt, T., and Peter-Katalinic, J. (2005) Glycoproteomics of N-glycosylation by in-gel deglycosylation and matrix-assisted laser desorption/ionisation-time of flight mass spectrometry mapping: application to congenital disorders of glycosylation. *Proteomics.* **5**, 2689–2701.
24. Ylonen, A., Kalkkinen, N., Saarinen, J., Bogwald, J., and Helin, J. (2001) Glycosylation analysis of two cysteine proteinase inhibitors from Atlantic salmon skin: di-O-acetylated sialic acids are the major sialic acid species on N-glycans. *Glycobiology* **11**, 523–531.
25. Ylonen, A., Helin, J., Bogwald, J., Jaakola, A., Rinne, A., and Kalkkinen, N., (2002) Purification and characterization of novel kininogens from spotted wolffish and Atlantic cod. *Eur. J. Biochem.* **269**, 2639–2646.
26. Liu, X., Afonso, L., Altman, E., Johnson, S., Brown, L., and Li, J. (2008) O-acetylation of sialic acids in N-glycans of Atlantic salmon (*Salmo salar*) serum is altered by handling stress. *Proteomics* **8**, 2849–2857.
27. Kang, P., Mechref, Y., Kyselova, Z., Goetz, J. A., and Novotny, M. V. (2007) Comparative glycomic mapping through quantitative permethylation and stable-isotope labeling. *Anal. Chem.* **79**, 6064–6073.

Chapter 19

Bioinformatics in Glycomics: Glycan Characterization with Mass Spectrometric Data Using SimGlycan™

Arun Apte and Ningombam Sanjib Meitei

Abstract

Mass spectrometry (MS), with its low sample requirement and high sensitivity, has been the predominantly used methodology for characterization and elucidation of glycan structures. However, manual interpretation of MS data is complex and tedious due to large number of product ions observed and also due to the variation in their m/z values under various experimental conditions. We present an automated tool, SimGlycan™, for this purpose, which accepts raw/standard MS data files as input and characterizes the associated glycan structure with high accuracy using database searching and scoring techniques. Not only does it predict the glycan structure using an MS/MS database searching technique, but it also facilitates predicting novel glycans by drawing a glycan and mapping it onto an experimental spectrum to check the degree of proximity between the theoretical and the experimental glycans. It serves as a platform for developing advanced tools that may be used for glycopeptide identification using MS data and 3D structural analysis of glycans with a few improvements in the existing features.

Key words: Glycan, bioinformatics, mass spectrometric data , SimGlycan™, MS/MS database searching.

1. Introduction

Characterization and structural elucidation of glycans are important for encoding information for specific molecular recognition and determination of protein folding, stability, and pharmacokinetics. However, the structural complexity of glycans, far greater than that of proteins and nucleic acids, provides for a truly challenging analytical problem to bio-analysts and biochemists. Mass spectrometry (MS) in combination with modern separation methodologies has become one of the most powerful and versa-

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8_19,

tile techniques for the structural analysis of glycans. Other current structural techniques cannot match MS for the range of structural problems that can be addressed and the complexity of samples that can be analyzed successfully (1). Many researchers have described the theoretical basis for the elucidation of unknown glycan structures using MS data (2–6). MS does have some drawbacks including the difficulty in manual interpretation of the large volumes of data it generates. This analysis requires information about the sugar sequences and anomeric linkages between the monosaccharide units and their fragmentation patterns under different experimental conditions. The need for developing bioinformatics tools to enable the comparison of MS data between unknown and known structures currently limits the interpretation of MS results (7). Many attempts have been made to develop automated tools for interpreting glycan MS data. STAT (8), StrOligo (9), GlycoSearchMS (10), and Cartoonist (11) are some of the tools for this purpose. A good review of the methods and tools developed before the year 2005 and the continuing challenge had been discussed in the reference article (12). We compile a list of tools in **Table 19.1** which are used for glycan analysis with specific classification of their usage.

Table 19.1
The list of informatics modules that are required for glycan characterization using mass spectrometry, available bioinformatics tools, and their respective references

S. No	Categories	Software tools	References
1.	Structural Glycan characterization tool using mass spectrometric data	STAT (de novo tool) Stroligo GlycoSearchMS OSCAR (de novo tool) Glych Glycosid iQ SimGlycan™	(8) (9) (10) (13) (14) (15)
2.	Glycan fragmentation tools	GlycoFragment Glycofrag SimGlycan™	(10) (15)
3.	Structure drawing and search tool	GlycanBuilder KCaM SimGlycan™	(16) (17)
4.	Spectra Annotation tool	Cartoonist Peptoonist Glyco-Peakfinder SimGlycan™	(11) (18) (19)
5.	Glycopeptide composition analysis tool (for known peptide sequence)	Glycomod Glycopep DB Glycomaster SimGlycan™	(20) (21) (22)

SimGlycan™ is a comprehensive desktop tool designed to address the challenges of glycan structure prediction from glycan mass spectrometric data. It can be used for

(i) Glycan structure prediction using release glycan mass spectrometric data.

(ii) Glycan structure prediction using glycopeptide mass spectrometric data (mass of the peptide or the peptide chain should be provided).

(iii) Novel glycan prediction by drawing a new glycan or editing an existing glycan and then comparing the theoretical fragments against the experimental data.

(iv) Annotation of the experimental spectrum with match fragment ions.

In the following sections, we discuss various aspects of glycan analysis using SimGlycan™ in detail.

2. Accurate Glycan Structure Prediction Using SimGlycan™

SimGlycan™ predicts the structure of a glycan from the MS/MS data acquired by mass spectrometry, facilitating the study of glycosylation and post-translational modification studies. SimGlycan™ accepts the experimental MS/MS data generated by a mass spectrometer, matches them with its own database of theoretical fragmentation of over 9000 glycans, and generates a list of probable glycan structures. Each structure is scored to reflect how closely it matches experimental data. Other biological information for the probable glycan structures such as the glycan class, reaction, pathway, and enzyme are also made available for easy reference. A pictorial depiction of how SimGlycan™ works is available in **Figs. 19.1 and 19.2**. **Figure 19.1** shows the steps involved in MS/MS data collection and uploading it onto SimGlycan™ along with initial search parameters. Image 1 in **Fig. 19.1** shows the precursor region of glycan at m/z 1087.5 on MS1 spectrum. Image 2 shows the MS/MS spectrum triggered for this m/z value. Image 3 shows the extracted peak list – m/z and intensity values – from the MS/MS spectrum and the experimental conditions that are used as initial search predicates by SimGlycan™.

Figure 19.1b depicts the SimGlycan™ mechanism for glycan prediction using database searching methods and scoring algorithms. Images in the **Fig. 19.2** are numbered in continuation to those of **Fig. 19.1**. Image 4 shows the retrieval of glycan structures which satisfy the search parameters from the

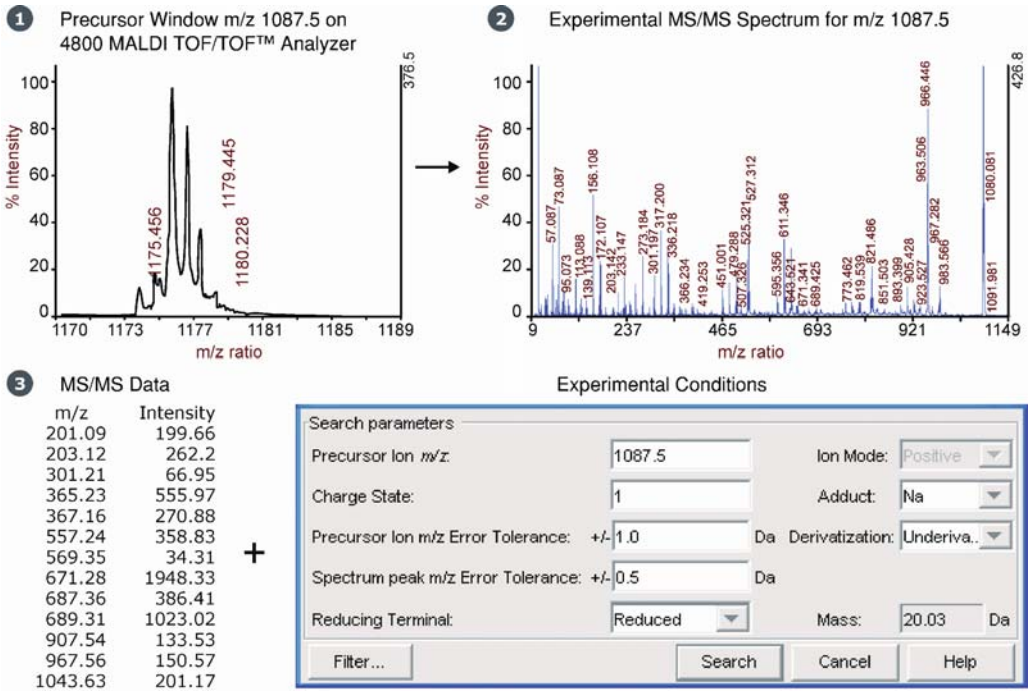


Fig. 19.1. Glycan Prediction Schema for SimGlycan™: MS/MS data collection mechanism and uploading it into SimGlycan™ along with search parameters.

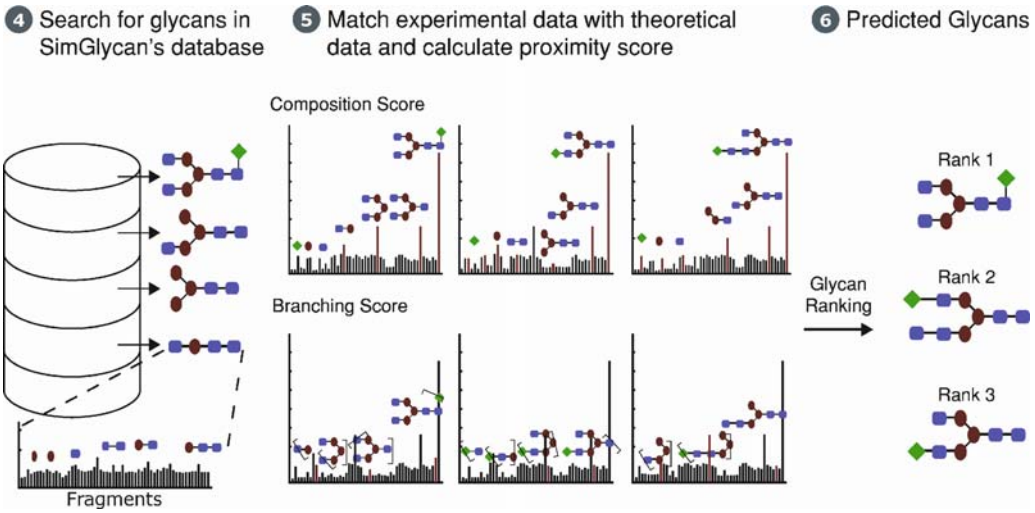


Fig. 19.2. Glycan Prediction Schema for SimGlycan™: MS/MS database searching, scoring and ranking of the probable glycans using SimGlycan™.

SimGlycan™ database. The calculated *m/z* values for theoretical fragments of these glycans are matched against the experimental data. Images 5 and 6 show the mechanism on how SimGlycan™ verifies the monosaccharides composition and branching pattern

of the unknown experimental glycan using the glycosidic, internal and cross-ring cleavages of theoretical glycans using propriety scoring algorithm. The theoretical glycan structure closest to the experimental glycan is ranked 1.

We further explain the input parameters, result output, and other facilities provided by SimGlycan™.

2.1. Input Data and Supported File Formats

SimGlycan™ accepts experimental *m/z* and intensity values (MS/MS data) of glycans. Manual entry of MS/MS data or uploading it as an input file is allowed. The standard file formats listed in **Table 19.2** are accepted as input MS/MS data:

Table 19.2
List of standard file formats accepted as input MS/MS data by SimGlycan™

File extension	Description
.txt	Text files
.xls	Microsoft Excel files®
.mzXML	Standard MS output file format
.mzData	Standard MS output file format

In addition, SimGlycan™ also accepts the native files from various MS instruments manufactured by different companies. **Table 19.3** presents the list of native file formats from various MS instrumentations that can be uploaded as input without converting it into standard data file formats.

Table 19.3
List of MS instruments native file formats, instruments, and corresponding manufacturer, accepted as input MS/MS data by SimGlycan™

File format	Instrument	Manufacturer
.wiff	QSTAR® and QTRAP®	Applied Biosystems/MDS Analytical Technologies
.t2d	4700 and 4800 MALDI TOFTOF™	Applied Biosystems
.baf	ESQUIRE™ instrument family	Bruker Daltonics
.fid	autoFlex™ and ultraFlex® instrument families.	Bruker Daltonics
.yep	APEX®, micrOTOF™, micOTOF-Q	Bruker Daltonics

Apart from the MS/MS data, the following parameters need to be specified to find theoretical glycan structures. These

parameters will be used as search predicates while matching the theoretical data from the database against the experimental spectrum.

1. Precursor ion *m/z* (as indicated in **Fig. 19.1**).
2. Charge state of the precursor *m/z* (specify 1 if it is singly charged, 2, if doubly and so on).
3. Precursor ion *m/z* error tolerance (e.g., 1 Da).
4. Spectrum peak *m/z* tolerance (e.g., 0.5 Da).
5. Glycan derivatization (specify whether the glycan is permethylated or underivatized).
6. Reducing terminal modification (select a reducing end modification or enter the mass if not in the list).
7. Ion mode (positive or negative).
8. Adduct (H, Na).
9. Advanced filter (optional, user can specify the Biological source of the glycan, type of linkage like N-link, tentative branching nature of the glycan like linear, biantennary, etc., to narrow down the search).

2.2. Project Management in SimGlycan™

SimGlycan™ provides comprehensive project management, associating results with input profile and search parameters. Each project can include multiple MS profiles. The name of the MS/MS data file is retained as the default name of a profile. For example, if the name of the MS/MS data file is test_heparin.txt then the name of the profile in the SimGlycan™ project will be "test_heparin." However, the profile name can be renamed or deleted. When mzXML or mzData files are uploaded as input files, all the scans with MS level 2 are uploaded automatically. The profiles created in the project are named after their respective scan numbers which are in turn retrieved from the mzData or mzXML files. SimGlycan™ enables projects to be classified on the basis of the source, the laboratory or the research goal which may be handy when conducting large-scale projects.

SimGlycan™ highlights the *m/z* values which match those of theoretical fragments. Upon their selection, it plots, annotates, and exports an MS profile in .png format for a laboratory notebook. While annotating the spectrum, the spectrum view can be switched between peaks with subsequent loss of monosaccharide units and peaks labeled with fragment names using fragment nomenclature (22) using a simple interface. The annotated spectrum can be zoomed into specific plot regions by specifying the *m/z* range to display specific plot sections. SimGlycan™ enables a user to assign their own rank to predicted structures and add comments and search fragments by their mass, name, and *m/z* values.

2.3. Generate Report

SimGlycan™ can create a report for the laboratory notebook, publication or sharing information with a third party. The report includes glycan characteristics such as ID, sequence, name, composition, mass, precursor ion *m/z*, class, reaction, pathway along with a graphical view of the glycan, and its possible theoretical fragments along with corresponding mass and *m/z* values. A user can either export the information to Microsoft Excel or to an HTML file. The images of the glycan structure, fragments, and plotted spectrum can be exported as JPEG/PNG files.

2.4. Identification of Glycopeptide when Peptide Mass or Sequence Is Known

Users can use SimGlycan™ for glycopeptide MS/MS data to predict the attached glycan structure if the mass of the peptide or the peptide sequence is known. The input parameters are the same as explained above except for the reducing terminal modification. The amino acid modifications such as cysteine modification, methionine modification, and the position of the modified amino acids can be specified. The mass of the peptide sequence along with specified modifications would be incorporated automatically while comparing the experimental MS/MS data against the theoretical data for glycan structures.

2.5. Identification of Novel Glycan Structures Using SimGlycan™

Most of the software tools reported in the literature use databases of known glycan structures so that the unknown glycan is identified by comparing the experimental tandem MS data against the theoretical database. Thus, the success of database approaches depends heavily on the completeness of the quality of the database. Another methodology for glycan identification is the use of de novo approaches, seeking to determine glycan structures successively, monosaccharide by monosaccharide, from the fragments observed. The de novo method has the advantage of reporting novel glycans. However, its accuracy deteriorates when larger masses are used in the search. In addition, the successful identification of glycans with non-standard cores such as O-linked glycans is a difficult task as observed in the case of OSCAR (13). No information regarding the predicted glycan would be available in this case.

SimGlycan™ predicts unknown glycans using MS/MS database searching techniques. However, upon finding that the percentage match between the theoretical and experimental product ions is very low – a user can select any other glycan from the list of probables or draw a new glycan altogether. A monosaccharide or a non-glycan compound such as HSO_3, ETN, etc. in the glycan can be added or deleted. Furthermore, branching points and anomeric linkages can also be modified. At each step, the percentage match of the theoretical and experimental product ions is displayed, enabling a user to see whether his modification brings the theoretical glycan closer to the experimental glycan.

This feature should greatly assist in resolving glycan structures, especially when data for glycans of interest are not available in the database. This methodology facilitates the identification of novel glycan structures regardless of the glycan mass.

3. Glycan Characterization Using Real Sample Data

3.1. Materials and Methods

DATA ACQUISITION: 4800 MALDI TOF/TOF™ Analyzer.

Carbohydrate samples in different concentrations were prepared using DHB as a matrix. Samples were analyzed using the 4800 MALDI TOF/TOF™ Analyzer in reflector-ion mode or MS/MS-ion mode. DHB matrix 20 mg/mL in deionized water. Dried droplet preparation with sugar solutions at concentrations of about 10^{-4} to 10^{-5} mol/L. MS/MS (positive ion mode of sodiated ion species) on 4800 MALDI TOF/TOF™ Analyzer using 2 kV collision energy and Argon as collision gas at a pressure of 4×10^{-6} Torr (operating pressure of ion source at 1×10^{-7} Torr).

3.2. SimGlycan™ Settings

SimGlycan™ accepts experimental *m/z* and intensity values (MS/MS data) of glycans. The following settings are used for the data: precursor ion *m/z* as indicated in **Fig. 19.1**, precursor ion tolerance of 1 Da, spectrum peak *m/z* tolerance of 0.5 Da, no glycan derivatization, reduced terminal reducing, ion mode (MALDI positive), and the adduct as Na (sodiated carbohydrates as $[M+Na]^+$).

3.3. SimGlycan™ Results for m/z 1087.5: ID for (Hex) 4 (Pent) 3 Sugar

A list of glycans with their corresponding scores are displayed in the Search Results Pane, in decreasing order of their probability. **Figure 19.3** shows the output glycans for the MS/MS data. The glycan predicted by SimGlycan™ was accurate. Apart from the structure, other important information such as the glycan mass, monosaccharide composition, glycan class, other database link where you can find further information regarding the glycan and most importantly, the percentage match of the glycan fragments against the experimental product ions, is displayed. If the percentage match is not satisfactory, then the user can opt for finding a novel glycan structure using the "Draw and edit" functionality provided, as explained in **Section 2.5**.

On the MS/MS annotation page, SimGlycan™ highlights the *m/z* values which match those of theoretical fragments. The matched theoretical fragments are annotated. Fragment nomenclature is based on the standard Domon and Costello rules (22) along with other rules (23). Upon their selection, it plots,

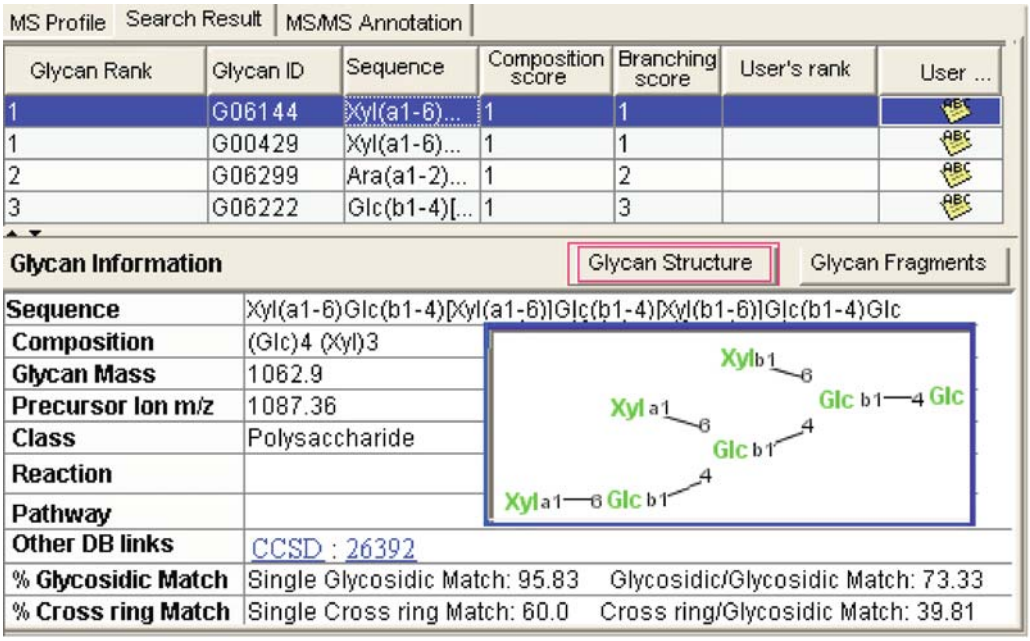


Fig. 19.3. SimGlycan™ MS/MS database search result page showing the list of probable glycan structures with rank 1 given to the closest structures.

annotates, and exports an MS profile in .png format for your laboratory notebook. The MS/MS annotation result page is shown in **Fig. 19.4**.

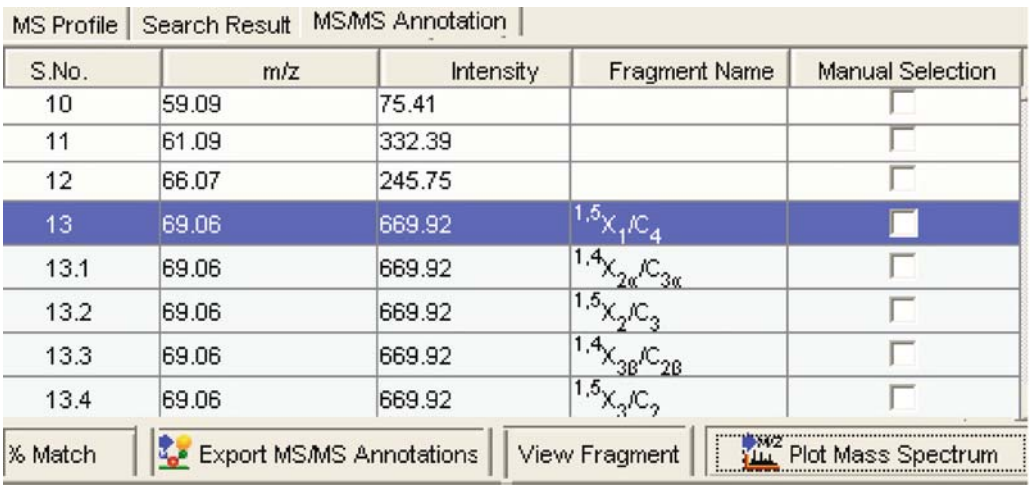


Fig. 19.4. SimGlycan™ result page showing annotated MS/MS data with matched glycan fragments.

The annotated spectrum is shown in **Fig. 19.5**. It displays the subsequent loss of monosaccharide units from the precursor ion. The loss of branches such as α and β is also displayed.

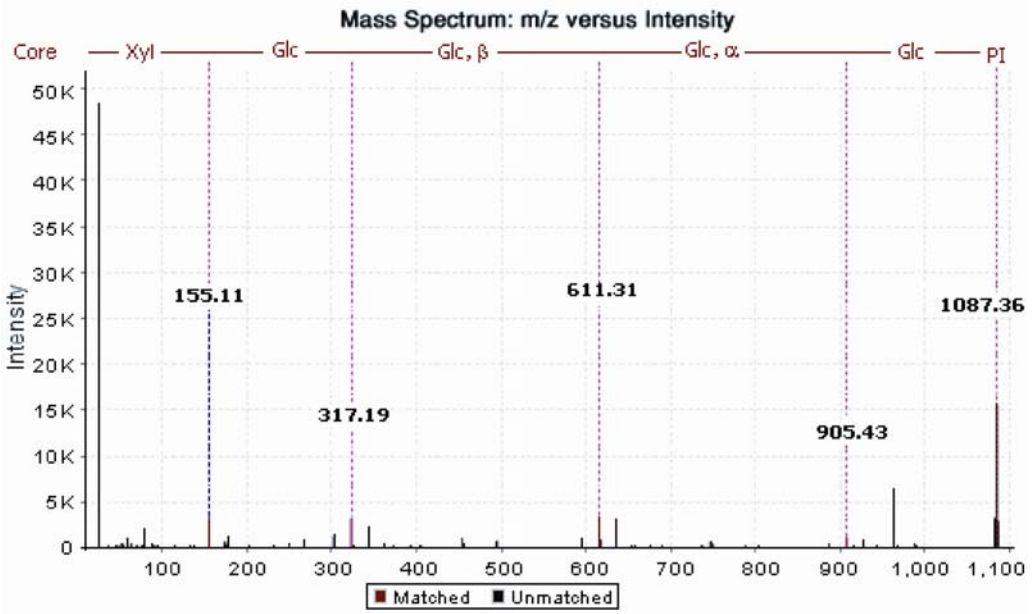


Fig. 19.5. SimGlycan^{TM} result page showing annotated MS/MS spectrum with subsequent loss of monosaccharides of the predicted glycan structure (**Fig. 19.3**).

3.4. Future Research and Updates in SimGlycan^{TM}

To address the broadening dimensions of glycomics and glycoproteomics research in general, SimGlycan^{TM} is being continually enhanced. In this section, we discuss two important enhancements: (a) complete glycopeptide identification using MS and (b) 3D structural identification of glycans, that will later be supported by SimGlycan^{TM}.

3.4.1. Complete Glycopeptide Identification Using MS

With the advances in technology, researchers have started taking a keen interest in developing bioinformatics tools with the ability to elucidate the glycan structure and attached amino acid composition directly using LC–MS/MS or multi-stage MS data of glycopeptides. These tools speed up the lengthy process of glycoprotein characterization. In the most common methodology, the proteolytically digested glycoprotein is subjected to liquid chromatography for separating peptides from glycopeptides. The separated glycopeptides are then subjected to mass spectrometry and identification of the molecule is done using the generated data. No separate procedure to release the glycans from intake amino acids such as using PNGase F is required before subjecting the sample to mass spectrometry analysis. Peterman and Mulholland (24) describe a novel approach for identification and characterization of glycoproteins using a hybrid Linear Ion Trap/FT-ICR Mass Spectrometer. Such approaches generate large data sets from multi-stage mass spectra. There is an urgent need for developing high-throughput bioinformatics tools that can handle such

large amounts of data and elucidate the glycopeptides directly. In future versions of SimGlycan^{TM}, it will allow users to upload mass spectrometric data of glycopeptides along with other experimental conditions such as the enzyme used to digest the glycoprotein, amino acid modifications (if any), etc. The uploaded data will be matched against theoretical database to determine the nearest matched glycopeptides. All the available features in the current version of SimGlycan^{TM} for structural elucidation of release glycans will be extended to incorporate the technical needs of glycopeptide elucidation. The following features are under active development:

(i) Identification of monosaccharide composition.
(ii) Identification of the amino acid composition.
(iii) Identification of non-carbohydrate compounds such as HSO_3, ETN, etc.
(iv) Identification of anomeric linkages.
(v) Identification of branching point.

3.4.2. 3D Structural Identification of Glycans

Three-dimensional structural identification of glycans is in high demand by structural biologists. Perhaps, it provides one of the most meaningful analyses in understanding the functions of glycoproteins. Yet, it is still one of the most challenging tasks for researchers. Typically, 3D structures of glycans are identified using X-Ray crystallography. However, the primary composition of the glycan should be verified using other techniques such as NMR, HPLC, MS, etc. It includes identification of basic monosaccharide composition, the anomeric linkages which in turn identify the branching pattern of glycans and other compound such as HSO_3, ETN, etc. Mass spectrometry in combination with modern separation methodologies has become one of the most powerful and versatile techniques for primary structural analysis of glycans. SimGlycan^{TM} facilitates the verification of primary composition of glycan mass spectrometric data. Using this functionality as a platform, SimGlycan^{TM} is being further developed to display the 3D structures of glycans.

4. Conclusion

The complexity of glycan structural characterization coupled with the lack of bioinformatics tools hinders the study of common glycan-based research such as cell-to-cell interaction, protein folding, stability, and drug discovery. SimGlycan^{TM}, with its ability to accept various data file formats, native mass spectrometric data or

standard data formats as input, well-organized project management systems, strong database support, accurate glycan identification, various options for annotating the spectrum, portability of outputted results/reports into various file formats, etc., addresses various needs of researchers pertaining to glycomics research. Results including glycan fragments, structure, sequence, composition, glycan mass, class, reaction, pathway, enzyme, and other database links for every probable glycan structure enable researchers to look beyond just predicting the glycan structure present in the database. It also serves as a strong platform for developing further bioinformatics tools that will support glycopeptide identification using mass spectrometry data, 3D structural analysis of glycans, etc.

4.1. Availability and Licensing

Product Home Page: SimGlycan™ can be downloaded from http://www.premierbiosoft.com/glycan

4.2. Systems Requirement for SimGlycan™

- Operating system(s): XP, Vista (32 and 64 bit) and other windows flavors.
- Programming language: Java, Swing API.
- Other requirements: JVM 1.4 or higher. Minimum CPU capacity with 800 MHz, 512 MB RAM, 200 MB free space hard disk drive and 800×600 screen resolution. Better performance of the software can be achieved using higher CPU, RAM, and hard drive access speed.

References

1. Morelle, W. and Michalski, J. C. (2005) The mass spectrometric analysis of glycoproteins and their glycan structures. *Curr. Anal. Chem.* **1**, 29–57.
2. Konig, S. and Leary, J. A. (1998) Evidence for linkage position determination in cobalt coordinated pentasaccharides using ion trap mass spectrometry. *J. Am. Soc. Mass Spectrom.* **9**, 1125–1134.
3. Sheeley, D. M. and Reinhold, V. N. (1998) Structural characterization of carbohydrate sequence, linkage, and branching in a quadrupole ion trap mass spectrometer: neutral oligosaccharides and N-linked glycans. *Anal. Chem.* **70**, 3053–3059.
4. Viseux, N., de Hoffmann, E., and Domon, B. (1998) Structural assignment of permethylated oligosaccharide subunits using sequential tandem mass spectrometry. *Anal. Chem.* **70**, 4951–4959.
5. Weiskopf, A. S, Vouros, P., and Harvey, D. J (1998) Electrospray ionization-ion trap mass spectrometry for structural analysis of complex N-linked glycoprotein oligosaccharides. *Anal. Chem.* **70**, 4441–4447.
6. Tseng, K., Hedrick, J. L, and Lebrilla, C. B (1999) Catalog-library approach for the rapid and sensitive structural elucidation of oligosaccharides. *Anal. Chem.* **71**, 3747–3754.
7. Zaia, J. (2004) Mass spectrometry of oligosaccharides. *Mass Spectrom. Rev.* **23**, 161–227.
8. Gaucher, S. P., Morrow, J., and Leary, J. A. (2000) STAT: A saccharide topology analysis tool used in combination with tandem mass spectrometry. *Anal. Chem.* **72**, 2331–2336.
9. Ethier, M, Saba, JA, Ens, W, Standing, KG, and Perreault, H. (2002) Automated structural assignment of derivatized complex N-linked oligosaccharides from tandem mass spectra. *Rapid Commun. Mass Spectrom.* **16**, 1743–1754.
10. Karl, L. K., Lieth, V.D, and Claus-W. (2004) GlycoFragment and GlycoSearchMS: web tools to support the interpretation of

mass spectra of complex carbohydrates. *Nucl. Acids Res.* **32** (1), W261–W266(1).
11. Goldberg, D, Sutton-Smith, M, Paulson, J, and Dell, A. (2005) Automatic annotation of matrix-assisted laser desorption/ionization of N-glycan spectra. *Proteomics* **5**(4):865–875.
12. Lieth, V. D, Lutteke, T., and Frank, M. (2006) The role of informatics in glycobiology research with special emphasis on automatic interpretation of MS spectra. *Biochim. Biophys. Acta.* **1760**(4), 568–577.
13. Lapadula, A. J, Hatcher, P. J, Hanneman, A. J, Ashline, D. J, Zhang, H., and Reinhold, V. N (2005) Congruent strategies for carbohydrate sequencing: 3. OSCAR: An algorithm for assigning oligosaccharide topology from MS(n) data. *Anal. Chem.* 77, 6271–6279.
14. Tang, H, Mechref, Y, and Novotny, MV (2005) Automated interpretation of MS/MS spectra of oligosaccharides. *Bioinformatics* **21**(Suppl 1), i431–i439.
15. Joshi, H. J, Harrison, M. J, Schulz, B. L, Cooper, C. A, Packer, N. H, and Karlsson, N. G (2004) Development of a mass fingerprinting tool for automated interpretation of oligosaccharide fragmentation data. *Proteomics* **4**, 1650–1664.
16. Ceroni, A., Dell, A., and Haslam, S. M. (2007) The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. *Source Code Biol. Med.* **2:3** doi:10.1186/1751-0473-2-3.
17. Aoki, K. F, Yamaguchi, A., Ueda, N, Akutsu, T., Mamitsuka, H, Goto, S., and Kanehisa, M. (2004) KCaM (KEGG Carbohydrate Matcher): a software tool for analyzing the structures of carbohydrate sugar chains. *Nucl. Acids Res.* **32**(Web Server issue), W267–W272, DOI: 10.1093/nar/gkh473.
18. Goldberg, D, Bern, M, Parry, S, Sutton-Smith, M, Panico, M, Morris, HR, and Dell, A. (2007) Automated N-glycopeptide identification using a combination of single- and tandem-MS. *J. Proteome. Res.* **6**(10), 3995–4005.
19. Maass, K., Ranzinger, R., Geyer, H., Lieth, V., D., C-W, and Geyer, R. (2007) "Glyco-Peakfinder" – de novo composition analysis of glycoconjugates. *Proteomics* **7**, 4435–4444.
20. Cooper, C. A, Gasteiger, E., and Packer, N. H (2001) GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* **1**, 340–349.
21. Go, E. P., Rebecchi, K. R., Dalpathado, D. S., Bandu, M. L., Zhang, Y., and Desaire, H. (2007) GlycoPep DB: a tool for glycopeptide analysis using a "Smart Search". *Anal Chem.* **79**(4), 1708–1713.
22. Domon, B., and Costello, C. E (1998) A systematic nomenclature for carbohydrate fragmentation in FAB-MS/MS spectra of Glycoconjugates. *Glycoconj. J.* **5**, 397–409.
23. Cooper, C. A, Wilkins, M. R, Williams, K. L, and Packer, N. H(1999) BOLD – a biological O-linked glycan database. *Electrophoresis* **20**, 3589–3598.
24. Peterman, C. M. and Mulholland, J. J (2006) A novel approach for identification and characterization of glycoproteins using a hybrid linear ion trap/FT-ICR mass spectrometer. *J. Am. Soc. Mass. Spectrom.* **17**, 168–179.

Subject Index

J. Li (ed.), *Functional Glycomics*, Methods in Molecular Biology 600,
DOI 10.1007/978-1-60761-454-8, © Humana Press, a part of Springer Science+Business Media, LLC 2010

G

L

M

N

P

S

V