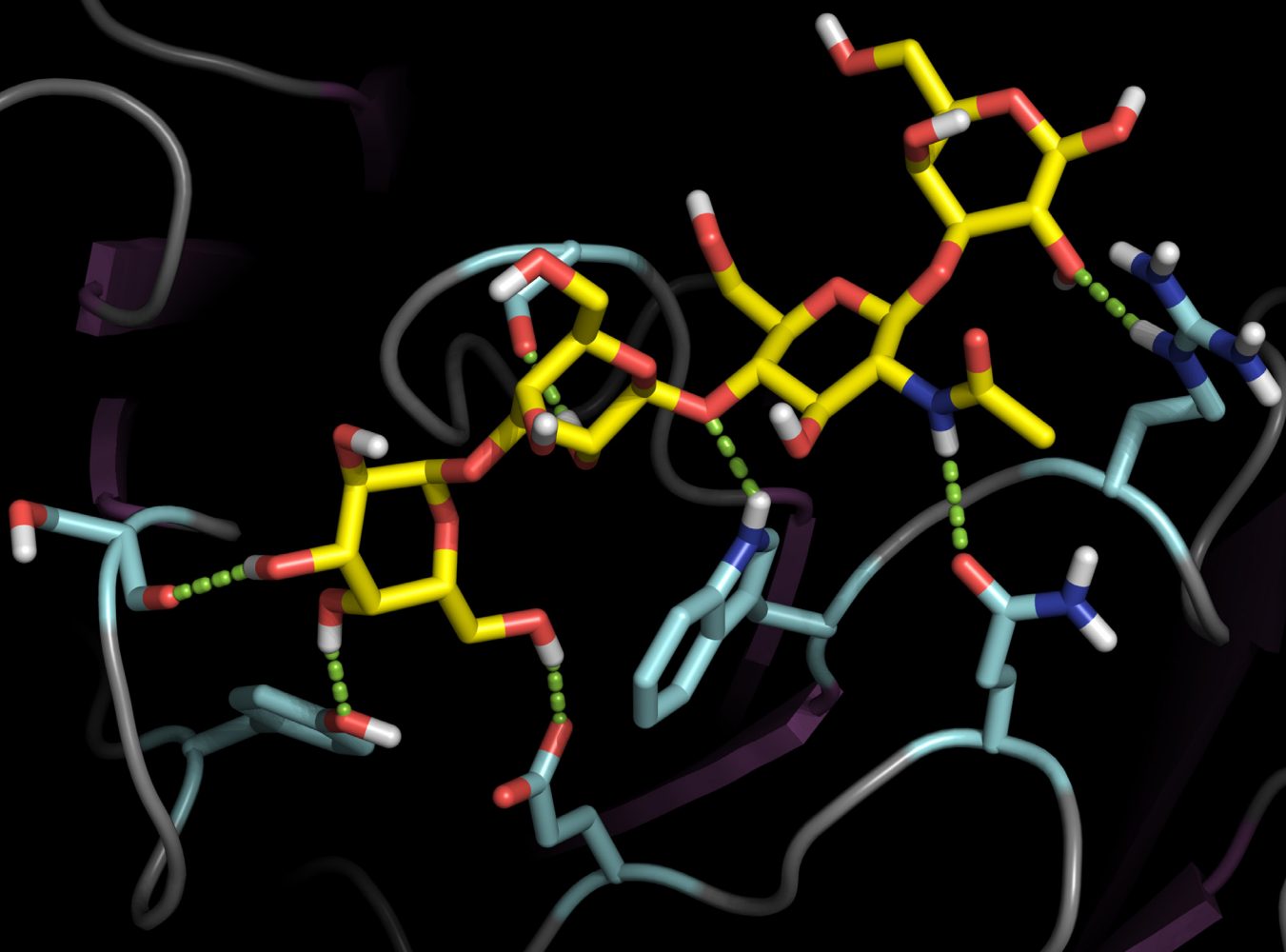


# STRUCTURAL AND COMPUTATIONAL GLYCOBIOLOGY: IMMUNITY AND INFECTION

EDITED BY : Elizabeth Yuriev and Mark Agostino  
PUBLISHED IN : Frontiers in Immunology





# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2015 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-638-8

DOI 10.3389/978-2-88919-638-8

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# STRUCTURAL AND COMPUTATIONAL GLYCOBIOLOGY: IMMUNITY AND INFECTION

Topic Editors:

**Elizabeth Yuriev**, Monash University, Melbourne, Australia

**Mark Agostino**, School of Biomedical Sciences, Curtin University, Perth, Australia



Computational methods play an increasingly crucial role in structural glycobiochemistry studies. This complex of a tetrasaccharide xenoantigen ( $\text{Gal}\alpha 1\text{-3Gal}\beta 1\text{-4GlcNAc}\beta 1\text{-3Gal}$ ) with the anti-Gal mAb 8.17 was predicted through a combination of molecular docking and computational site mapping techniques. Based on Agostino et al. *Glycobiology*. 2010;20:724-735.

Interest in understanding the biological role of carbohydrates has increased significantly over the last 20 years. The use of structural techniques to understand carbohydrate-protein recognition is still a relatively young area, but one that is of emerging importance. The high flexibility of carbohydrates significantly complicates the determination of high quality structures of their complexes with proteins. Specialized techniques are often required to understand the complexity of carbohydrate recognition by proteins. In this Research Topic, we focus on structural and computational approaches to understanding carbohydrate recognition by proteins involved in immunity and infection. Particular areas of focus include cancer immunotherapeutics, carbohydrate-lectin interactions, glycosylation and glycosyltransferases.

**Citation:** Elizabeth Yuriev and Mark Agostino, eds. (2015). *Structural and computational glycobiochemistry: immunity and infection*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-638-8

# Table of Contents

- 04 Editorial: Structural and computational glycobiology – immunity and infection**  
Mark Agostino and Elizabeth Yuriev
- 06 Molecular recognition of gangliosides and their potential for cancer immunotherapies**  
Ute Krengel and Paula A. Bousquet
- 17 Structure based refinement of a humanized monoclonal antibody that targets tumor antigen disialoganglioside GD2**  
Mahiuddin Ahmed, Jian Hu and Nai-Kong V. Cheung
- 23 Carbohydrate-mimetic peptides for pan anti-tumor responses**  
Thomas Kieber-Emmons, Somdutta Saha, Anastas Pashov, Behjatolah Monzavi-Karbassi and Ramachandran Murali
- 35 Predicting the origins of anti-blood group antibody specificity: a case study of the ABO A- and B-antigens**  
Spandana Makeneni, Ye Ji, David C. Watson, N. Martin Young and Robert J. Woods
- 44 Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants**  
Ling Yen Lee, Chi-Hung Lin, Susan Fanayan, Nicolle H. Packer and Morten Thaysen-Andersen
- 57 Crossroads between bacterial and mammalian glycosyltransferases**  
Inka Brockhausen
- 78 MCL and Mincle: C-type lectin receptors that sense damaged self and pathogen-associated molecular patterns**  
Mark B. Richardson and Spencer J. Williams
- 87 Computational and experimental prediction of human C-type lectin receptor druggability**  
Jonas Aretz, Eike-Christian Wamhoff, Jonas Hanske, Dario Heymann and Christoph Rademacher
- 99 Carbohydrates in cyberspace**  
Elizabeth Yuriev and Paul A. Ramsland



# Editorial: Structural and computational glycobiology – immunity and infection

Mark Agostino<sup>1,2\*</sup> and Elizabeth Yuriev<sup>3\*</sup>

<sup>1</sup> CHIRI Biosciences and Curtin Institute for Computation, School of Biomedical Sciences, Curtin University, Perth, WA, Australia,

<sup>2</sup> Centre for Biomedical Research, Burnet Institute, Melbourne, VIC, Australia, <sup>3</sup> Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Melbourne, VIC, Australia

**Keywords:** glycobiology, structural biology, infection, cancer immunotherapy, molecular modeling, molecular recognition, lectins, signaling

Historically deemed as the realm of the brave or the foolhardy, glycobiology has grown considerably as a discipline over the last 50 years. Carbohydrates, which were once considered to be mere “decorations” on proteins and lipid membranes, are increasingly demonstrated to afford specific roles in signaling and communication (1).

Although the rate of structures deposited into the Protein Data Bank continues to grow at an exponential rate, the characterization of new structures of carbohydrate–protein complexes is growing more modestly, still being very challenging and prone to errors (2). Computational methods are increasingly being pursued to provide structural insight into carbohydrate–protein interactions. The complex structure and high flexibility of carbohydrates, as well as difficulties associated with accurately computing binding energies for these interactions, present considerable challenges for the use of these methods in both understanding the carbohydrate–protein recognition and the structure-aided design of carbohydrate-based therapeutics. However, numerous computational approaches have been developed in recent years that address some of these issues (3–9). The Opinion piece in this Research Topic further highlights some computational resources that have been developed specifically for glycobiology (10).

Several carbohydrate classes, most notably gangliosides, Lewis antigens, and Thomsen–Friedenreich antigen, are of considerable interest for the development of cancer immunotherapeutics. Krenzel and Bousquet (11) present a comprehensive review on the importance of gangliosides not only to cancer therapeutics but also their relevance for signaling and in mediating infection by pathogens, as well as how their structure and presentation on glycolipids and glycoproteins influences their function and potential to be exploited in therapeutics. Ahmed et al. (12) describe the use of molecular modeling to optimize framework regions of an anti-ganglioside antibody, resulting in the identification of a new construct with enhanced stability, antigen binding, and cytotoxic properties. Kieber-Emmons et al. (13) discuss the challenges and frontiers associated with the development of peptides as immunogenic mimics of carbohydrates, particularly focusing on mimics of tumor-associated carbohydrate antigens.

Despite considerable advances in the understanding of many aspects of glycobiology, several fundamental processes remain only partially understood. An excellent example of this is the structural basis of antibody recognition of the blood group antigens (A, B, H). Makeneni et al. (14) combine docking with a recently developed carbohydrate-specific scoring function and molecular dynamics simulation to demonstrate the structural basis of A vs. B specificity of an anti-A antibody. Lee et al. (15) performed LC-MS/MS-based glycomics and proteomics, combined with structural analyses, of a wide range of glycosylated proteins in order to understand the differences in the glycosylation of secreted cell surface and intracellular proteins. The study correlates the presence of specific *N*-glycan terminations with their subcellular location, providing insight into pathophysiological conditions

## OPEN ACCESS

### Edited and reviewed by:

Kendall Arthur Smith,  
Weill Medical College of Cornell  
University, USA

### \*Correspondence:

Mark Agostino  
mark.agostino@curtin.edu.au;  
Elizabeth Yuriev  
elizabeth.yuriev@monash.edu

### Specialty section:

This article was submitted to  
Immunotherapies and Vaccines,  
a section of the journal  
Frontiers in Immunology

**Received:** 19 June 2015

**Accepted:** 01 July 2015

**Published:** 14 July 2015

### Citation:

Agostino M and Yuriev E (2015)  
Editorial: Structural and  
computational glycobiology –  
immunity and infection.  
Front. Immunol. 6:359.  
doi: 10.3389/fimmu.2015.00359

caused by glycosylation disorders. Brockhausen (16) provides a comprehensive review detailing known glycosyltransferases with overlapping activities between bacteria and mammals. In many cases, similar catalytic mechanisms between bacterial and mammalian glycosyltransferases can be identified, despite limited sequence similarity.

Lectins, particularly C-type lectins, are of considerable importance for immunity, mediating cell–cell recognition, and representing potential targets for the development of therapeutics. Notable C-type lectins include DC-SIGN and the selectins, known for their roles in the progression of HIV and cancer, respectively. Richardson and Williams (17) review the discovery and characterization of the macrophage C-type lectin (MCL) and

the macrophage-inducible C-type lectin (Mincle), their roles in initiating the immune response to infection, and the identification of activating ligands for these receptors. Aretz et al. (18) predict the druggability of a panel of C-type lectins, as well as perform fragment-based screening by nuclear magnetic resonance spectroscopy against DC-SIGN, langerin, and MCL. Their work highlights limitations in the application of computational methods to predict the druggability of this class of proteins.

The work presented in this Research Topic illustrates a small selection of the wide ranging research in this area and the considerable challenges associated with both understanding glycan function and targeting glycan interactions for the development of therapeutic agents.

## References

- Varki A. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (2009).
- Agirre J, Davies G, Wilson K, Cowtan K. Carbohydrate anomalies in the PDB. *Nat Chem Biol* (2015) 11:303. doi:10.1038/nchembio.1798
- Agostino M, Mancera RL, Ramsland PA, Yuriev E. AutoMap: a tool for analyzing protein–ligand recognition using multiple ligand binding modes. *J Mol Graph Model* (2013) 40:80–90. doi:10.1016/j.jmgm.2013.01.001
- Tessier MB, Grant OC, Heimburg-Molinaro J, Smith D, Jadey S, Gulick AM, et al. Computational screening of the human TF-glycome provides a structural definition for the specificity of anti-tumor antibody JAA-F11. *PLoS One* (2013) 8:e54874. doi:10.1371/journal.pone.0054874
- Kerzmann A, Fuhrmann J, Kohlbacher O, Neumann D. BALLDock/SLICK: a new method for protein–carbohydrate docking. *J Chem Inf Model* (2008) 48(8):1616–25. doi:10.1021/ci800103u
- Eid S, Saleh N, Zalewski A, Vedani A. Exploring the free-energy landscape of carbohydrate–protein complexes: development and validation of scoring functions considering the binding-site topology. *J Comput Aided Mol Des* (2014) 28:1191–204. doi:10.1007/s10822-014-9794-3
- Agostino M, Yuriev E, Ramsland PA. Antibody recognition of cancer-related gangliosides and their mimics investigated using in silico site mapping. *PLoS One* (2012) 7:e35457. doi:10.1371/journal.pone.0035457
- Pérez S, Sarkar A, Rivet A, Breton C, Imbert A. Glyco3D: a portal for structural glycosciences. *Methods Mol Biol* (2015) 1273:241–58. doi:10.1007/978-1-4939-2343-4\_18
- Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* (2008) 29:622–55. doi:10.1002/jcc.20820
- Yuriev E, Ramsland PA. Carbohydrates in cyberspace. *Front Immunol* (2015) 6:300. doi:10.3389/fimmu.2015.00300
- Krengel U, Bousquet PA. Molecular recognition of gangliosides and their potential for cancer immunotherapies. *Front Immunol* (2014) 5:325. doi:10.3389/fimmu.2014.00325
- Ahmed M, Hu J, Cheung N-K. Structure based refinement of a humanized monoclonal antibody that targets tumor antigen disialoganglioside GD2. *Front Immunol* (2014) 5:372. doi:10.3389/fimmu.2014.00372
- Kieber-Emmons T, Pashov A, Saha S, Monzavi-Karbassi B, Murali R. Carbohydrate mimetic peptides for pan anti-tumor responses. *Front Immunol* (2014) 5:308. doi:10.3389/fimmu.2014.00308
- Makeneni S, Ji Y, Watson DC, Young NM, Woods RJ. Predicting the origins of anti-blood group antibody specificity: a case study of the ABO A- and B-antigens. *Front Immunol* (2014) 5:397. doi:10.3389/fimmu.2014.00397
- Lee LY, Lin C-H, Fanayan S, Packer NH, Thaysen-Andersen M. Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants. *Front Immunol* (2014) 5:404. doi:10.3389/fimmu.2014.00404
- Brockhausen I. Crossroads between bacterial and mammalian glycosyltransferases. *Front Immunol* (2014) 5:492. doi:10.3389/fimmu.2014.00492
- Richardson MB, Williams SJ. MCL and mincle: C-type lectin receptors that sense damaged self and pathogen associated molecular patterns. *Front Immunol* (2014) 5:288. doi:10.3389/fimmu.2014.00288
- Aretz J, Wamhoff E-C, Hanske J, Heymann D, Rademacher C. Computational and experimental prediction of human C-type lectin receptor druggability. *Front Immunol* (2014) 5:323. doi:10.3389/fimmu.2014.00323

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Agostino and Yuriev. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Molecular recognition of gangliosides and their potential for cancer immunotherapies

Ute Krengel\* and Paula A. Bousquet\*

Department of Chemistry, University of Oslo, Oslo, Norway

**Edited by:**

Elizabeth Yuriev, Monash University, Australia

**Reviewed by:**

Paul A. Ramsland, Burnet Institute, Australia

Anne Imberty, CNRS, France

**\*Correspondence:**

Ute Krengel and Paula A. Bousquet, Department of Chemistry, University of Oslo, P.O. 1033 Blindern, NO-0315 Oslo, Norway

e-mail: ute.krengel@kjemi.uio.no; paula.bousquet@kjemi.uio.no

Gangliosides are sialic-acid-containing glycosphingolipids expressed on all vertebrate cells. They are primarily positioned in the plasma membrane with the ceramide part anchored in the membrane and the glycan part exposed on the surface of the cell. These lipids have highly diverse structures, not the least with respect to their carbohydrate chains, with *N*-acetylneuraminic acid (NeuAc) and *N*-glycolylneuraminic acid (NeuGc) being the two most common sialic-acid residues in mammalian cells. Generally, human healthy tissue is deficient in NeuGc, but this molecule is expressed in tumors and in human fetal tissues, and was hence classified as an onco-fetal antigen. Gangliosides perform important functions through carbohydrate-specific interactions with proteins, for example, as receptors in cell–cell recognition, which can be exploited by viruses and other pathogens, and also by regulating signaling proteins, such as the epidermal growth factor receptor (EGFR) and the vascular endothelial growth factor receptor (VEGFR), through lateral interaction in the membrane. Through both mechanisms, tumor-associated gangliosides may affect malignant progression, which makes them attractive targets for cancer immunotherapies. In this review, we describe how proteins recognize gangliosides, focusing on the molecular recognition of gangliosides associated with cancer immunotherapy, and discuss the importance of these molecules in cancer research.

**Keywords:** biological membranes, cancer immunotherapy, cell signaling, gangliosides, protein–carbohydrate interactions, glycosphingolipids, sialic acid, tumor-associated antigens

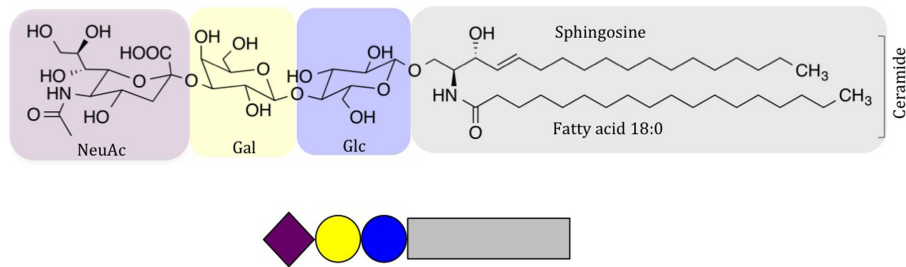
## INTRODUCTION

Few lipid species included in biological membranes have received as much attention as glycosphingolipids (GSLs), and especially gangliosides, sialic-acid-containing GSLs. They were discovered by Ernst Klenk in the 1940s, who proposed the term “ganglioside” due to the abundance of these molecules in “Ganglionzellen” (neurons). Gangliosides were later classified by Svennerholm according to the number of sialic-acid residues and chromatographic mobility (1). In contrast to glycerolipids, the lipid anchor in sphingolipids builds on the long-chain amino alcohol sphingosine, which is coupled *via* its amino group to a fatty acid to form ceramide (**Figure 1**). In gangliosides, the ceramide anchor is linked to a hydrophilic glycan head group, which is characterized by the presence of one or more sialic-acid residues (carbohydrates with a nine-carbon backbone and a carboxylic acid group); however, there is large variability of this structure. One example, the GM3 ganglioside, abundant in almost all healthy tissues, is shown in **Figure 1**. The large structural variability is related to developmental stage and cell type, and hundreds of gangliosides are known today (3–5). Variations in carbohydrate structure alone account for over a 100 different structures, and this number significantly increases, when ceramide variations are taken into account (4–7). Accumulating evidence indicates that many cellular events, including differentiation, growth, signaling, interactions, and immune reactions are highly influenced by gangliosides, and that these molecules may also cause malignancies. Positioned in the plasma membrane, gangliosides interact with other lipids and proteins,

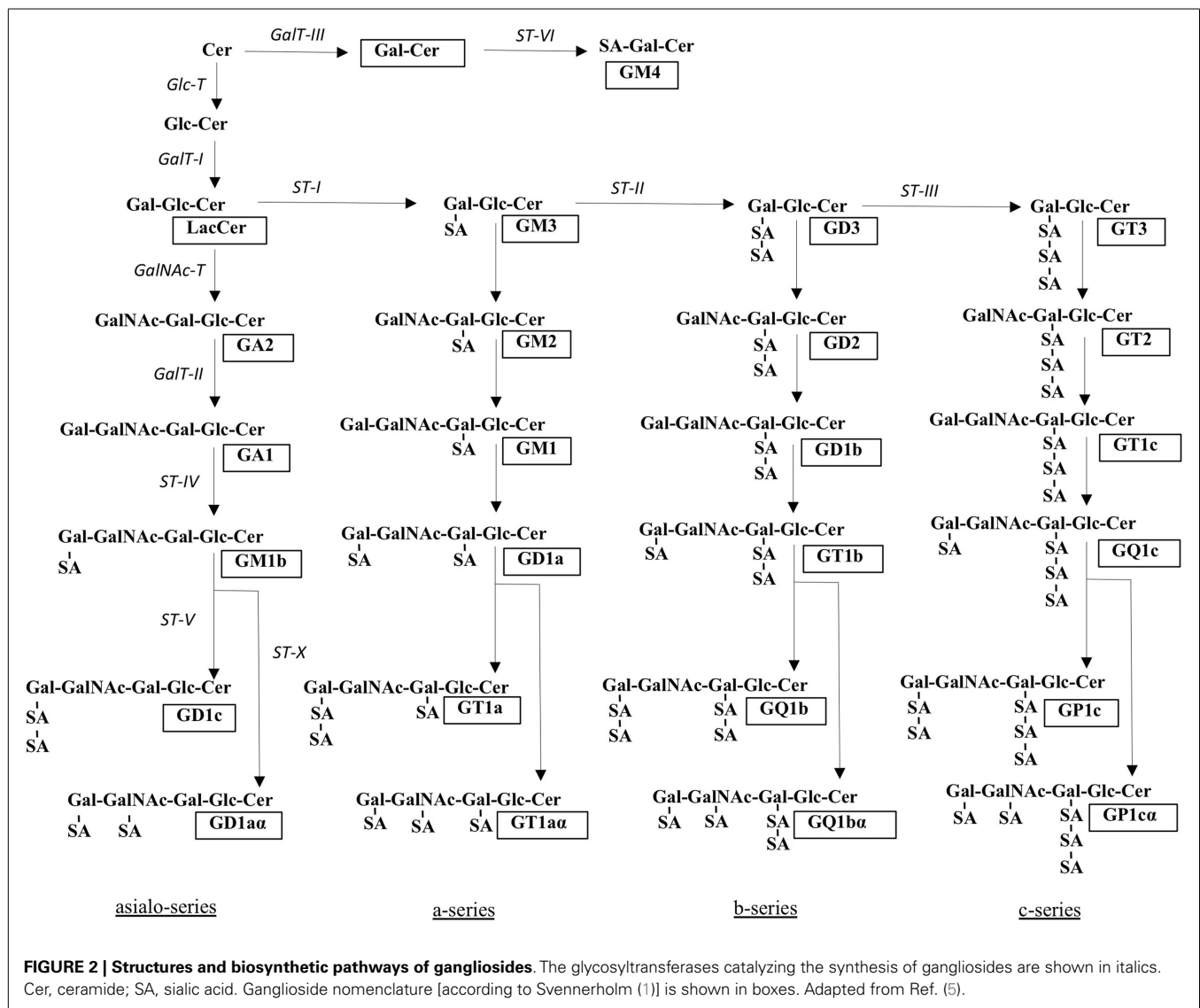
both laterally in the membrane and *via* their head groups, acting as cellular receptors that can be recognized by antibodies and other ganglioside-binding molecules. Here, we highlight the function and molecular interactions of gangliosides with high clinical significance.

## GANGLIOSIDES – GENERAL ARCHITECTURE, CELLULAR LOCALIZATION, AND BIOSYNTHESIS

Gangliosides consist of a lipid anchor, the ceramide, decorated by a glycan head group of various complexity. In cells, gangliosides are mainly found in the outer leaflets of the plasma membrane. Together with sphingomyelin and cholesterol, they form membrane microdomains, which play important roles in cell–cell communication and signal transduction (8–10). The synthesis of gangliosides starts in the ER compartment with the synthesis of the ceramide, the common precursor of all GSLs. Aided by the ceramide-transfer protein, CERT, ceramide is then transferred to the Golgi apparatus, and thereafter converted to glucosylceramide (GlcCer) (11). Subsequently, other carbohydrate residues are attached, one by one, catalyzed by glycosyltransferases, as described below (12, 13). The glycosyltransferases are specific to the sugar residues that they transfer and are grouped into families according to their specificity. Interestingly, all glycosyltransferase promoters lack the TATA sequence, and hence do not have any core promoter element characteristic for house-keeping genes. Although some indications relate their transcription to complex developmental and tissue-specific regulation, very



**FIGURE 1 | Schematic drawing of NeuAc GM3, a common ganglioside in vertebrate tissues.** Carbohydrate symbols follow the nomenclature of the Consortium for Functional Glycomics (2); purple diamond – *N*-acetylneuraminic acid; yellow circle – D-galactose; blue circle – D-glucose.



**FIGURE 2 | Structures and biosynthetic pathways of gangliosides.** The glycosyltransferases catalyzing the synthesis of gangliosides are shown in italics. Cer, ceramide; SA, sialic acid. Ganglioside nomenclature [according to Svennerholm (1)] is shown in boxes. Adapted from Ref. (5).

little is known about how glycosyltransferases are regulated (14). The molecular products are further subject to remodeling, by sialidases, sialyltransferases, and other enzymes, followed by vesicle sorting and fusion with the plasma membrane (15). Gangliosides are assumed to recycle to the plasma membrane from early

endosomes, and a degradation process is thought to take place at the late endosomal level (16).

The biosynthetic pathways of gangliosides are shown in **Figure 2**. After formation of the initial glucosylceramide, a galactose moiety is added to GlcCer to yield lactosylceramide

(LacCer), the common precursor for almost all gangliosides (except GM4). Addition of one sialic-acid residue to LacCer subsequently converts this precursor molecule to GM3. This reaction is catalyzed by sialyltransferase I (ST-I) or GM3 synthase. In the same manner, GD3 and GT3 can be generated by further addition of sialic-acid residues, catalyzed by ST-II or GD3 synthase and ST-III or GT3 synthase, respectively. The number of sialic-acid residues linked to the inner galactose residue (0, 1, 2, or 3) classify the gangliosides into asialo, a-, b-, or c-series (**Figure 2**), however, only trace amounts of gangliosides from the asialo- and c-series are found in adult human tissue (17).

## GANGLIOSIDES – BIOLOGICAL FUNCTION AND EXPLOITATION BY PATHOGENS

Gangliosides are key molecules in cellular recognition and signaling. They are primarily present in the plasma membranes of vertebrates, but have recently also been found in nuclear membranes, recognized as functionally important constituents (18, 19). Knock-out studies in mice have been essential for revealing the functions of gangliosides, especially in embryonic development and differentiation. For example, Yamashita et al. observed that mouse embryos carrying a knock-out in the glycosylceramide synthase enzyme did not survive more than 7.5 days (20). Other examples are studies of mice with a knock-down of GM3 synthase and GM2/GD2 synthase, which exhibit increased insulin sensitivity and decreased ability to repair nervous tissues, respectively (21, 22).

Because of the tight packing of lipids in membranes, gangliosides associate with other types of lipids, forming membrane subcompartments such as lipid rafts, to which specific proteins can associate (8, 23, 24). The organization of gangliosides in membranes will be further discussed in the Section “Organization and Presentation of Gangliosides in Biological Membranes.” Since gangliosides have the ability to interact with both sugars and proteins (see Sections “Gangliosides – Structure and Molecular Recognition”, “Organization and Presentation of Gangliosides in Biological Membranes”, and “Effect of Gangliosides on Membrane Proteins and Cellular Signaling”), a large range of events can be triggered or inhibited by these molecules. Cell growth, migration, differentiation, adhesion, and apoptosis are some examples (25, 26). The terminal sialic-acid residue(s) in particular are targets for many important intercellular interactions, but can also be exploited by pathogens that use these residues as a docking station to enter the cell (27).

Various pathogens, from viruses to bacteria and parasites, recognize sialic-acid residues on host cell membranes, several of these known to cause cancer. The most common recognition module is NeuAc; in addition, NeuGc and 9-O-acetylated sialic acids are also well-known receptors (28, 29). Examples of viral pathogens recognizing gangliosides are the influenza virus (30), simian virus 40 (SV40) (31), and polyomavirus (32, 33). Bacteria interact with gangliosides *via* toxins and adhesins, with the cholera toxin (34) and the Sialic-acid binding adhesin from the Class 1 carcinogen *Helicobacter pylori*, SabA (35, 36), being prominent examples. Gangliosides may also suppress natural killer (NK) cell cytotoxicity, through interaction with Siglec-7 (sialic-acid binding

immunoglobulin-like lectin 7), as elaborated further in the Section “Gangliosides and Cancer.”

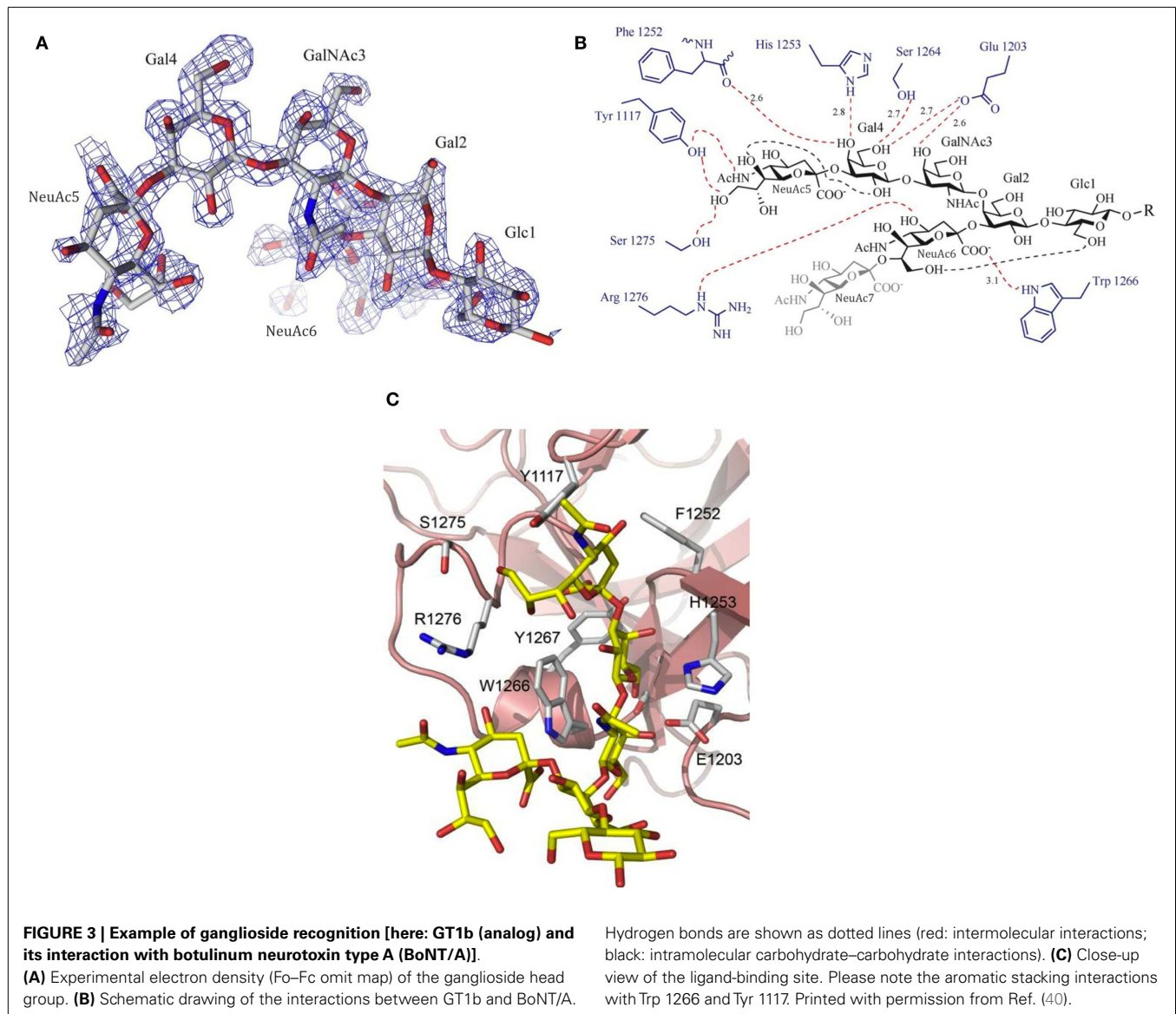
## GANGLIOSIDES – STRUCTURE AND MOLECULAR RECOGNITION

The molecular recognition of carbohydrates, with their large number of hydroxyl groups, is dominated by hydrogen bonds, with the binding specificity determined by the recognition of the characteristic OH-scaffolds of different sugars (37, 38). Many of these interactions are water-mediated, and sometimes, metal ions are involved. In addition, hydrophobic interactions contribute significantly to carbohydrate recognition, which may involve methyl groups such as in the monosaccharide fucose or the stacking against exposed hydrophobic patches of the sugar rings. A particularly typical molecular recognition mechanism of carbohydrates involves the CH- $\pi$  stacking of sugar rings against the side chains of aromatic amino acids (so-called “aromatic stacking interactions”), promoted by weak hydrogen bonds (39) (**Figure 3**).

Gangliosides are characterized by the presence of at least one sialic-acid residue, which in contrast to many other sugars is charged. This charge can be exploited by salt bridges with positively charged residues, but this is not necessarily the case (and in fact quite rare). The carboxylate group is often not even the most important recognition motif. For example, the fingerprint of the most common sialic acid, *N*-acetylneuraminic acid (NeuAc), which is derived from pyruvate and *N*-acetylmannosamine, generally involves the recognition of the *N*-acetyl group and the adjacent 4-OH-group, originating from mannose (which corresponds to 3-OH in hexoses) (41). Further H-bonding interactions are provided by the sialic-acid glycerol chain (also originating from mannose), which is recognized by a conserved binding motif common to a number of viral and bacterial sialic-acid binding proteins (42). In addition, conformer selection and clustering play important roles for the molecular recognition of gangliosides, as shown for example for the recognition of GM1 by the cholera toxin or galectin-1 (34, 43–45).

Carbohydrates in general are flexible molecules, but due to internal carbohydrate-carbohydrate interactions, the influence of the lipid anchor, or due to interactions with other molecules in the immediate neighborhood, rigid molecular epitopes may arise. As gangliosides are localized in the plasma membrane, the presentation of the carbohydrate epitopes in particular depends on the interaction with other lipids (8). However, the structural characterization of anchored gangliosides is difficult to achieve. State-of-the-art lipid simulations are described by Vattulainen and Róg (46), but these often fail to take the glycan head groups into account. Nevertheless a few studies have been undertaken that do just that. One interesting example is the atomic-resolution conformational analysis of GM3 in a bilayer composed of dimyristoylphosphatidylcholine (DMPC) (47). Two known GM3-binding proteins [sialoadhesin, PDB ID: 1QFO (48), and wheat germ agglutinin, PDB ID: 2CWG (49)] were studied in order to evaluate the importance of carbohydrate accessibility and ganglioside recognition. Probing the presentation and dynamics of the glycan head group, DeMarco and Woods observed significantly altered accessibility of the less exposed carbohydrate residues Gal and Glc, even though the internal structural properties for





membrane-bound versus soluble GM3 were unchanged. On the other hand, the terminal NeuAc-residue remained almost fully exposed. The difference in accessibility is likely of considerable importance for the initial recognition of GM3 by a receptor protein, although subsequent recognition events may include the glycan residues embedded deeper in the membrane. The less exposed residues may also indirectly affect recognition, by ceramide-Glc and Glc-Gal rotations, altering NeuAc presentation. Furthermore, the hydrophobic ceramide together with the polar Glc residue may regulate the insertion depth.

### ORGANIZATION AND PRESENTATION OF GANGLIOSIDES IN BIOLOGICAL MEMBRANES

Cellular membranes serve both as segregation barriers and as facilitators of cellular communication. Positioned in the cell membrane, lipids interact laterally with other membrane components (lipids or membrane proteins), and also serve as cellular receptors,

through their exposed head groups. In the past decade many studies have focused on the lateral characterization of membranes and it is now well-established that highly unsaturated components, like glycerophospholipids, provide the membrane with flexibility, while saturated components, such as GSLs, create order in biological membranes (10). Furthermore, the shape and length of the lipids determine the shape, size, and stability of cellular membranes (50). The ceramide part of gangliosides is characterized by a rigid and planar structure, composed of saturated acyl chains, which can be more tightly packed. Together with other membrane sphingolipids and cholesterol, they can segregate and form dynamic nanoscale “clusters”, also called lipid rafts (8, 24, 51), to which specific proteins associate, hitching a ride.

Apparently, the density of GSLs can also influence their structure, affecting antigen specificity. For example, an antibody established by immunizing mice with syngenic B16 melanoma, named M2590, reacted only with melanoma and not with healthy

tissues (52). Remarkably, the target epitope was later identified as GM3, an abundant ganglioside in membranes of normal cells (53). Further studies showed that a ganglioside density above a threshold value was required for reactivity, suggesting that this antibody recognized more densely packed GM3 (54). These results indicate that ganglioside antigens can be differently organized in tumor cells compared to normal cells and that some ganglioside antigens are fully antigenic when organized in clusters, but fail to bind antibodies when their density is under a threshold value (54, 55).

How can this be explained? This brings us back to the structural characterization of GSLs in biological membranes. One example has already been described [GM3 in DMPC bilayer; (47)]. Two other interesting studies evaluate the effect of cholesterol on GSL structure (56, 57), building on earlier work by Pascher and coworkers (58). Notably, cholesterol was found to introduce a tilt in the glycolipid head group from a conformation almost perpendicular to the membrane surface to an alignment parallel to the membrane (Figure 4). The culprit appears to be an H-bonding network involving the cholesterol OH-group, the sphingosine amide, and the oxygen of the glycosidic bond (56). Similar lipid-raft-specific conformational changes of GSLs may be critical for the entry of bacterial toxins or viruses into host cells (8, 59).

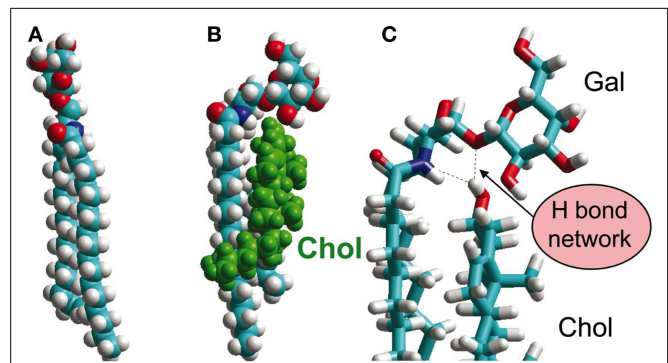
Glycosphingolipids are not always fully accessible, however. Their short head groups may be hidden in the “jungle” of membrane proteins or even masked by sialic-acid binding proteins positioned near the GSLs in the membranes (i.e., in *cis*). Such a scenario is postulated, e.g., for Siglecs, a family of lectins that modulate innate and adaptive immune functions. *Trans* interactions may still occur, e.g., for higher-affinity ligands that can out-compete the *cis* ligands, however, in general, accessibility will be reduced.

### EFFECT OF GANGLIOSIDES ON MEMBRANE PROTEINS AND CELLULAR SIGNALING

It has been suggested that also the activation of membrane proteins can be influenced by lipid cluster association. In addition to lateral interaction with the lipid tails in the cell membrane, such interactions may exploit the unique properties of sphingolipids, bearing a carbonyl oxygen, a hydroxyl group, and an amide nitrogen, thus being able to act as both H-bond donors and acceptors (60). As described in the previous section, gangliosides and other GSLs may further cause conformational changes of the glycan head group, which may either interact directly with amino acids of the extracellular part of the protein or alternatively interact with the sugar residues of a glycosylated protein, affecting protein activity.

Most growth factor receptors are known to be regulated by gangliosides (9). Here, we will discuss two examples of membrane proteins important for cancer research and immunotherapy: the epidermal growth factor receptor (EGFR) and the vascular endothelial growth factor receptor (VEGFR) (Table 1). A number of cancers are characterized by hyper-activated EGFRs, either caused by mutations or over-expression (61–63). Another important factor for tumor progression is the growth of new blood vessels. Tumor cells produce and release the growth factor VEGF, stimulating the VEGFR, and ultimately resulting in proliferation and migration of vascular endothelial cells (64).

The EGFR is known to undergo ligand-dependent dimerization, resulting in an autophosphorylation of tyrosine residues at



**FIGURE 4 | Glycosphingolipid interaction with cholesterol, an important constituent of lipid rafts. (A)** GalCer, extended conformation. **(B,C)** GalCer, tilted conformation, induced by H-bonding interactions with cholesterol OH-group, shown in **(C)** [(A,B): space-filling representation, **(C)**: stick representation]. Printed with permission from Ref. (56), in an extension of earlier work by Nyholm et al. (58).

**Table 1 | Gangliosides affecting the growth factor receptors EGFR and VEGFR.**

Ganglioside	Growth factor receptor	Reference
GM3	EGFR	(65–68)
GM1	EGFR	(68, 69)
GM2	EGFR	(70, 71)
GM4	EGFR	(70)
GD3	EGFR	(70, 72)
GD1a	EGFR	(68, 73)
GT1b	EGFR	(68)
GM3	VEGFR	(74, 75)
GD1a	VEGFR	(75, 76)
GD3	VEGFR	(77)

the C-terminal tail of the protein (78). This initiates downstream signaling, leading to adhesion, cell migration, and proliferation (79). More recently, the EGFR has also been shown to undergo ligand-independent dimerization, a phenomenon that is poorly understood (80). Such ligand-free dimers can also be functionally active, but this is not always the case.

Several membrane ligands have been shown to affect signaling by the EGFR and the VEGFR. The GM3 ganglioside, a well-known regulator of the insulin receptor (81), has an inhibitory effect on both the EGFR and the VEGFR, while the ganglioside GD1a strongly induces VEGFR-2 activation (26, 66, 70, 75, 82, 83). Moreover, the proangiogenic effects of GD1a can be efficiently reduced by GM3 (75). GM3 has been suggested to inhibit VEGFR-2 activation by blocking both growth factor binding and receptor dimerization through direct interaction with the extracellular domain of the VEGFR (74). The molecular interaction between the EGFR and GM3 is not fully elucidated, although it has been studied extensively. It has been shown that the inhibition of EGFR activation by GM3 involves the binding of the ganglioside to the GlcNAc-terminated N-glycans on the EGFR, suggesting carbohydrate–carbohydrate interactions (65, 67, 84,



85). In addition, increasing evidence points to the integral importance of ganglioside organization in the membrane for signal transduction (affecting the localization and activation of growth factor receptors). For example, recent computer simulations of the EGFR embedded in the membrane suggest that membrane lipids, especially anionic species, interact extensively with the EGFR (86). These interactions are more pronounced for the inactive EGFR, due to electrostatic interactions with the EGFR's intracellular domain, which may explain the inhibitory effect of GM3 on EGFR activation.

Cellular biological membranes are complex and the dynamics difficult to study. Even small modifications like the fluorescent labeling of lipids may critically affect bulk membrane properties as well as ligand–receptor interactions in biological environments (87). To generate a more controllable system, Coskun et al. reconstituted EGFR into proteoliposomes with defined lipid composition, with either uniform liquid-disordered (ld) membrane phases or a combination of disordered and ordered (ld/lo) domains. Adding gangliosides to this system, they found that GM3 had a strong inhibitory effect on EGFR activation, without interfering with ligand-binding, but in ld/lo proteoliposomes only (66). It would be of significant clinical interest to investigate how targeting GM3 by immunotherapy affects EGFR and VEGFR signaling, and whether the presence of both targets (GM3 clusters and EGFR/VEGFR) affect antibody efficiency and affinity.

## GANGLIOSIDES AND CANCER

Gangliosides play important roles in many normal physiological processes, such as cell growth, differentiation, and embryogenesis (20), but also in pathological events like cellular malignancy and metastasis (88) (see **Table 2** for examples of gangliosides expressed in human cancer cells). Tumor formation results from autonomous uncontrolled proliferation of neoplastic cells, while metastasis occurs when tumor cells are released from the primary tumor and continue to proliferate at a distant site. Multiple factors affect these processes, in which gangliosides may serve both as inhibitory and stimulating molecules. For example, it has been

shown that highly metastatic melanoma cells have high expression levels of GD3. This is in contrast to poorly metastatic cells or the normal counterpart, melanocytes, which express very low levels of GD3 (89–91), suggesting a role of GD3 in transforming melanocytes into melanomas and promotion of metastasis. Gangliosides may suppress NK cell cytotoxicity through interaction with Siglec-7, which preferentially binds to gangliosides of the b-series, as found for cells engineered to overexpress GD3 (92). The high expression levels of the GD3 ganglioside in melanoma may hence reflect the suppressed efficiency of NK cell cytotoxicity against these tumor cells. The function of gangliosides as suppressors of the anti-tumor immune response is well-documented in many studies, with tumor-associated gangliosides reported to down-regulate the activity of T and B cells, NK cytotoxicity and active dendritic cells, among others (93–95). For instance, T-cell dysfunction is promoted by the GM2 ganglioside, however, an antibody targeting GM2 was able to block 50–60% of T-cell apoptosis (94).

Gangliosides are also shed from the tumor to the microenvironment in greater quantities than normal cells. Shed gangliosides can interact with proteins or be incorporated into the membrane of other cells, leading to signaling events or interactions with healthy cells (112–114). For example, the addition of exogenous GD3 to the culture medium of glioma cells was found to stimulate the release of VEGF (115). Taken together, these observations suggest a multitude of mechanisms by which tumor-associated gangliosides may contribute to malignancy and cancer progression.

Many of the tumor-associated gangliosides are also found in normal healthy tissues, but are over-expressed in tumors, while other antigens are only found in cancer cells. An interesting example is the sialic-acid NeuGc, which is found in several tumor types, such as melanoma and breast cancer (116). Among all variants of sialic acids, NeuAc and NeuGc are the most abundant; however, humans are a notable exception. Due to a 92-bp deletion in the gene coding for CMP-NeuAc hydroxylase (*cmah*), humans lack a functional enzyme required for generation of NeuGc (117, 118). Nevertheless, NeuGc is present in fetal tissues and malignant cells

**Table 2 | Gangliosides expressed in human cancer cells.**

Ganglioside	Structure	Cancer type	Reference
NeuAc GM3	$\alpha$ Neu5Ac(2-3) $\beta$ DGal(1-4) $\beta$ DGlc(1-1)Cer	Melanoma, NSCLC, breast carcinoma, renal carcinoma	(89, 96–100)
NeuGc GM3	$\alpha$ Neu5Gc(2-3) $\beta$ DGal(1-4) $\beta$ DGlc(1-1)Cer	Colon cancer, retinoblastoma, melanoma, breast carcinoma, neuroectodermal cancer, Wilms tumor	(98, 99, 101–104)
GM2	$\beta$ DGalNAc(1-4)[ $\alpha$ Neu5Ac(2-3)] $\beta$ DGal(1-4) $\beta$ DGlc(1-1)Cer	Melanoma, neuroblastoma, SCLC, t-ALL, breast carcinoma, renal carcinoma	(74, 96, 99, 100, 105–107)
GM1	$\beta$ DGal(1-3) $\beta$ DGalNAc[ $\alpha$ Neu5Ac(2-3)] $\beta$ DGal(1-4) $\beta$ DGlc(1-1)Cer	SCLC, renal carcinoma	(99, 106)
GD3	$\alpha$ Neu5Ac(2-8) $\alpha$ Neu5Ac(2-3) $\beta$ DGal(1-4) $\beta$ DGlc(1-1)Cer	Melanoma, neuroblastoma, glioma, SCLC, t-ALL, breast carcinoma	(25, 89, 96, 97, 105, 107–111)
GD2	$\beta$ DGalNAc(1-4)[ $\alpha$ Neu5Ac(2-8) $\alpha$ Neu5Ac(2-3)] $\beta$ DGal(1-4) $\beta$ DGlc(1-1)Cer	Melanoma, neuroblastoma, glioma, SCLC, t-ALL	(89, 96, 97, 105–109)

$\alpha$ NeuAc = 5-acetyl- $\alpha$ -neuraminic acid,  $\alpha$ NeuGc = 5-glycolyl- $\alpha$ -neuraminic acid,  $\beta$ DGal =  $\beta$ -D-galactopyranose,  $\beta$ DGalNAc = N-acetyl- $\beta$ -D-galactopyranose,  $\beta$ DGlc =  $\beta$ -D-glucopyranose, Cer = ceramide, NSCLC = non-small-cell lung carcinoma, SCLC = small-cell lung carcinoma.

(99, 119, 120). For this reason, NeuGc was assumed to classify as an “onco-fetal” antigen, being expressed in the fetus, suppressed during adult life and re-expressed in malignant cells. However, since humans lack the putative active site of the enzyme, other explanations must lie at the heart of this change in carbohydrate profile. Diet incorporation, hypoxic conditions, and endogenous metabolic mechanisms are currently being discussed as possible origins of the increased levels of NeuGc (116, 121–124). Getting to grips with the high NeuGc-ganglioside levels is important, since this property appears to correlate with a poor prognosis. Specifically, recent studies indicate that non-small-cell lung cancer (NSCLC) patients with high NeuGc-ganglioside expression exhibit a low overall survival rate and a significantly lower progression-free survival rate (125). These findings are consistent with recent experiments demonstrating that the silencing of the *cmah* gene in NeuGc GM3-expressing L1210 mouse lymphocytic leukemia B cells caused a shift to NeuAc GM3 expression and a concomitant reduction of tumorigenicity (126).

Interestingly, it has been shown that serum from healthy humans contains antibodies recognizing glycoconjugates exhibiting NeuGc (127, 128). These antibodies are called Hanganutziu–Deicher (HD) antibodies, and were first described by Hanganutziu (129) and Deicher (130) [as cited in Ref. (131)] independently in the 1920s. HD antibodies attract complement molecules to malignant cells (132, 133). The level decreases with age, which may correlate with an increased cancer risk at higher age (133). Characteristic for natural antibodies is that they recognize highly conserved antigens (134). Importantly, auto-antibodies against tumor-associated antigens can arise and be detected early, before symptoms occur, and hence have potential for early diagnosis (135–137). In line with this hypothesis, a recent study reported that healthy donors exhibited low levels of anti-NeuGc GM3 antibodies (decreasing with age), while these antibodies were absent in NSCLC patients (138).

## GANGLIOSIDE-BASED THERAPY

Cancer immunotherapy is a highly promising approach to cancer treatment, which has been gaining grounds only recently (139). In contrast to traditional therapies like chemo- or radiation-therapy, immunotherapies constitute a much more targeted approach that promises higher specificity while eliciting fewer side effects. As the name states, this type of therapy uses the immune system to treat cancer. There are two main approaches (139, 140): (i) tumor-associated antigens or derivatives or mimics of these may be used as active therapeutic vaccines, priming the body to launch an immune attack against these molecules and hence the tumor cells (overcoming the body's tolerance of self-antigens); (ii) alternatively, antibodies may be used for passive immunotherapy, either coupled to toxins, radioactivity or on their own, relying on processes like antibody-dependent cell-mediated cytotoxicity (ADCC) or complement-dependent cytotoxicity (CDC). In both cases, effective immunotherapy relies on the choice of the antigen. Notably, in a recent project for prioritization of cancer antigens, 4 of the 75 selected antigens were gangliosides (GD2, GD3, fucosyl-GM1, and *N*-acetyl GM3), and additional targets, like the EGFR and the VEGFR, are known to interact with gangliosides (141).

Several antibodies targeting tumor-associated gangliosides are currently under investigation in pre-clinical or clinical studies, also including molecular vaccines. One example, the antibody 3F8, targets GD2, which is highly expressed in aggressive cancer, such as pediatric neuroblastoma (142). Other examples are 14F7 and chP3, both of which specifically recognize NeuGc GM3, discriminating it from the highly similar NeuAc GM3. So far, no crystal structures of these complexes have been reported, however, computer docking studies, *in silico* site mapping and phage display studies are contributing to reveal the recognition mechanisms of these promising tools (143–146). In addition, two NeuGc-ganglioside-based vaccines are currently tested in clinical trials (phase III); these are Racotumomab, an anti-idiotypic antibody<sup>1</sup> registered and launched in Cuba and Argentina under the trade name Vaxira (147) and NeuGc GM3/VSSP, a NeuGc GM3 ganglioside conjugated into very small proteoliposomes. In the ongoing clinical trials, the NeuGc GM3/VSSP and Racotumomab vaccines show efficacy and are well-tolerated by patients with advanced cutaneous melanoma (148) and NSCLC (149), respectively. This represents a significant step forward from the first, unsuccessful, attempt of developing a ganglioside-based vaccine – the GMK (GM2-based) vaccine for melanoma (150, 151). These molecules are part of a growing arsenal of targeted molecular weapons against cancer, which may be used as stand-alone therapy, but will more likely be employed as adjuvant therapy, in combination with or following standard treatment such as surgery, radiation, or chemotherapy. For example, based on the important roles of NeuGc GM3 and the EGFR for tumor cell immune evasion and proliferation, a combination therapy targeting both molecules may provide a rationale for fighting tumor cells. This combination is currently tested using Racotumomab and a vaccine targeting EGF in NSCLC patients, showing, so far, promising clinical results (152).

## CONCLUSION

Today, we are still far from fully understanding the roles, structures, and mechanisms of gangliosides in biological systems, and only at the beginning of the exploitation of these molecules in potential therapies. However, the importance of these molecules is evident, and technology development is picking up pace (7, 46, 153, 154). We are looking forward to a bright future, in which gangliosides are fully appreciated, and unfold their full potential in targeted therapies.

## ACKNOWLEDGMENT

We would like to thank Steffi Munack for improving the quality of Figure 1.

## REFERENCES

1. Svennerholm L. Chromatographic separation of human brain gangliosides. *J Neurochem* (1963) 10:613–23. doi:10.1111/j.1471-4159.1963.tb08933.x
2. Nomenclature Committee, Consortium for Functional Glycomics. *Symbol and Text Nomenclature for Representation of Glycan Structure*. (2014). Available

<sup>1</sup>Explanation: Antibodies (Ab2) raised against the primary antibody (Ab1, generated in the original immune response against cancer antigens) are also called “anti-idiotypic” antibodies. They may have potential as therapeutic vaccines if they give rise to a third class of antibodies (Ab3) that resembles the primary antibody (Ab1).

- from: <http://www.functionalglycomics.org/static/consortium/Nomenclature.shtml>.
3. Schwarz A, Futerman AH. The localization of gangliosides in neurons of the central nervous system: the use of anti-ganglioside antibodies. *Biochim Biophys Acta* (1996) **1286**(3):247–67. doi:10.1016/S0304-4157(96)00011-1
  4. Yu RK, Yanagisawa M, Ariga T. Glycosphingolipid structures. In: Kamerling JP, editor. *Comprehensive Glycoscience* (Vol. 1), Oxford: Elsevier (2007). p. 73–122.
  5. Yu RK, Tsai Y-T, Ariga T, Yanagisawa M. Structures, biosynthesis, and functions of gangliosides – an overview. *J Oleo Sci* (2011) **60**(10):537–44. doi:10.5650/jos.60.537
  6. Futerman AH, Hannun YA. The complex life of simple sphingolipids. *EMBO Rep* (2004) **5**(8):777–82. doi:10.1038/sj.embor.7400208
  7. Shevchenko A, Simons K. Lipidomics: coming to grips with lipid diversity. *Nat Rev Mol Cell Biol* (2010) **11**(8):593–8. doi:10.1038/nrm2934
  8. Simons K, Sampaio JL. Membrane organization and lipid rafts. *Cold Spring Harb Perspect Biol* (2011) **3**(10):a004697. doi:10.1101/cshperspect.a004697
  9. Lopez PHH, Schnaar RL. Gangliosides in cell recognition and membrane protein regulation. *Curr Opin Struct Biol* (2009) **19**(5):549–57. doi:10.1016/j.sbi.2009.06.001
  10. Sonnino S, Mauri L, Chigorno V, Prinetti A. Gangliosides as components of lipid membrane domains. *Glycobiology* (2007) **17**(1):1R–13R. doi:10.1093/glycob/cwl052
  11. Hanada K, Kumagai K, Yasuda S, Miura Y, Kawano M, Fukasawa M, et al. Molecular machinery for non-vesicular trafficking of ceramide. *Nature* (2003) **426**(6968):803–9. doi:10.1038/nature02188
  12. Merrill AH Jr. *De novo* sphingolipid biosynthesis: a necessary, but dangerous, pathway. *J Biol Chem* (2002) **277**(29):25843–6. doi:10.1074/jbc.R200009200
  13. Kolter T, Proia RL, Sandhoff K. Combinatorial ganglioside biosynthesis. *J Biol Chem* (2002) **277**(29):25859–62. doi:10.1074/jbc.R200001200
  14. Zeng G, Yu RK. Cloning and transcriptional regulation of genes responsible for synthesis of gangliosides. *Curr Drug Targets* (2008) **9**(4):317–24. doi:10.2174/138945008783954925
  15. Maccioni HJF. Glycosylation of glycolipids in the Golgi complex. *J Neurochem* (2007) **103**(Suppl 1):81–90. doi:10.1111/j.1471-4159.2007.04717.x
  16. Tettamanti G. Ganglioside/glycosphingolipid turnover: new concepts. *Glycoconj J* (2004) **20**(5):301–17. doi:10.1023/B:GLYC.0000033627.02765.cc
  17. Sandhoff K, Kolter T. Biosynthesis and degradation of mammalian glycosphingolipids. *Philos Trans R Soc Lond B Biol Sci* (2003) **358**(1433):847–61. doi:10.1098/rstb.2003.1265
  18. Saito M, Sugiyama K. Characterization of nuclear gangliosides in rat brain: concentration, composition, and developmental changes. *Arch Biochem Biophys* (2002) **398**(2):153–9. doi:10.1006/abbi.2001.2725
  19. Ledeen RW, Wu G. Nuclear lipids: key signaling effectors in the nervous system and other tissues. *J Lipid Res* (2004) **45**(1):1–8. doi:10.1194/jlr.R300015-JLR200
  20. Yamashita T, Wada R, Sasaki T, Deng C, Bierfreund U, Sandhoff K, et al. A vital role for glycosphingolipid synthesis during development and differentiation. *Proc Natl Acad Sci U S A* (1999) **96**(16):9142–7. doi:10.1073/pnas.96.16.9142
  21. Yamashita T, Hashiramoto A, Haluzik M, Mizukami H, Beck S, Norton A, et al. Enhanced insulin sensitivity in mice lacking ganglioside GM3. *Proc Natl Acad Sci U S A* (2003) **100**(6):3445–9. doi:10.1073/pnas.0635898100
  22. Takamiya K, Yamamoto A, Furukawa K, Yamashiro S, Shin M, Okada M, et al. Mice with disrupted GM2/GD2 synthase gene lack complex gangliosides but exhibit only subtle defects in their nervous system. *Proc Natl Acad Sci U S A* (1996) **93**(20):10662–7. doi:10.1073/pnas.93.20.10662
  23. Simons K, Toomre D. Lipid rafts and signal transduction. *Nat Rev Mol Cell Biol* (2000) **1**(1):31–9. doi:10.1038/35036052
  24. Simons K, Ikonen E. Functional rafts in cell membranes. *Nature* (1997) **387**(6633):569–72. doi:10.1038/42408
  25. Hakomori S-i, Igarashi Y. Functional role of glycosphingolipids in cell recognition and signaling. *J Biochem* (1995) **118**(6):1091–103.
  26. Bremer EG, Schlessinger J, Hakomori S-i. Ganglioside-mediated modulation of cell growth. Specific effects of GM<sub>3</sub> on tyrosine phosphorylation of the epidermal growth factor receptor. *J Biol Chem* (1986) **261**(5):2434–40.
  27. Imberty A, Varrot A. Microbial recognition of human cell surface glycoconjugates. *Curr Opin Struct Biol* (2008) **18**(5):567–76. doi:10.1016/j.sbi.2008.08.001
  28. Lehmann F, Tiralongo E, Tiralongo J. Sialic acid-specific lectins: occurrence, specificity and function. *Cell Mol Life Sci* (2006) **63**(12):1331–54. doi:10.1007/s00018-005-5589-y
  29. Schauer R, Srinivasan GV, Wipfler D, Knipf B, Schwartz-Albiez R. O-acetylated sialic acids and their role in immune defense. In: Wu AM, editor. *The Molecular Immunology of Complex Carbohydrates-3*. Advances in Experimental Medicine and Biology (Vol. 705). New York: Springer (2011). p. 525–48. doi:10.1007/978-1-4419-7877-6\_28
  30. Suzuki Y, Nagao Y, Kato H, Matsumoto M, Nerome K, Nakajima K, et al. Human influenza A virus hemagglutinin distinguishes sialylloligosaccharides in membrane-associated gangliosides as its receptor which mediates the adsorption and fusion processes of virus infection. Specificity for oligosaccharides and sialic acids and the sequence to which sialic acid is attached. *J Biol Chem* (1986) **261**(36):17057–61.
  31. Neu U, Woellner K, Gauglitz G, Stehle T. Structural basis of GM1 ganglioside recognition by simian virus 40. *Proc Natl Acad Sci U S A* (2008) **105**(13):5219–24. doi:10.1073/pnas.0710301105
  32. Erickson KD, Garcea RL, Tsai B. Ganglioside GT1b is a putative host cell receptor for the Merkel cell polyomavirus. *J Virol* (2009) **83**(19):10275–9. doi:10.1128/JVI.00949-09
  33. Neu U, Maginnis MS, Palma AS, Ströhl LJ, Nelson CDS, Feizi T, et al. Structure-function analysis of the human JC polyomavirus establishes the LSTc pentasaccharide as a functional receptor motif. *Cell Host Microbe* (2010) **8**(4):309–19. doi:10.1016/j.chom.2010.09.004
  34. Merritt EA, Sarfaty S, van den Akker F, L'Hoir C, Martial JA, Hol WGJ. Crystal structure of cholera toxin B-pentamer bound to receptor GM<sub>1</sub> pentasaccharide. *Prot Sci* (1994) **3**(2):166–75. doi:10.1002/pro.5560030202
  35. Mahdavi J, Sondén B, Hurtig M, Olfat FO, Forsberg L, Roche N, et al. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* (2002) **297**(5581):573–8. doi:10.1126/science.1069076
  36. Roche N, Ångström J, Hurtig M, Larsson T, Borén T, Teneberg S. *Helicobacter pylori* and complex gangliosides. *Infect Immun* (2004) **72**(3):1519–29. doi:10.1128/IAI.72.3.1519-1529.2004
  37. Drickamer K. Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins. *Structure* (1997) **5**(4):465–8. doi:10.1016/S0969-2126(97)00202-5
  38. Gabius H-J, André S, Jiménez-Barbero J, Romero A, Solís D. From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem Sci* (2011) **36**(6):298–313. doi:10.1016/j.tibs.2011.01.005
  39. del Carmen Fernández-Alonso M, Cañada FJ, Jiménez-Barbero J, Cuevas G. Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J Am Chem Soc* (2005) **127**(20):7379–86. doi:10.1021/ja051020+
  40. Stenmark P, Dupuy J, Imamura A, Kiso M, Stevens RC. Crystal structure of botulinum neurotoxin type A in complex with the cell surface co-receptor GT1b – insight into the toxin–neuron interaction. *PLoS Pathog* (2008) **4**(8):e1000129. doi:10.1371/journal.ppat.1000129
  41. Lis H, Sharon N. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem Rev* (1998) **98**(2):637–74. doi:10.1021/cr940413g
  42. Baker HM, Basu I, Chung MC, Caradoc-Davies T, Fraser JD, Baker EN. Crystal structures of the staphylococcal toxin SSL5 in complex with sialyl Lewis X reveal a conserved binding site that shares common features with viral and bacterial sialic acid binding proteins. *J Mol Biol* (2007) **374**(5):1298–308. doi:10.1016/j.jmb.2007.09.091
  43. Lauer S, Goldstein B, Nolan RL, Nolan JP. Analysis of cholera toxin–ganglioside interactions by flow cytometry. *Biochemistry* (2002) **41**(6):1742–51. doi:10.1021/bi0112816
  44. Siebert H-C, André S, Lu S-Y, Frank M, Kaltner H, van Kuik JA, et al. Unique conformer selection of human growth-regulatory lectin galectin-1 for ganglioside GM<sub>1</sub> versus bacterial toxins. *Biochemistry* (2003) **42**(50):14762–73. doi:10.1021/bi035477c
  45. Gabius H-J, Wu AM. The emerging functionality of endogenous lectins: a primer to the concept and a case study on galectins including medical implications. *Chang Gung Med J* (2006) **29**(1):37–62.
  46. Vattulainen I, Rog T. Lipid simulations: a perspective on lipids in action. *Cold Spring Harb Perspect Biol* (2011) **3**(4). doi:10.1101/cshperspect.a004655
  47. DeMarco ML, Woods RJ. Atomic-resolution conformational analysis of the GM<sub>3</sub> ganglioside in a lipid bilayer and its implications for ganglioside-protein recognition at membrane surfaces. *Glycobiology* (2009) **19**(4):344–55. doi:10.1093/glycob/cwn137
  48. May AP, Robinson RC, Vinson M, Crocker PR, Jones EY. Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at

- 1.85 Å resolution. *Mol Cell* (1998) **1**(5):719–28. doi:10.1016/S1097-2765(00)80071-4
49. Wright CS. 2.2 Å resolution structure analysis of two refined N-acetylneuraminyl-lactose – wheat germ agglutinin isolectin complexes. *J Mol Biol* (1990) **215**(4):635–51. doi:10.1016/S0022-2836(05)80174-3
  50. Frolov VA, Shnyrova AV, Zimmerberg J. Lipid polymorphisms and membrane shape. *Cold Spring Harb Perspect Biol* (2011) **3**(11):a004747. doi:10.1101/cshperspect.a004747
  51. Coskun Ü, Simons K. Membrane rafting: from apical sorting to phase segregation. *FEBS Lett* (2010) **584**(9):1685–93. doi:10.1016/j.febslet.2009.12.043
  52. Taniguchi M, Wakabayashi S. Shared antigenic determinant expressed on various mammalian melanoma cells. *Gann* (1984) **75**(5):418–26.
  53. Hirabayashi Y, Hamaoka A, Matsumoto M, Matsubara T, Tagawa M, Wakabayashi S, et al. Syngeneic monoclonal antibody against melanoma antigen with interspecies cross-reactivity recognizes GM<sub>3</sub>, a prominent ganglioside of B16 melanoma. *J Biol Chem* (1985) **260**(24):13328–33.
  54. Nores GA, Dohi T, Taniguchi M, Hakomori S-i. Density-dependent recognition of cell surface GM<sub>3</sub> by a certain anti-melanoma antibody, and GM<sub>3</sub> lactone as a possible immunogen: requirements for tumor-associated antigen and immunogen. *J Immunol* (1987) **139**(9):3171–6.
  55. Hakomori S-i, Nudelman E, Levery S, Solter D, Knowles BB. The hapten structure of a developmentally regulated glycolipid antigen (SSEA-1) isolated from human erythrocytes and adenocarcinoma: a preliminary note. *Biochem Biophys Res Commun* (1981) **100**(4):1578–86. doi:10.1016/0006-291X(81)90699-9
  56. Yahi N, Aulas A, Fantini J. How cholesterol constrains glycolipid conformation for optimal recognition of Alzheimer's β amyloid peptide (Aβ<sub>1–40</sub>). *PLoS One* (2010) **5**(2):e9079. doi:10.1371/journal.pone.0009079
  57. Lingwood D, Binnington B, Róg T, Vattulainen I, Grzybek M, Coskun Ü, et al. Cholesterol modulates glycolipid conformation and receptor activity. *Nat Chem Biol* (2011) **7**(5):260–2. doi:10.1038/nchembio.551
  58. Nyholm P-G, Pascher I, Sundell S. The effect of hydrogen bonds on the conformation of glycosphingolipids. Methylated and unmethylated cerebroside studied by X-ray single crystal analysis and model calculations. *Chem Phys Lipids* (1990) **52**(1):1–10. doi:10.1016/0009-3084(90)90002-9
  59. Sandvig K, Bergan J, Kavaliuskiene S, Skotland T. Lipid requirements for entry of protein toxins into cells. *Prog Lipid Res* (2014) **54**C:1–13. doi:10.1016/j.plipres.2014.01.001
  60. Pascher I. Molecular arrangements in sphingolipids. Conformation and hydrogen bonding of ceramide and their implication on membrane stability and permeability. *Biochim Biophys Acta* (1976) **455**(2):433–51. doi:10.1016/0005-2736(76)90316-3
  61. Walker F, Abramowitz L, Benabderrahmane D, Duval X, Descatoire V, Hénin D, et al. Growth factor receptor expression in anal squamous lesions: modifications associated with oncogenic human papillomavirus and human immunodeficiency virus. *Hum Pathol* (2009) **40**(11):1517–27. doi:10.1016/j.humpath.2009.05.010
  62. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* (2004) **350**(21):2129–39. doi:10.1056/NEJMoa040938
  63. Kuan C-T, Wikstrand CJ, Bigner DD. EGF mutant receptor vIII as a molecular target in cancer therapy. *Endocr Relat Cancer* (2001) **8**(2):83–96. doi:10.1677/erc.0.0080083
  64. Ferrara N, Kerbel RS. Angiogenesis as a therapeutic target. *Nature* (2005) **438**(7070):967–74. doi:10.1038/nature04483
  65. Kawashima N, Yoon S-J, Itoh K, Nakayama K-i. Tyrosine kinase activity of epidermal growth factor receptor is regulated by GM3 binding through carbohydrate to carbohydrate interactions. *J Biol Chem* (2009) **284**(10):6147–55. doi:10.1074/jbc.M808171200
  66. Coskun Ü, Grzybek M, Drechsel D, Simons K. Regulation of human EGF receptor by lipids. *Proc Natl Acad Sci U S A* (2011) **108**(22):9044–8. doi:10.1073/pnas.1105666108
  67. Yoon S-J, Nakayama K-i, Hikita T, Handa K, Hakomori S-i. Epidermal growth factor receptor tyrosine kinase is modulated by GM3 interaction with N-linked GlcNAc termini of the receptor. *Proc Natl Acad Sci U S A* (2006) **103**(50):18987–91. doi:10.1073/pnas.0609281103
  68. Mirkin BL, Clark SH, Zhang C. Inhibition of human neuroblastoma cell proliferation and EGF receptor phosphorylation by gangliosides GM<sub>1</sub>, GM<sub>3</sub>, GD<sub>1A</sub> and GT<sub>1B</sub>. *Cell Prolif* (2002) **35**(2):105–15. doi:10.1046/j.1365-2184.2002.00228.x
  69. Hofman EG, Bader AN, Gerritsen HC, van Bergen en Henegouwen PMP. EGF induces rapid reorganization of plasma membrane microdomains. *Commun Integr Biol* (2009) **2**(3):213–4. doi:10.4161/cib.2.3.7877
  70. Miljan EA, Meuillet EJ, Mania-Farnell B, George D, Yamamoto H, Simon H-G, et al. Interaction of the extracellular domain of the epidermal growth factor receptor with gangliosides. *J Biol Chem* (2002) **277**(12):10108–13. doi:10.1074/jbc.M111669200
  71. Zurita AR, Crespo PM, Koritschoner NP, Daniotti JL. Membrane distribution of epidermal growth factor receptors in cells expressing different gangliosides. *Eur J Biochem* (2004) **271**(12):2428–37. doi:10.1111/j.1432-1033.2004.04165.x
  72. Wang J, Yu RK. Interaction of ganglioside GD3 with an EGF receptor sustains the self-renewal ability of mouse neural stem cells in vitro. *Proc Natl Acad Sci U S A* (2013) **110**(47):19137–42. doi:10.1073/pnas.1307224110
  73. Liu Y, Li R, Ladisch S. Exogenous ganglioside GD<sub>1A</sub> enhances epidermal growth factor receptor binding and dimerization. *J Biol Chem* (2004) **279**(35):36481–9. doi:10.1074/jbc.M402880200
  74. Chung T-W, Kim S-J, Choi H-J, Kim K-J, Kim M-J, Kim S-H, et al. Ganglioside GM3 inhibits VEGF/VEGFR-2-mediated angiogenesis: direct interaction of GM3 with VEGFR-2. *Glycobiology* (2009) **19**(3):229–39. doi:10.1093/glycob/cwn114
  75. Mukherjee P, Faber AC, Shelton LM, Baek RC, Chiles TC, Seyfried TN. *Thematic Review Series: Sphingolipids*. Ganglioside GM3 suppresses the proangiogenic effects of vascular endothelial growth factor and ganglioside GD<sub>1A</sub>. *J Lipid Res* (2008) **49**(5):929–38.
  76. Liu Y, McCarthy J, Ladisch S. Membrane ganglioside enrichment lowers the threshold for vascular endothelial cell angiogenic signaling. *Cancer Res* (2006) **66**(21):10408–14. doi:10.1158/0008-5472.CAN-06-1572
  77. Zeng G, Gao L, Birklé S, Yu RK. Suppression of ganglioside GD3 expression in a rat F-11 tumor cell line reduces tumor growth, angiogenesis, and vascular endothelial growth factor production. *Cancer Res* (2000) **60**(23):6670–6.
  78. Downward J, Parker P, Waterfield MD. Autophosphorylation sites on the epidermal growth factor receptor. *Nature* (1984) **311**(5985):483–5. doi:10.1038/311483a0
  79. Oda K, Matsuoka Y, Funahashi A, Kitano H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol* (2005) **1**(1):1–17. doi:10.1038/msb4100014
  80. Endres NF, Engel K, Das R, Kovacs E, Kuriyan J. Regulation of the catalytic activity of the EGF receptor. *Curr Opin Struct Biol* (2011) **21**(6):777–84. doi:10.1016/j.sbi.2011.07.007
  81. Kabayama K, Sato T, Saito K, Loberto N, Prinetti A, Sonnino S, et al. Dissociation of the insulin receptor and caveolin-1 complex by ganglioside GM3 in the state of insulin resistance. *Proc Natl Acad Sci U S A* (2007) **104**(34):13678–83. doi:10.1073/pnas.0703650104
  82. Nojiri H, Stroud M, Hakomori S-i. A specific type of ganglioside as a modulator of insulin-dependent cell growth and insulin receptor tyrosine kinase activity. Possible association of ganglioside-induced inhibition of insulin receptor function and monocytic differentiation induction in HL-60 cells. *J Biol Chem* (1991) **266**(7):4531–7.
  83. Wang X-Q, Sun P, Paller AS. Ganglioside GM3 blocks the activation of epidermal growth factor receptor induced by integrin at specific tyrosine sites. *J Biol Chem* (2003) **278**(49):48770–8. doi:10.1074/jbc.M308818200
  84. Yoon S-J, Nakayama K-i, Takahashi N, Yagi H, Utkina N, Wang HY, et al. Interaction of N-linked glycans, having multivalent GlcNAc termini, with GM3 ganglioside. *Glycoconj J* (2006) **23**(9):639–49. doi:10.1007/s10719-006-9001-4
  85. Handa K, Hakomori S-i. Carbohydrate to carbohydrate interaction in development process and cancer progression. *Glycoconj J* (2012) **29**(8–9):627–37. doi:10.1007/s10719-012-9380-7
  86. Arkhipov A, Shan Y, Das R, Endres NF, Eastwood MP, Wemmer DE, et al. Architecture and membrane interactions of the EGF receptor. *Cell* (2013) **152**(3):557–69. doi:10.1016/j.cell.2012.12.030
  87. Sezgin E, Levental I, Grzybek M, Schwarzmann G, Mueller V, Honigsmann A, et al. Partitioning, diffusion, and ligand binding of raft lipid analogs in model and cellular plasma membranes. *Biochim Biophys Acta* (2012) **1818**(7):1777–84. doi:10.1016/j.bbmem.2012.03.007

88. Hakomori S-i. Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer Res* (1996) **56**(23):5309–18.
89. Ravindranath MH, Tsuchida T, Morton DL, Irie RF. Ganglioside GM3:GD3 ratio as an index for the management of melanoma. *Cancer* (1991) **67**(12):3029–35. doi:10.1002/1097-0142(19910615)67:12<3029::AID-CNCR2820671217>3.0.CO;2-8
90. Carubia JM, Yu RK, Macala LJ, Kirkwood JM, Varga JM. Gangliosides of normal and neoplastic human melanocytes. *Biochem Biophys Res Commun* (1984) **120**(2):500–4. doi:10.1016/0006-291X(84)91282-8
91. Merzak A, Koochekpour S, Pilkington GJ. Cell surface gangliosides are involved in the control of human glioma cell invasion in vitro. *Neurosci Lett* (1994) **177**(1–2):44–6. doi:10.1016/0304-3940(94)90040-X
92. Nicoll G, Avril T, Lock K, Furukawa K, Bovin N, Crocker PR. Ganglioside GD3 expression on target cells can modulate NK cell cytotoxicity via siglec-7-dependent and -independent mechanisms. *Eur J Immunol* (2003) **33**(6):1642–8. doi:10.1002/eji.200323693
93. Ando I, Hoon DS, Suzuki Y, Saxton RE, Golub SH, Irie RF. Ganglioside GM2 on the K562 cell line is recognized as a target structure by human natural killer cells. *Int J Cancer* (1987) **40**(1):12–7. doi:10.1002/ijc.2910400104
94. Biswas K, Richmond A, Rayman P, Biswas S, Thornton M, Sa G, et al. GM2 expression in renal cell carcinoma: potential role in tumor-induced T-cell dysfunction. *Cancer Res* (2006) **66**(13):6816–25. doi:10.1158/0008-5472.CAN-06-0250
95. Grayson G, Ladisch S. Immunosuppression by human gangliosides. II. Carbohydrate structure and inhibition of human NK activity. *Cell Immunol* (1992) **139**(1):18–29. doi:10.1016/0008-8749(92)90096-8
96. Morton DL, Barth A. Vaccine therapy for malignant melanoma. *CA Cancer J Clin* (1996) **46**(4):225–44. doi:10.3322/canjclin.46.4.225
97. Pukel CS, Lloyd KO, Travassos LR, Dippold WG, Oettgen HF, Old LJ. GD<sub>3</sub>, a prominent ganglioside of human melanoma. Detection and characterization by mouse monoclonal antibody. *J Exp Med* (1982) **155**(4):1133–47. doi:10.1084/jem.155.4.1133
98. van Cruijssen H, Ruiz MG, van der Valk P, de Gruijl TD, Giaccone G. Tissue micro array analysis of ganglioside N-glycolyl GM3 expression and signal transducer and activator of transcription (STAT)-3 activation in relation to dendritic cell infiltration and microvessel density in non-small cell lung cancer. *BMC Cancer* (2009) **9**:180. doi:10.1186/1471-2407-9-180
99. Marquina G, Waki H, Fernandez LE, Kon K, Carr A, Valiente O, et al. Gangliosides expressed in human breast cancer. *Cancer Res* (1996) **56**(22):5165–71.
100. Kudo D, Rayman P, Horton C, Cathcart MK, Bukowski RM, Thornton M, et al. Gangliosides expressed by the renal cell carcinoma cell line SK-RC-45 are involved in tumor-induced apoptosis of T cells. *Cancer Res* (2003) **63**(7):1676–83.
101. Higashi H, Hirabayashi Y, Fukui Y, Naiki M, Matsumoto M, Ueda S, et al. Characterization of N-glycolylneuraminic acid-containing gangliosides as tumor-associated Hanganutziu-Deicher antigen in human colon cancer. *Cancer Res* (1985) **45**(8):3796–802.
102. Hirabayashi Y, Higashi H, Kato S, Taniguchi M, Matsumoto M. Occurrence of tumor-associated ganglioside antigens with Hanganutziu-Deicher antigenic activity on human melanomas. *Jpn J Cancer Res* (1987) **78**(6):614–20.
103. Scursoni AM, Galluzzo L, Camarero S, Lopez J, Lubieniecki F, Sampor C, et al. Detection of N-Glycolyl GM3 ganglioside in neuroectodermal tumors by immunohistochemistry: an attractive vaccine target for aggressive pediatric cancer. *Clin Dev Immunol* (2011) **2011**:245181. doi:10.1155/2011/245181
104. Scursoni AM, Galluzzo L, Camarero S, Pozzo N, Gabri MR, de Acosta CM, et al. Detection and characterization of N-glycosylated gangliosides in Wilms tumor by immunohistochemistry. *Pediatr Dev Pathol* (2009) **13**(1):18–23. doi:10.2350/08-10-0544.1
105. Hettmer S, Ladisch S, Kaucic K. Low complex ganglioside expression characterizes human neuroblastoma cell lines. *Cancer Lett* (2005) **225**(1):141–9. doi:10.1016/j.canlet.2004.11.036
106. Brezicka F-T, Olling S, Nilsson O, Bergh J, Holmgren J, Sörensen S, et al. Immunohistological detection of fucosyl-G<sub>M1</sub> ganglioside in human lung cancer and normal tissues with monoclonal antibodies. *Cancer Res* (1989) **49**(5):1300–5.
107. Okada M, Furukawa K, Yamashiro S, Yamada Y, Haraguchi M, Horibe K, et al. High expression of ganglioside  $\alpha$ -2,8-Sialyltransferase (GD3 Synthase) gene in adult T-cell leukemia cells unrelated to the gene expression of human T-lymphotropic virus type I. *Cancer Res* (1996) **56**(12):2844–8.
108. Cheung N-K, Saarinen UM, Neely JE, Landmeier B, Donovan D, Coccia PF. Monoclonal antibodies to a glycolipid antigen on human neuroblastoma cells. *Cancer Res* (1985) **45**(6):2642–9.
109. Mujoo K, Cheresch DA, Yang HM, Reisfeld RA. Disialoganglioside GD<sub>2</sub> on human neuroblastoma cells: target antigen for monoclonal antibody-mediated cytotoxicity and suppression of tumor growth. *Cancer Res* (1987) **47**(4):1098–104.
110. Steenackers A, Vanbeselaere J, Cazet A, Bobowski M, Rombouts Y, Colomb F, et al. Accumulation of unusual gangliosides GQ<sub>3</sub> and GP<sub>3</sub> in breast cancer cells expressing the GD<sub>3</sub> synthase. *Molecules* (2012) **17**(8):9559–72. doi:10.3390/molecules17089559
111. Bobowski M, Vincent A, Steenackers A, Colomb F, Van Seuningen I, Julien S, et al. Estradiol represses the GD<sub>3</sub> synthase gene ST8SIA1 expression in human breast cancer cells by preventing NF $\kappa$ B binding to ST8SIA1 promoter. *PLoS One* (2013) **8**(4):e62559. doi:10.1371/journal.pone.0062559
112. Rusnati M, Tanghetti E, Urbinati C, Tulipano G, Marchesini S, Ziche M, et al. Interaction of fibroblast growth factor-2 (FGF-2) with free gangliosides: biochemical characterization and biological consequences in endothelial cell cultures. *Mol Biol Cell* (1999) **10**(2):313–27. doi:10.1091/mbc.10.2.313
113. Chang F, Li R, Ladisch S. Shedding of gangliosides by human medulloblastoma cells. *Exp Cell Res* (1997) **234**(2):341–6. doi:10.1006/excr.1997.3619
114. Li RX, Ladisch S. Shedding of human neuroblastoma gangliosides. *Biochim Biophys Acta* (1991) **1083**(1):57–64. doi:10.1016/0005-2760(91)90124-Z
115. Koochekpour S, Merzak A, Pilkington GJ. Vascular endothelial growth factor production is stimulated by gangliosides and TGF- $\beta$  isoforms in human glioma cells in vitro. *Cancer Lett* (1996) **102**(1–2):209–15. doi:10.1016/0304-3835(96)04161-4
116. Malykh YN, Schauer R, Shaw L. N-glycolylneuraminic acid in human tumours. *Biochimie* (2001) **83**(7):623–34. doi:10.1016/S0300-9084(01)01303-7
117. Chou H-H, Takematsu H, Diaz S, Iber J, Nickerson E, Wright KL, et al. A mutation in human CMP-sialic acid hydroxylase occurred after the *Homo-Pan* divergence. *Proc Natl Acad Sci U S A* (1998) **95**(20):11751–6. doi:10.1073/pnas.95.20.11751
118. Irie A, Koyama S, Kozutsumi Y, Kawasaki T, Suzuki A. The molecular basis for the absence of N-glycolylneuraminic acid in humans. *J Biol Chem* (1998) **273**(25):15866–71. doi:10.1074/jbc.273.25.15866
119. Hirabayashi Y, Kasakura H, Matsumoto M, Higashi H, Kato S, Kasai N, et al. Specific expression of unusual GM2 ganglioside with Hanganutziu-Deicher antigen activity on human colon cancers. *Jpn J Cancer Res* (1987) **78**(3):251–60.
120. Kawachi S, Saida T. Analysis of the expression of Hanganutziu-Deicher (HD) antigen in human malignant melanoma. *J Dermatol* (1992) **19**(11):827–30.
121. Bardor M, Nguyen DH, Diaz S, Varki A. Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells. *J Biol Chem* (2005) **280**(6):4228–37. doi:10.1074/jbc.M412040200
122. Tangvoranuntakul P, Gagneux P, Diaz S, Bardor M, Varki N, Varki A, et al. Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc Natl Acad Sci U S A* (2003) **100**(21):12045–50. doi:10.1073/pnas.2131556100
123. Yin J, Hashimoto A, Izawa M, Miyazaki K, Chen G-Y, Takematsu H, et al. Hypoxic culture induces expression of sialin, a sialic acid transporter, and cancer-associated gangliosides containing non-human sialic acid on human cancer cells. *Cancer Res* (2006) **66**(6):2937–45. doi:10.1158/0008-5472.CAN-05-2615
124. Yin J, Miyazaki K, Shaner RL, Merrill AH Jr., Kannagi R. Altered sphingolipid metabolism induced by tumor hypoxia – new vistas in glycolipid tumor markers. *FEBS Lett* (2010) **584**(9):1872–8. doi:10.1016/j.febslet.2009.11.019
125. Hayashi N, Chiba H, Kuronuma K, Go S, Hasegawa Y, Takahashi M, et al. Detection of N-glycosylated gangliosides in non-small-cell lung cancer using GMR8 monoclonal antibody. *Cancer Sci* (2013) **104**(1):43–7. doi:10.1111/cas.12027
126. Casadesús AV, Fernández-Marrero Y, Clavell M, Gómez JA, Hernández T, Moreno E, et al. A shift from N-glycolyl- to N-acetyl-sialic acid in the GM3 ganglioside impairs tumor development in mouse lymphocytic leukemia cells. *Glycoconj J* (2013) **30**(7):687–99. doi:10.1007/s10719-013-9473-y
127. Higashi H, Naiki M, Matuo S, Okouchi K. Antigen of “serum sickness” type of heterophile antibodies in human sera: identification as gangliosides with N-glycolylneuraminic acid. *Biochem Biophys Res Commun* (1977) **79**(2):388–95. doi:10.1016/0006-291X(77)90169-3
128. Merrick JM, Zadarlik K, Milgrom F. Characterization of the Hanganutziu-Deicher (serum-sickness) antigen as gangliosides containing

- n-glycolylneuraminic acid. *Int Arch Allergy Appl Immunol* (1978) **57**(5):477–80. doi:10.1159/000232140
129. Hanganutziu M. Hémagglutinines hétérogénétiques après injection de serum de cheval. *C R Seances Soc Biol Fil* (1924) **91**:1457.
  130. Deicher H. Über die Erzeugung heterospezifischer Hämagglutinine durch Injektion artfremden Serums. *Z Hyg Infektionskr* (1926) **106**:561. doi:10.1007/BF02176298
  131. Kasukawa R, Kano K, Bloom ML, Milgrom F. Heterophile antibodies in pathologic human sera resembling antibodies stimulated by foreign species sera. *Clin Exp Immunol* (1976) **25**(1):122–32.
  132. Zhu A, Hurst R. Anti-N-glycolylneuraminic acid antibodies identified in healthy human serum. *Xenotransplantation* (2002) **9**(6):376–81. doi:10.1034/j.1399-3089.2002.02138.x
  133. Ravindranath MH, Yesowitch P, Sumobay C, Morton DL. Glycoimmunomics of human cancer: current concepts and future perspectives. *Future Oncol* (2007) **3**(2):201–14. doi:10.2217/14796694.3.2.201
  134. Cojocaru M, Silosi I. The significance of natural autoantibodies. *Maedica (Buchar)* (2009) **4**(1):22–5.
  135. Chapman CJ, Murray A, McElveen JE, Sahin U, Luxemburger U, Türeci Ö, et al. Autoantibodies in lung cancer: possibilities for early detection and subsequent cure. *Thorax* (2008) **63**(3):228–33. doi:10.1136/thx.2007.083592
  136. Zhang J-Y, Casiano CA, Peng X-X, Koziol JA, Chan EKL, Tan EM. Enhancement of antibody detection in cancer using panel of recombinant tumor-associated antigens. *Cancer Epidemiol Biomarkers Prev* (2003) **12**(2):136–43.
  137. Storr SJ, Chakrabarti J, Barnes A, Murray A, Chapman CJ, Robertson JF. Use of autoantibodies in breast cancer screening and diagnosis. *Expert Rev Anticancer Ther* (2006) **6**(8):1215–23. doi:10.1586/14737140.6.8.1215
  138. Rodríguez-Zhurbenko N, Martínez D, Blanco R, Rondón T, Griñán T, Hernández AM. Human antibodies reactive to NeuGcGM3 ganglioside have cytotoxic antitumor properties. *Eur J Immunol* (2013) **43**:826–37. doi:10.1002/eji.201242693
  139. Couzin-Frankel J. Breakthrough of the year 2013. Cancer immunotherapy. *Science* (2013) **342**(6165):1432–3. doi:10.1126/science.342.6165.1432
  140. Schuster M, Nechansky A, Loibner H, Kircheis R. Cancer immunotherapy. *Biotechnol J* (2006) **1**(2):138–47. doi:10.1002/biot.200500044
  141. Cheever MA, Allison JP, Ferris AS, Finn OJ, Hastings BM, Hecht TT, et al. The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research. *Clin Cancer Res* (2009) **15**(17):5323–37. doi:10.1158/1078-0432.CCR-09-0737
  142. Modak S, Cheung NK. Disialoganglioside directed immunotherapy of neuroblastoma. *Cancer Invest* (2007) **25**(1):67–77. doi:10.1080/07357900601130763
  143. Krengel U, Olsson L-L, Martínez C, Talavera A, Rojas G, Mier E, et al. Structure and molecular interactions of a unique antitumor antibody specific for N-glycolyl GM3. *J Biol Chem* (2004) **279**(7):5597–603. doi:10.1074/jbc.M311693200
  144. Talavera A, Eriksson A, Ökvist M, López-Requena A, Fernández-Marrero Y, Pérez R, et al. Crystal structure of an anti-ganglioside antibody, and modeling of the functional mimicry of its NeuGc-GM3 antigen by an anti-idiotypic antibody. *Mol Immunol* (2009) **46**(16):3466–75. doi:10.1016/j.molimm.2009.07.032
  145. Agostino M, Yuriev E, Ramsland PA. Antibody recognition of cancer-related gangliosides and their mimics investigated using *in silico* site mapping. *PLoS One* (2012) **7**(4):e35457. doi:10.1371/journal.pone.0035457
  146. Rojas G, Pupo A, Gómez S, Krengel U, Moreno E. Engineering the binding site of an antibody against N-Glycolyl GM3: from functional mapping to novel anti-ganglioside specificities. *ACS Chem Biol* (2013) **8**(2):376–86. doi:10.1021/cb3003754
  147. Gajdosik Z. Racotumomab – a novel anti-idiotypic monoclonal antibody vaccine for the treatment of cancer. *Drugs Today (Barc)* (2014) **4**:301–7. doi:10.1358/dot.2014.50.4.2116670
  148. Osorio M, Gracia E, Reigosa E, Hernandez J, de la Torre A, Saurez G, et al. Effect of vaccination with N-glycolyl GM3/VSSP vaccine by subcutaneous injection in patients with advanced cutaneous melanoma. *Cancer Manag Res* (2012) **4**:341–5. doi:10.2147/CMAR.S226174
  149. Alfonso S, Valdés-Zayas A, Santiesteban ER, Flores YI, Areces F, Hernández M, et al. A randomized, multicenter, placebo-controlled clinical trial of racotumomab-alum vaccine as switch maintenance therapy in advanced non-small-cell-lung cancer patients. *Clin Cancer Res* (2014) **20**(14):1–12. doi:10.1158/1078-0432.CCR-13-1674
  150. Eggermont AM, Suci S, Ruka W, Marsden J, Testori A, Corrie P, et al. Post-operative adjuvant ganglioside GM2-KLH21 vaccination treatment vs observation in stage II (T3-T4N0M0) melanoma: 2nd interim analysis led to an early disclosure of the results. *J Clin Oncol* (2008) **26**.
  151. Eggermont AM, Suci S, Rutkowski P, Marsden J, Santinami M, Corrie P, et al. Adjuvant ganglioside GM2-KLH/QS-21 vaccination versus observation after resection of primary tumor >1.5 mm in patients with stage II melanoma: results of the EORTC 18961 randomized phase III trial. *J Clin Oncol* (2013) **31**:3831–7. doi:10.1200/JCO.2012.47.9303
  152. Herrera ZM, Ramos TC. Pilot study of a novel combination of two therapeutic vaccines in advanced non-small-cell lung cancer patients. *Cancer Immunol Immunother* (2014) **63**(7):737–47. doi:10.1007/s00262-014-1552-9
  153. Tessier MB, DeMarco ML, Yongye AB, Woods RJ. Extension of the GLYCAM06 biomolecular force field to lipids, lipid bilayers and glycolipids. *Mol Simul* (2008) **34**(4):349–63. doi:10.1080/08927020701710890
  154. Czogalla A, Grzybek M, Jones W, Coskun Ü. Validity and applicability of membrane model systems for studying interactions of peripheral membrane proteins with lipids. *Biochim Biophys Acta* (2014) **1841**(8):1049–59. doi:10.1016/j.bbalip.2013.12.012

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 May 2014; accepted: 27 June 2014; published online: 21 July 2014.

Citation: Krengel U and Bousquet PA (2014) Molecular recognition of gangliosides and their potential for cancer immunotherapies. *Front. Immunol.* **5**:325. doi:10.3389/fimmu.2014.00325

This article was submitted to *Immunotherapies and Vaccines*, a section of the journal *Frontiers in Immunology*.

Copyright © 2014 Krengel and Bousquet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Structure based refinement of a humanized monoclonal antibody that targets tumor antigen disialoganglioside GD2

Mahiuddin Ahmed, Jian Hu and Nai-Kong V. Cheung\*

Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

## Edited by:

Mark Agostino, Curtin University, Australia

## Reviewed by:

Paul A. Ramsland, Burnet Institute, Australia

An-Suei Yang, Academia Sinica, Taiwan

## \*Correspondence:

Nai-Kong V. Cheung, Department of Pediatrics, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA  
e-mail: cheungn@mskcc.org

Disialoganglioside GD2 is an important target on several pediatric and adult cancer types including neuroblastoma, retinoblastoma, melanoma, small-cell lung cancer, brain tumors, sarcomas, and cancer stem cells. We have utilized structural and computational methods to refine the framework of humanized monoclonal antibody 3F8, the highest affinity anti-GD2 antibody in clinical development. Two constructs (V3 and V5) were designed to enhance stability and minimize potential immunogenicity. Construct V3 contained 12 point mutations and had higher thermal stability and comparable affinity and *in vitro* tumor cells killing as the parental hu3F8. Construct V5 had nine point mutations to minimize potential immunogenicity, but resulted in weaker thermal stability, weaker antigen binding, and reduced tumor killing potency. When construct V3 was combined with the single point mutation HC:G54I, the resulting V3-Ile construct had enhanced stability, antigen binding, and a nearly sixfold increase in tumor cell killing. The resulting product is a lead candidate for clinical development for the treatment of GD2-positive tumors.

**Keywords:** antibody engineering, ganglioside, neuroblastoma, melanoma, structure, computational chemistry

## INTRODUCTION

GD2 is a ganglioside expressed in several pediatric and adult cancer types and has been actively targeted by cancer immunotherapy approaches (see Ref. (1) for recent review). GD2 is a member of the b-series gangliosides, which are normally expressed during fetal development and are highly restricted to the central nervous system in healthy adults, with low levels of expression on peripheral nerves and skin melanocytes (2). GD2 has been found to be expressed in neuroectoderm-derived tumors and sarcomas, including neuroblastoma, retinoblastoma, melanoma, small-cell lung cancer, brain tumors, and sarcomas (3–5). Recent evidence has also shown that GD2 can be found on breast cancer stem cells (6, 7), as well as on neuroectodermal (8) and mesenchymal stem cells (9, 10).

Because of its surface expression on tumor cells and restricted normal expression in the brain and low levels in the periphery, GD2 has been an ideal target for the development of monoclonal antibodies (MoAbs), which cannot cross the blood–brain barrier. Several anti-GD2 antibodies have been developed and tested in the clinic over the past 20 years, primarily in pediatric neuroblastoma patients. 3F8 was the first anti-GD2 MoAb to be tested in patients with neuroblastoma (3, 11, 12). MoAb 3F8 is a murine IgG3 with the highest reported affinity for GD2 ( $K_D = 5$  nM) (13). It binds specifically to the pentasaccharide epitope on GD2. Phase II clinical data have demonstrated that 3F8 when combined with the cytokine GM-CSF can significantly improve the survival of high-risk stage 4 children with metastatic neuroblastoma (14). Murine 3F8 was more recently humanized (hu3F8) based on complementarity determining region (CDR) grafting (13), and is currently in Phase

I clinical trials (clinical trials.gov NCT01419834, NCT01757626, and NCT01662804).

We have previously solved the crystal structure of murine 3F8 to 1.65 Å resolution (protein data bank 3VFG) and used completely *in silico* methods to find a single point mutation (HC:G54I) that could significantly enhance the antibody-dependent cell-mediated cytotoxicity (ADCC) of hu3F8 (15). Based on computational modeling, we have developed two additional hu3F8 frameworks, named V3 and V5, which were designed to optimize the properties of hu3F8. More specifically, V3 was designed to maximize stability and V5 was designed to minimize potential immunogenicity. We present here the computational methods used to derive the hu3F8 V3 and V5 frameworks along with their experimental properties of antigen binding, thermal stability, and *in vitro* ADCC.

## MATERIALS AND METHODS

### MOLECULAR MODELING

Molecular modeling, energy calculations, and image renderings were done using Discovery Studio 4.0 (Accelrys, San Diego, CA, USA). The crystal structure of m3F8 Fab (pdb 3VFG) and the homology model of hu3F8 Fab were simulated using CHARMM (CHemistry at Harvard Molecular mechanics) force fields, and the effects of point mutations were calculated from the difference between the folding free energies of the mutated structure and the parental protein. Generalized Born approximation was used to account for the effect of the solvent and all electrostatic terms were calculated as a sum of coulombic interactions and polar contributions to the solvation energy. A weighted sum of the van der Waals, electrostatic, entropy, and non-polar terms was calculated for each point mutation.



### CONSTRUCTION AND EXPRESSION OF hu3F8 CONSTRUCTS

Humanized 3F8 genes were synthesized for CHO cells (Blue Heron Biotechnology or Genscript) as previously described (13). Using the bluescript vector, these heavy and light chain genes of hu3F8 were transfected into DG44 cells and selected with G418 (Invitrogen, CA, USA). Hu3F8 producer lines were cultured in Optich serum free medium (Invitrogen) and the mature supernatant was harvested as previously described (13). Protein A affinity column was pre-equilibrated with 25 mM sodium citrate buffer with 0.15 M NaCl, pH 8.2. Bound hu3F8 was eluted with 0.1 M citric acid/sodium citrate buffer, pH 3.9 and alkalized (1:10 v/v ratio) in 25 mM sodium citrate, pH 8.5. It was passed through a Sartobind-Q membrane and concentrated to 5–10 mg/mL in 25 mM sodium citrate, 0.15 M NaCl, pH 8.2.

### THERMAL STABILITY MEASUREMENTS

The thermal stabilities of MoAbs were measured by differential scanning fluorimetry using the Protein Thermal Shift assay (Life Technologies). MoAbs (0.2 mg/mL) were mixed with Sypro Orange dye and fluorescence was monitored using a StepOne-Plus quantitative PCR machine (Applied Biosystems) with a 1% thermal gradient from 25 to 99°C. Data were analyzed using Protein Thermal Shift Software (Applied Biosystems) to calculate the  $T_m$  using the derivative method. Fab and F(ab')<sub>2</sub> preparations of hu3F8 were used to correctly assign the Fab peak for the hu3F8 samples. All samples were prepared in triplicate. Statistical significance was calculated using a student's *T* test.

### BINDING KINETICS BY SURFACE PLASMON RESONANCE

*In vitro* binding kinetics were measured using Biacore T-100 (GE Healthcare) as previously described (13). In brief, gangliosides were directly immobilized onto the CM5 sensor chip via hydrophobic interaction. Purified anti-GD2 MoAbs were diluted in HBS-E buffer containing 250 mM NaCl at increasing concentrations (50–1600 nM) prior to analysis. Samples (60  $\mu$ L) were injected over the sensor surface at a flow rate of 30  $\mu$ L/min over 2 min. Following completion of the association phase, dissociation was monitored in HBS-E buffer containing 250 mM NaCl for 300 s at the same flow rate. At the end of each cycle, the surface was regenerated using 50  $\mu$ L 20 mM NaOH at a flow rate of 50  $\mu$ L/min over 1 min and 100  $\mu$ L 4 M MgCl<sub>2</sub> at a flow rate of 50  $\mu$ L/min over 2 min. The data were analyzed by the bivalent analyte model and default parameter setting for the rate constants using the Biacore T-100 evaluation software, and the apparent association on rate constant ( $k_{on}$ ), dissociation off rate constant ( $k_{off}$ ), and equilibrium dissociation constant ( $K_D = k_{off}/k_{on}$ ) were calculated.

### ANTIBODY-DEPENDENT CELL-MEDIATED CYTOTOXICITY BY <sup>51</sup>CHROMIUM RELEASE

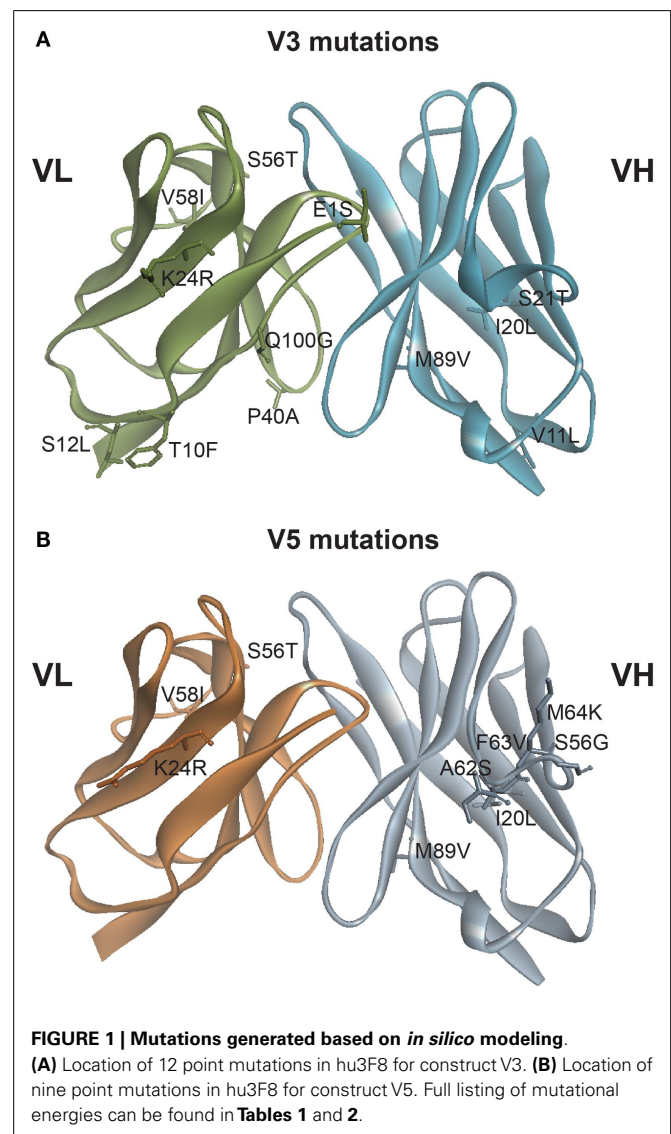
Human neuroblastoma cell line LAN-1 was provided by Dr. Robert Seeger (Children's Hospital of Los Angeles). LAN-1 cells were grown in F10 RPMI 1640 medium supplemented with 10% fetal bovine serum (Hyclone, South Logan, UT, USA), 2 mM glutamine, 100 U/mL penicillin, and 100  $\mu$ g/mL streptomycin at 37°C in a 5% CO<sub>2</sub> incubator. ADCC assays were performed using NK-92MI cells stably transfected with the human CD16 Fc receptor as previously described (13). LAN-1 target cells were detached with 2 mM

EDTA in Ca<sup>2+</sup> Mg<sup>2+</sup> free PBS and washed in F10, before radiolabeling with <sup>51</sup>Cr for ADCC assays. All samples were prepared in triplicate. Dose–response curves were fitted by non-linear regression to a sigmoidal dose–response (variable slope) model, using GraphPad Prism software, to allow for determination of EC<sub>50</sub>. For comparison of curves, best-fit values for EC<sub>50</sub> were analyzed for significance using F tests.

## RESULTS

### DESIGN OF CONSTRUCTS V3 AND V5

Constructs V3 and V5 (see Figure 1) were designed utilizing completely *in silico* methods, based on both the crystal structure of murine 3F8 Fab (pdb 3VFG) and a homology model of hu3F8 Fab that was built using MODELLER followed by CHARMm energy minimizations. The original hu3F8 that was built by CDR grafting methods utilized the human germline sequences IGHV3-33 for the heavy chain template and IGKV3-15 for the light chain template (www.imgt.org). These same templates were utilized in



deciding which mutations to incorporate into V3 and V5, in order to minimize potentially immunogenic sequences.

**Table 1** shows the 12 mutations that were incorporated into hu3F8 resulting in construct V3, along with their predicted mutational energies. *In silico* mutagenesis was done on every potential humanizing mutation in the murine 3F8 structure that was not directly predicted to be involved in antigen recognition, based on our previous 3F8:GD2 docked model (15). In addition, potential back mutations and humanizing mutations were analyzed in the homology model of hu3F8. **Table 1** shows both of these sets of calculations. In choosing which mutations to incorporate into construct V3, more emphasis was placed in the first set of calculations for stabilizing the murine 3F8 structure, since this was the experimentally verified high-resolution crystal structure, in comparison to the homology model of hu3F8, which can contain inherent error. Another consideration in placing emphasis on the native structure of murine 3F8 was the fact that murine 3F8 had consistently shown higher antigen-binding affinity than hu3F8 (13).

The analysis showed that five mutations that were made in the original hu3F8 were destabilizing, and so for construct V3, those mutations were reverted back to the murine sequence (LC:E1S, LC:T10F, LC:S12L, HC:V11L, and HC:S21T). Two additional back mutations were made (LC:P40A and LC:Q100G) because they involved Gly or Pro residues that can affect protein backbone conformation. To offset the potential immunogenicity of these seven back mutations in the V3 construct, five humanizing mutations were added (LC:K24R, LC:S56T, LC:V58I, HC:I20L, and HC:M89V), which had either enhanced stability or had a negligible effect ( $< 0.5$  kcal/mol). Two of these mutations involved mutating CDR residues (LC:K24R and LC:S56T). The net result of all 12 mutations was predicted to have a stabilizing mutational energy of  $-3.55$  kcal/mol to the murine 3F8 structure. However, this same set of mutations in the model of hu3F8 was predicted to have a destabilizing mutational energy of  $+4.61$  kcal/mol.

**Table 1 | Mutation energies associated with the design of construct V3.**

Mutation	Location	Mutation energy (kcal/mol) in m3F8 Fab	Mutation energy (kcal/mol) in hu3F8 Fab	Resulting phenotype
LC: E1S	Framework	-0.48	+0.28	Murine
LC: T10F	Framework	+0.17	-0.95	Murine
LC: S12L	Framework	-1.26	+0.79	Murine
LC: K24R	CDR L1	-0.21	-0.17	Human
LC: P40A	Framework	+0.22	+1.41	Murine
LC: S56T	CDR L2	-0.05	+0.04	Human
LC: V58I	Framework	-0.64	+0.01	Human
LC: Q100G	Framework	+1.06	+2.94	Murine
HC: V11L	Framework	-1.61	+0.99	Murine
HC: I20L	Framework	+0.43	-0.86	Human
HC: S21T	Framework	-0.68	+0.42	Murine
HC: M89V	Framework	-0.50	-0.29	Human
<b>Net result</b>		<b>-3.55</b>	<b>+4.61</b>	

**Table 2** shows the nine point mutations were incorporated into hu3F8 to make construct V5, in an effort to minimize potential immunogenicity. In addition to the five humanizing mutations from construct V3 (LC:K24R, LC:S56T, LC:V58I, HC:I20L, and HC:M89V), construct V5 also includes four additional humanizing mutations (HC:A62S, HC:F63V, HC:M64K, and HC:S65G), which are located on CDR H2. These four CDR residues were predicted to be a part of a potentially moderate affinity MHC class II T-cell epitope, which can result in enhanced immunogenicity (as identified using the NN-align method on the Immune Epitope Database (<http://www.iedb.org/>)). The net mutational energy of all nine mutations in construct V5 was predicted to be a moderately destabilizing  $+1.62$  kcal/mol for the murine 3F8 structure, and  $+0.90$  kcal/mol for the hu3F8 model.

Potential immunogenicity of constructs V3 and V5 as compared to hu3F8 was analyzed using the T20 score analyzer (16), a new *in silico* tool that can predict the “humanness” content of antibody variable regions derived from a database of  $\sim 38,700$  human antibody variable sequences. **Table 3** shows the T20 scores for hu3F8, V3, and V5. As expected, the net two additional murine mutations in V3 compared to hu3F8 resulted in slightly lower T20 scores, and the net nine humanizing mutations in V5 resulted in higher T20 scores, a characteristic of low immunogenicity MoAbs.

**Table 2 | Mutation energies associated with the design of construct V5.**

Mutation	Location	Mutation energy (kcal/mol) in m3F8 Fab	Mutation energy (kcal/mol) in hu3F8 Fab	Resulting phenotype
LC: K24R	CDR L1	-0.21	-0.17	Human
LC: S56T	CDR L2	-0.05	+0.04	Human
LC: V58I	Framework	-0.64	+0.01	Human
HC: I20L	Framework	+0.43	-0.86	Human
HC: A62S	CDR H2	-0.24	-0.50	Human
HC: F63V	CDR H2	+1.91	+1.80	Human
HC: M64K	CDR H2	+0.10	-0.23	Human
HC: S65G	CDR H2	+0.82	+1.10	Human
HC: M89V	Framework	-0.50	-0.29	Human
<b>Net result</b>		<b>+1.62</b>	<b>+0.90</b>	

**Table 3 | Humanness content based on T20 score analyzer (16).**

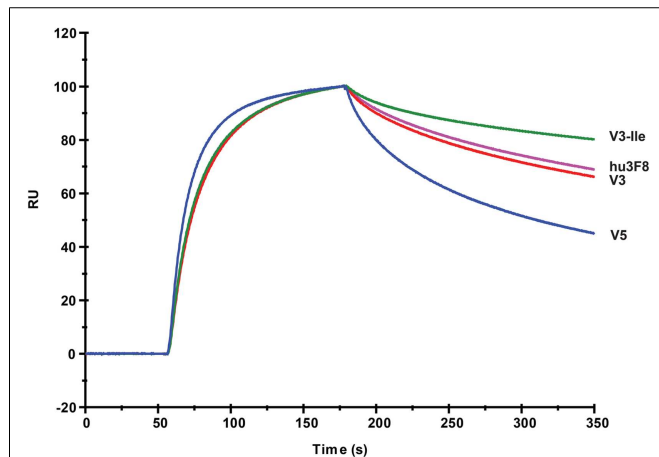
Domain	Construct	T20 (CDR + framework)	T20 (framework)
VL	hu3F8	76.0	84.8
VL	V3	74.8	80.2
VL	V5	78.9	85.9
VH	hu3F8	71.6	82.2
VH	V3	71.6	82.0
VH	V5	76.4	84.3

Scale is on the order of 0–100, with 100 being the most human in sequence.

**Table 4 | Thermal stability of hu3F8 constructs.**

Construct	Fab T <sub>m</sub> (°C)	ΔT <sub>m</sub> Fab (°C)	p-Value
hu3F8	73.6 ± 0.3		
V3	75.4 ± 0.5	1.8	0.006
V3-Ile	75.1 ± 0.1	1.5	0.001
V5	64.5 ± 0.3	−9.1	< 0.001

Samples were prepared in triplicate and measured by differential scanning fluorimetry. Values are shown as mean ± standard deviation.



**FIGURE 2 | Composite surface plasmon resonance sensorgram of the binding hu3F8, V3, V3-Ile, and V5 to ganglioside GD2.** Analysis of binding kinetics is shown in Table 5. Full sensorgrams are presented in Figure S1 in Supplementary Material.

### THERMAL STABILITY

The thermal stability of hu3F8, V3, and V5 was measured using differential scanning fluorimetry (see Table 4). The Fab domain of construct V3, which was designed to be more stable, had a nearly 2°C increase in T<sub>m</sub> compared to hu3F8 ( $p = 0.006$ ). Construct V5, on the other hand, had substantially lower thermal stability (9°C lower T<sub>m</sub> than hu3F8). Based on the enhanced stability, V3 was chosen as a lead candidate, and the HC:G54I mutation, which we had previously shown to enhance tumor cell killing, was incorporated to make construct V3-Ile. The measured thermal stability of V3-Ile was nearly identical to V3.

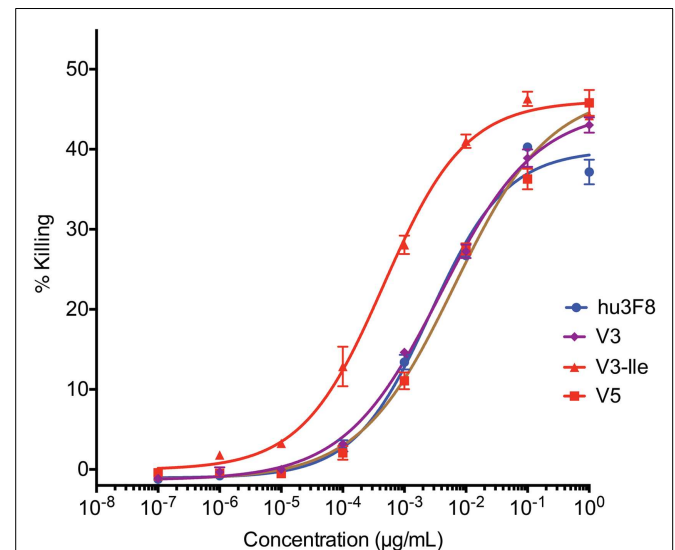
### ANTIGEN BINDING KINETICS

The GD2 binding kinetics of hu3F8, V3, V3-Ile, and V5 were measured by surface plasmon resonance (see Figure 2 for normalized composite sensorgram, Figure S1 in Supplementary Material for complete sensorgrams, and Table 5 for analysis). Construct V3 (11.5 nM  $K_D$ ) had similar binding properties to hu3F8 (9.1 nM  $K_D$ ). Construct V5, on the other hand, had an almost twofold loss in binding (19.1 nM  $K_D$ ), which may have resulted from the additional CDR mutations and/or the weakened thermal stability. Construct V3-Ile had the highest GD2 affinity (3.7 nM  $K_D$ ). Interestingly, this enhancement in affinity is higher than what we had previously observed with the HC:G54I mutation in the parental hu3F8 framework.

**Table 5 | Analysis of binding kinetics measured by surface plasmon resonance.**

Construct	$K_{on}$ (S <sup>-1</sup> M <sup>-1</sup> )	$K_{off}$ (S <sup>-1</sup> )	$K_D$ (nM)
hu3F8	$1.15 \times 10^5$	$1.04 \times 10^{-3}$	9.1
V3	$1.09 \times 10^5$	$1.25 \times 10^{-3}$	11.5
V3-Ile	$1.28 \times 10^5$	$0.48 \times 10^{-3}$	3.7
V5	$1.73 \times 10^5$	$3.30 \times 10^{-3}$	19.1

$K_{on}$  and  $K_{off}$  were determined from individual sensorgrams, shown in Figure S1 in Supplementary Material.



**FIGURE 3 | In vitro antibody-dependent cell-mediated cytotoxicity assay of human neuroblastoma LAN-1 target cells.** Antibodies were applied to LAN-1 cells in the presence of NK-92MI cells stably transfected with the human CD16 Fc receptor, at an effector:target ratio of 20:1. Samples were prepared in triplicate, and cytotoxicity was measured by <sup>51</sup>chromium release. Values are shown as mean ± standard error.

### IN VITRO ANTIBODY-DEPENDENT CELL-MEDIATED CYTOTOXICITY

Antibody-dependent cell-mediated cytotoxicity assays were done to test the effectiveness of hu3F8, V3, V3-Ile, and V5 on human neuroblastoma LAN-1 cells (see Figure 3 and Table 6). Cytotoxicity of an isotype matched non-targeting control is shown in Figure S2 in Supplementary Material. Consistent with the antigen binding data, V3 had similar binding to hu3F8 ( $EC_{50}$  of  $3.83 \pm 0.51 \times 10^{-3}$  μg/mL for V3 compared to  $EC_{50}$  of  $2.61 \pm 0.48 \times 10^{-3}$  μg/mL for hu3F8,  $p = 0.1138$ ). Construct V5 had significantly weaker killing ( $EC_{50}$  of  $6.55 \pm 1.45 \times 10^{-3}$  μg/mL,  $p = 0.0025$ ). Construct V3-Ile had the highest level of killing ( $EC_{50}$  of  $0.46 \pm 0.08 \times 10^{-3}$  μg/mL), a nearly sixfold increase in killing relative to hu3F8 ( $p < 0.0001$ ) and eightfold increase over its parental V3.

### DISCUSSION

Aberrant glycosylation has long been considered to be a hallmark of cancer (17). Ganglioside markers such as GD2 have become an attractive target in recent years because of the number of tumor types and cancer stem cells that express it on their surface, as

**Table 6 | Analysis of *in vitro* antibody-dependent cell-mediated cytotoxicity of neuroblastoma LAN-1 cells.**

Construct	EC50 ( $\times 10^{-3}$ $\mu\text{g/mL}$ )	Relative potency	p-Value
hu3F8	2.61 $\pm$ 0.48	1	
V3	3.83 $\pm$ 0.51	0.7	0.1138
V3-Ile	0.46 $\pm$ 0.08	5.7	<0.0001
V5	6.55 $\pm$ 1.45	0.4	0.0025

Samples were prepared in triplicate. Values are shown as mean  $\pm$  standard error.

well as GD2's restricted expression in normal tissue. Monoclonal antibody 3F8 is a lead therapeutic candidate in this area, and its derivatives are being tested in a number of different targeting strategies including bispecific T-cell engaging antibodies, pre-targeted radio-immunotherapy, drug/toxin conjugates, nanoparticles, and even chimeric antigen receptors for use in adoptive cell therapy.

As in all antibody therapeutics, *in vivo* efficacy is affected by antigen affinity, antibody stability, immunogenicity, as well as a number of serum stability and pharmacokinetics related factors. In this study, we have investigated a structural and computational approach to refine the stability and to reduce computationally predicted immunogenicity of a humanized form of the anti-GD2 MoAb 3F8. By introducing site-specific mutations based on force-field simulations of the antibody crystal structure, we generated construct V3, which had significantly higher thermal stability, and comparable antigen binding and *in vitro* ADCC.

We have additionally attempted to minimize potential immunogenicity in designing construct V5, which had major mutations to CDR residues to eliminate a predicted T-cell epitope. Immunogenicity is a major component of clinical efficacy. A large percentage of neuroblastoma patients treated with the murine 3F8 developed a human anti-mouse antibody response, limiting repeated administrations of this antibody. In the case of designing construct V5, however, the mutations were too stringent, resulting in lowered thermal stability, weaker antigen binding, and weaker tumor cell killing. Finally, we combined our stability enhanced V3 construct with the cytotoxicity enhancing mutation (HC:G54I), which resulted in enhanced stability, antigen binding, and *in vitro* tumor cell killing, compared to the parental hu3F8.

While there are several examples of using computational methods to enhance the properties of antibodies [see Ref. (18) for review], there are few examples of using site-specific *in silico* based framework mutations to enhance thermal stability profiles. Wang and Duan (19) did suggest mutations to the VH–VL interface of anti-VEGF single-chain variable fragment (scFv) to enhance thermal stability based on molecular dynamics simulations, but with no experimental validation. We have recently shown that disulfide stabilization at the VH–VL interface of the anti-GD2 scFv 5F11 in the context of a GD2xCD3 tandem scFv bispecific antibody resulted in a 10°C increase in thermal stability and a nearly 150-fold increase in tumor killing potency (20). Enhancing thermal stability can also lead to less aggregation and less immunogenicity. Liu et al. (21) have shown that disulfide stabilization of an anti-CD22 antibody–toxin fusion protein resulted in enhanced thermal stability and less immunogenicity in mice. What is novel and less

obvious in this investigation is that we used *in silico* predictions to make site-specific framework mutations which resulted in a nearly 2°C increase in thermal stability to the V3 framework, and when combined with a cytotoxicity enhancing mutation also derived by *in silico* methods, resulted in enhancement of both antigen binding affinity and tumor cell killing potency. In fact, we had previously shown that cytotoxicity enhancing Ile mutation (HC:G54I) had nearly the same binding to GD2 as the parental hu3F8 (15), but when the same mutation was inserted into the more stable V3 framework in this investigation, there was a greater than twofold enhancement to GD2 binding. We have therefore demonstrated that structural and computational methods can be used to refine MoAbs that bind to complex carbohydrate targets such as GD2, with further *in vivo* validation necessary to progress toward clinical development.

## ACKNOWLEDGMENTS

The authors would like to thank Ms. Hong-fen Guo and Ms. Yi Feng for excellent technical assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/Journal/10.3389/fimmu.2014.00372/abstract>

## REFERENCES

- Ahmed M, Cheung NK. Engineering anti-GD2 monoclonal antibodies for cancer immunotherapy. *FEBS Lett* (2014) **588**(2):288–97. doi:10.1016/j.febslet.2013.11.030
- Lammie GA, Cheung NKV, Gerald W, Rosenblum M, Cordon-Cardo C. Ganglioside GD2 expression in the human nervous system and in neuroblastomas – an immunohistochemical study. *Int J Oncol* (1993) **3**:909–15.
- Cheung NK, Saarinen UM, Neely JE, Landmeier B, Donovan D, Coccia PF. Monoclonal antibodies to a glycolipid antigen on human neuroblastoma cells. *Cancer Res* (1985) **45**(6):2642–9.
- Chang HR, Cordon-Cardo C, Houghton AN, Cheung NK, Brennan MF. Expression of disialogangliosides GD2 and GD3 on human soft tissue sarcomas. *Cancer* (1992) **70**(3):633–8. doi:10.1002/1097-0142(19920801)70:3<633::AID-CNCR2820700315>3.0.CO;2-F
- Grant SC, Kostakoglu L, Kris MG, Yeh SD, Larson SM, Finn RD, et al. Targeting of small-cell lung cancer using the anti-GD2 ganglioside monoclonal antibody 3F8: a pilot trial. *Eur J Nucl Med* (1996) **23**(2):145–9. doi:10.1007/BF01731837
- Liang YJ, Ding Y, Lavery SB, Lobaton M, Handa K, Hakomori SI. Differential expression profiles of glycosphingolipids in human breast cancer stem cells vs. cancer non-stem cells. *Proc Natl Acad Sci U S A* (2013) **110**(13):4968–73. doi:10.1073/pnas.1302825110
- Battula VL, Shi Y, Evans KW, Wang RY, Spaeth EL, Jacamo RO, et al. Ganglioside GD2 identifies breast cancer stem cells and promotes tumorigenesis. *J Clin Invest* (2012) **122**(6):2066–78. doi:10.1172/JCI59735
- Yanagisawa M, Yoshimura S, Yu RK. Expression of GD2 and GD3 gangliosides in human embryonic neural stem cells. *ASN Neuro* (2011) **3**(2):doi:10.1042/AN20110006
- Martinez C, Hofmann TJ, Marino R, Dominici M, Horwitz EM. Human bone marrow mesenchymal stromal cells express the neural ganglioside GD2: a novel surface marker for the identification of MSCs. *Blood* (2007) **109**(10):4245–8. doi:10.1182/blood-2006-08-039347
- Jin HJ, Nam HY, Bae YK, Kim SY, Im IR, Oh W, et al. GD2 expression is closely associated with neuronal differentiation of human umbilical cord blood-derived mesenchymal stem cells. *Cell Mol Life Sci* (2010) **67**(11):1845–58. doi:10.1007/s00018-010-0292-z
- Saito M, Yu RK, Cheung NKV. Ganglioside GD2 specificity of monoclonal antibodies to human neuroblastoma cell. *Biochem Biophys Res Commun* (1985) **127**(1):1–7. doi:10.1016/S0006-291X(85)80117-0
- Cheung NK, Lazarus H, Miraldi FD, Abramowsky CR, Kallick S, Saarinen UM, et al. Ganglioside GD2 specific monoclonal antibody 3F8: a phase I study in

- patients with neuroblastoma and malignant melanoma. *J Clin Oncol* (1987) **5**(9):1430–40.
13. Cheung NK, Guo H, Hu J, Tassev DV, Cheung IY. Humanizing murine IgG3 anti-GD2 antibody m3F8 substantially improves antibody-dependent cell-mediated cytotoxicity while retaining targeting in vivo. *Oncol Immunology* (2012) **1**(4):477–86. doi:10.4161/onci.19864
  14. Cheung NK, Cheung IY, Kushner BH, Ostrovnya I, Chamberlain E, Kramer K, et al. Murine anti-GD2 monoclonal antibody 3F8 combined with granulocyte-macrophage colony-stimulating factor and 13-*cis*-retinoic acid in high-risk patients with stage 4 neuroblastoma in first remission. *J Clin Oncol* (2012) **30**(26):3264–70. doi:10.1200/JCO.2011.41.3807
  15. Ahmed M, Goldgur Y, Hu J, Guo HF, Cheung NK. In silico driven redesign of a clinically relevant antibody for the treatment of GD2 positive tumors. *PLoS One* (2013) **8**(5):e63359. doi:10.1371/journal.pone.0063359
  16. Gao SH, Huang K, Tu H, Adler AS. Monoclonal antibody humanness score and its applications. *BMC Biotechnol* (2013) **13**(55):doi:10.1186/1472-6750-13-55
  17. Hakomori S. Role of gangliosides in tumor progression. In: Svennerholm L, Asbury AK, Reisfeld RA, Sandhoff K, Suzuki K, Tettamanti G, et al. editors. *Progress in Brain Research*. Netherlands: Elsevier Science BV (1994).
  18. Kuroda D, Shirai H, Jacobson MP, Nakamura H. Computer-aided antibody design. *Protein Eng Des Sel* (2012) **25**(10):507–21. doi:10.1093/protein/gzs024
  19. Wang T, Duan Y. Probing the stability-limiting regions of an antibody single-chain variable fragment: a molecular dynamics simulation study. *Protein Eng Des Sel* (2011) **24**(9):649–57. doi:10.1093/protein/gzr029
  20. Cheng M, Ahmed M, Xu H, Cheung N-KV. Structural design of disialoganglioside GD2 and CD3-bispecific antibodies to redirect T cells for tumor therapy. *Int J Cancer* (2014):doi:10.1002/ijc.29007
  21. Liu W, Onda M, Kim C, Xiang L, Weldon JE, Lee B, et al. A recombinant immunotoxin engineered for increased stability by adding a disulfide bond has decreased immunogenicity. *Protein Eng Des Sel* (2012) **25**(1):1–6. doi:10.1093/protein/gzr053

**Conflict of Interest Statement:** Mahiuddin Ahmed and Nai-Kong V. Cheung were named as inventors in patents related to antibody 3F8 filed by Memorial Sloan Kettering Cancer Center. Jian Hu declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 May 2014; accepted: 21 July 2014; published online: 14 August 2014.

Citation: Ahmed M, Hu J and Cheung N-KV (2014) Structure based refinement of a humanized monoclonal antibody that targets tumor antigen disialoganglioside GD2. *Front. Immunol.* 5:372. doi: 10.3389/fimmu.2014.00372

This article was submitted to *Immunotherapies and Vaccines*, a section of the journal *Frontiers in Immunology*.

Copyright © 2014 Ahmed, Hu and Cheung. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Carbohydrate-mimetic peptides for pan anti-tumor responses

Thomas Kieber-Emmons<sup>1\*</sup>, Somdutta Saha<sup>1</sup>, Anastas Pashov<sup>2</sup>, Behjatolah Monzavi-Karbassi<sup>1</sup> and Ramachandran Murali<sup>3</sup>

<sup>1</sup> Department of Pathology and Winthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, AR, USA

<sup>2</sup> Stephan Angelov Institute of Microbiology, Bulgarian Academy of Sciences, Sofia, Bulgaria

<sup>3</sup> Research Division of Immunology, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA

## Edited by:

Elizabeth Yuriev, Monash University, Australia

## Reviewed by:

Mark Agostino, Curtin University, Australia

Mauro Sergio Sandrin, University of Melbourne, Australia

## \*Correspondence:

Thomas Kieber-Emmons, University of Arkansas for Medical Sciences, 4301 West Markham Street, #824, Little Rock, AR 72205, USA  
e-mail: tke@uams.edu

Molecular mimicry is fundamental to biology and transcends to many disciplines ranging from immune pathology to drug design. Structural characterization of molecular partners has provided insight into the origins and relative importance of complementarity in mimicry. Chemical complementarity is easy to understand; amino acid sequence similarity between peptides, for example, can lead to cross-reactivity triggering similar reactivity from their cognate receptors. However, conformational complementarity is difficult to decipher. Molecular mimicry of carbohydrates by peptides is often considered one of those. Extensive studies of innate and adaptive immune responses suggests the existence of carbohydrate mimicry, but the structural basis for this mimicry yields confounding details; peptides mimicking carbohydrates in some cases fail to exhibit both chemical and conformational mimicry. Deconvolution of these two types of complementarity in mimicry and its relationship to biological function can nevertheless lead to new therapeutics. Here, we discuss our experience examining the immunological aspects and implications of carbohydrate-peptide mimicry. Emphasis is placed on the rationale, the lessons learned from the methodologies to identify mimics, a perspective on the limitations of structural analysis, the biological consequences of mimicking tumor-associated carbohydrate antigens, and the notion of reverse engineering to develop carbohydrate-mimetic peptides in vaccine design strategies to induce responses to glycan antigens expressed on cancer cells.

**Keywords:** glycans, carbohydrate-mimetic peptide, mimotope, vaccines, structural design, cancer

## INTRODUCTION

Among the most challenging of antigen targets for vaccine design are glycans (1). They are ubiquitous in nature and can be considered as one of the unique antigens expressed across pathogens and cancer cells. Glycans are fundamental to the biological functions of cell-cell communication, cell proliferation, and differentiation, and they mediate cell attachment, as well as mediating pathogen attachment and infection. Cancer cells, in particular, are noted for their aberrant glycosylation profiles that affect the metastatic process. Consequently, certain carbohydrate forms profoundly affect both the pathophysiology of infection and neoplasia (Table 1). A unique advantage in targeting tumor-associated carbohydrate antigens (TACAs) is that multiple proteins and lipids on cancer cells can be modified with the same carbohydrate structure which might be shared with bacterial antigens (2). Thus, targeting TACAs has the potential to broaden the spectrum of antigens recognized by the immune response, thereby lowering the risk of developing resistant tumors due to the loss of a given protein antigen.

We have come to learn that the manner a TACA is expressed will dictate how an immune effector mechanism will be invoked (8). Antibodies against glycolipids and globular glycoproteins are found to mediate complement-dependent cytotoxicity (CDC) because they extend less than 100 angstroms from the cell

membrane while antibodies to mucins that extend up to 5000 angstroms from the cell surface do not (8). But TACAs are also associated with cell signaling activities whereby anti-TACA antibodies are capable of direct induction of cell death of number of tumor cell lines, but this activity has not been investigated in great detail (9, 10). In this context, TACAs are pan-targets on tumor cells because they are collectively and intimately involved in cell-death signaling pathways. Strategies that target TACAs have, therefore, potential clinical benefit as cell-death therapies. Anti-TACA antibodies can mediate significant reprogramming of signaling events, with profound anti-tumor activities. The ability to induce antibodies reactive with multiple TACAs is relevant as heterogeneity of antigen expression in different cancers of the same type, as well as different cells of the same cancer, and heterogeneity of immune response in different patients make it likely that maximal anti-cancer effect may not result from immunization against a single antigen.

The success of carbohydrate-based vaccines against pathogens has led to technological advances in vaccine design, but they have typically been developed as mono or singular vaccine types requiring a polyvalent formulation to induce responses across carbohydrate types (11). While glycans are diverse in expression patterns and in their composition, the structural commonalities among glycans provide a template to target, at least some of them collectively,

**Table 1 | Glycosphingolipid Constituents Shared Among Bacteria and Tumor cells.**

GSL series type	Structure	Bacterial species
Lacto	Gal $\beta$ 1 $\rightarrow$ 4GlcNAc $\beta$ 1 $\rightarrow$ 3Gal $\beta$ 1 $\rightarrow$ 4Glc $\beta$ 1 $\rightarrow$ Cer	<i>N. gonorrhoeae</i> <i>N. meningitidis</i> <i>Moraxella catarrhalis</i> <i>Helicobacter pylori</i> <i>H. influenzae</i> <i>Campylobacter jejuni</i> <i>H. ducreyi</i>
Globo	Gal $\alpha$ 1 $\rightarrow$ 4Gal $\beta$ 1 $\rightarrow$ 4Glc $\beta$ 1 $\rightarrow$ Cer	<i>N. gonorrhoeae</i> <i>N. meningitidis</i> <i>H. influenzae</i> type b <i>H. influenzae</i> NT <i>B. catarrhalis</i>
Ganglio	GalNAc $\beta$ 1 $\rightarrow$ 4Gal $\beta$ 1 $\rightarrow$ 4Glc $\beta$ 1 $\rightarrow$ Cer Gal $\beta$ 1 $\rightarrow$ 3GalNAc $\beta$ 1 $\rightarrow$ 4Gal $\beta$ 1 $\rightarrow$ 4Glc $\beta$ 1 GalNAc $\beta$ 1 $\rightarrow$ 3Gal $\beta$ 1 $\rightarrow$ 4GlcNAc $\beta$ 1 $\rightarrow$ 3Gal $\beta$ 1 $\rightarrow$ 4Glc $\beta$ 1 $\rightarrow$ Cer	<i>N. gonorrhoeae</i>

*Lacto series, neolactoseries, Globosides, and Ganglioside antigens are found on tumor cells (3, 4) and on LOS of multiple bacteria (5–7).*

by directing the immune response toward these commonalities. Therefore, it is logical to target glycans in vaccine design, which can lead to the interruption of disease processes (11).

Among potential technological strategies is using carbohydrate-mimetic peptides (CMP) to induce responses to glycans on pathogens and cancer cells (12). Peptides can substitute as immunogens to target pathways involving protein–carbohydrate interactions and in carbohydrate-specific immunological reactions. However, there is a noted distinction between the ideas of antigenic mimicry versus the ability of a mimic to induce a response cross-reactive with a carbohydrate/glycan moiety.

Antigenic mimicry, in simple terms, is when one ligand competes with another for antibody binding. The origin of cross-reactivity involves thermodynamic and structural interpretations (13–15). The notion of immunological mimicry is less precise. Does it mean that the mimic generates the same antibody subset as the nominal antigen or just that it induces a response that cross-reacts with the nominal antigen?

Early on CMPs were shown to function as antigenic mimics (16, 17) but more importantly they were shown to induce serum antibodies in a variety of systems, having utility in directing responses to cancer cells and against pathogens (18–26). Most of all, unlike carbohydrate antigens, CMPs can prime for memory responses to TACAs (27) suggesting that the CMPs facilitate cognate interactions between B cells and T cells, which is something that carbohydrates/polysaccharides do not facilitate, but surrogate antigens of carbohydrates such as anti-idiotypic antibodies

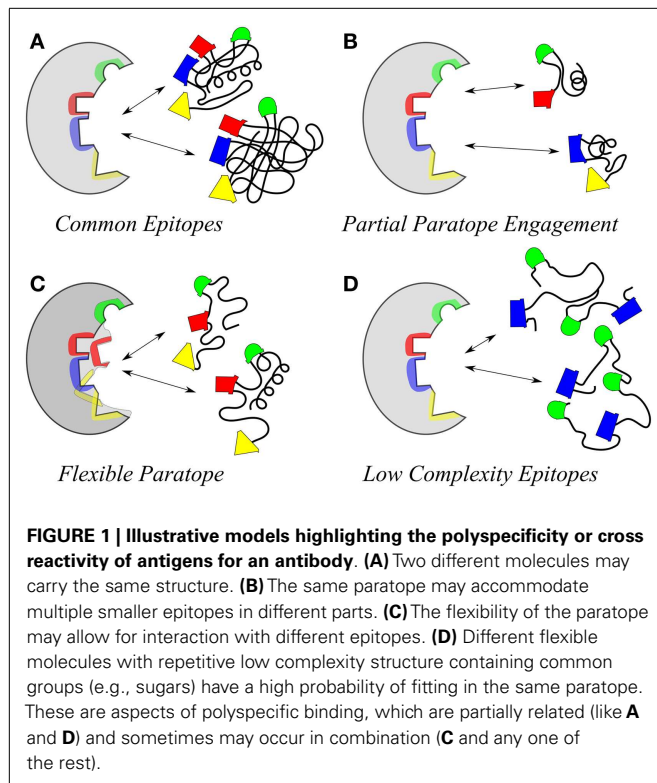
and CMPs should and can do. CMPs are not only a functional strategy to induce carbohydrate-reactive responses, but also they can function as probes to understand the structural basis for the dual recognition properties of antibodies, lectins, and T cells (12, 14, 15, 28, 29). Understanding the structural requirements for antibody and T-cell recognition provides a basis for identifying potentially new sets of immunogens that may have both fundamental immunological and clinical value. However, it has been argued that translation of such information into viable vaccines is still a long way off (30–32). Here, we briefly discuss the various perspectives and elements of CMPs useful to translate them into the clinic in tumor vaccine design applications to target glycans.

## MOLECULAR MIMICRY AT A GLANCE

Molecular mimicry is now firmly considered as the basis of many autoimmune disorders, proposed as a pathogenic mechanism for autoimmune disease, as well as a probe useful in uncovering its etiologic agents (33). On the other hand, self-limiting autoimmunity may underlie some of the pathogenic mechanisms in infectious disease. This hypothesis is based in part on observed cross-reactivity of immune reagents with host “self” antigens and microbial determinants (33). Molecular mimicry is also suggested as a means to regulate immune homeostasis and to elicit responses against target antigens as evidenced by studies on anti-idiotypic antibodies (34). This model suggests that conventional T-cell/B-cell collaboration can explain communication between complementary idotype [Id(+)] and anti-Id antibody at the cellular level that integrates present and previous data on B-cell regulation. Furthermore, this model provides a tool to probe carbohydrate immunology paradigms because the synergistic interaction of effector T and B cells require common recognition of identical tumor-associated antigen(s) (35). Anti-idiotypic antibodies have been proposed to mimic carbohydrate antigens and have been tested in the clinic (36–40).

At one level, an explanation for molecular mimicry is when a foreign antigen shares sequence or structural similarities with self-antigens. But on another level what defines the recognition and interaction basis for antigenic mimicry that ties to a functional immune response? Molecular mimicry in the context of antibody–antigen recognition is interpreted at several levels (Figure 1). The work by Hoffmuler et al. (41) suggests that a common epitope can be preserved among an ensemble of peptide variants. They demonstrated that the binding modes of intermediate conformation of selected peptides were characterized using complete sets of substitution analogs, revealing that a number of sequential substitutions accumulated without changing the pattern of key interacting residues. At a distinct step, however, one single amino acid exchange induces a change in the binding mode, indicating a flip in specificity and conformation (41). Regions of proteins with biased amino acid composition [so-called Low-Complexity Regions (LCRs)] are abundant in the protein universe (42). LCR-containing proteins tend to have more binding partners across different networks than proteins that have no LCRs. LCRs may be involved in flexible binding associated with specific functions, but also that their positions within a sequence may be important in determining both their binding properties and their biological roles (42).





Intrinsically disordered regions of proteins have also been associated with molecular mimicry (43), indicating the potential of highly flexible peptides as mimics. Such peptides may be attractive to induce pan anti-tumor responses, due to their potential ability to mimic multiple TACAs *in situ*. However, the structural diversity inherent to such peptides makes defining the precise nature of their mimicry of any or multiple TACAs even more challenging. Geometrical shape complementarity, the “lock and key” hypothesis, between antigen–antibody interaction, has long dominated immunological thinking. However, studies demonstrating the existence of a large number of monoclonal antibodies that can bind to a variety of totally unrelated self and foreign antigens (i.e., polyreactive antibodies) have modified this view. Consequently the lock and key model has been supplemented with an explanation focusing on the flexibility of antibody binding sites that can change conformation to accommodate different antigens (44).

Antibodies induced by a CMP to the meningococcal group C capsular polysaccharide (18) were shown to be reactive with the Lewis Y antigen (20). Carbohydrate-reactive antibodies show the potential cross-reactivity for dissimilar carbohydrate forms that highlight the common epitope basis for cross-reactivity (Figure 2); Figure 2A shows that a common epitope is formed between  $\alpha$ 2–8 sialic acid and the neolactoseries antigen Lewis Y (18, 20). The potential of antibodies recognizing three hydroxyl groups might be cross-reactive with three hydroxyl groups displayed on two glycosyl groups (Figure 2B). This level of recognition leads to the idea that antibodies can recognize carbohydrate in the context of pan-recognition. The cases discussed above relate as much to the common epitope mechanism as to the low-complexity epitopes, which seems to be often the case in carbohydrate recognition.

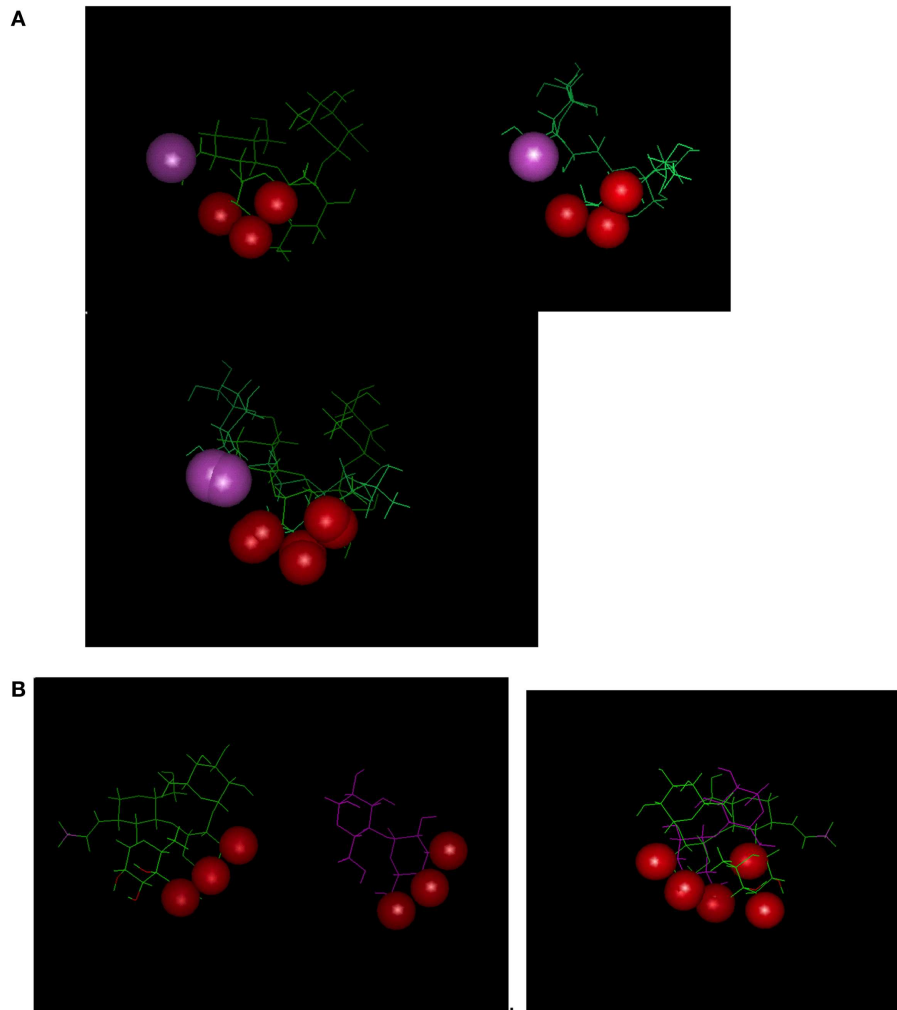
What is discussed here, strictly speaking, is molecular interaction at the atomic level, while recognition is rather the system level processing of information relevant to immune function, i.e., self/non-self distinction and identification of previously met danger (32, 45). Specificity of interaction serves these purposes only in some aspects while others favor polyspecific binding. For T-cell receptors (TCR) antigen specificity is an emerging property of the system rather than a characteristic of the individual receptor (46). On a molecular level, TCRs are a rather promiscuous binder. Furthermore, in terms of pre-immune antibodies, polyspecificity has also the role of ensuring a complete repertoire. It funnels antigen/pre-immune antibody interactions into the somatic hypermutation process of refining specificity.

An interesting twist to this topic is the emerging notion of reverting specific antibodies to polyspecific binding or induced polyspecificity as a physiological mechanism operating for instance at the sites of inflammation (47–49). Yet, perhaps to most, typical polyspecific immune binding makes use of pattern recognition to generalize a danger context (50, 51). Functionally, the boundary between pattern recognition receptors and natural antibodies is fuzzy (52–54). Intrinsically, prone to polyspecificity by several mechanisms, antibody recognition of carbohydrates conceptually merges antigen and pattern recognition. In this regard, carbohydrate mimotopes (e.g., CMP) instead of mimicking one particular structure by another come about rather as mimics of patterns, not unlike synthetic TLR agonists. But carbohydrate mimotopes are not exclusively artificial. CMPs from natural proteins are known for some time (55, 56). Peptides from Mucin 1 cell surface receptor (MUC1) are the most interesting because they are considered mimics of the Gal-epitope (56). Natural peptides can adopt structures similar to carbohydrate antigens (21) and can exhibit binding kinetics similar to the nominal antigens that they mimic (21, 28).

Often times CMPs share no obvious consensus sequence but their amino acid sequences often contain aromatic and hydrophobic residues but also amino acids having cyclic side chains, including proline and glycine that affects the conformational properties of the mimic (13, 57). The predominance of aromatic residues in CMPs invokes interaction scenarios that include stacking and hydrophobic interactions. A basis for this is the notion that carbohydrate recognition by antibodies use hydrophobic faces on carbohydrate antigens (58). It is important to note that cohesive solvent–solvent interactions are the major driving force behind apolar association in solution (59). Consequently, interaction models that implicate important roles for dispersion forces in molecular recognition events should be interpreted with caution in solvent-accessible systems (59). In addition, other antibody recognition systems also suggest that dual antigen recognition could involve divergent antibody conformations of nearly equivalent energetic states (60). Therefore, developing high-affinity-binders might make use of antibody structural plasticity to mediate the recognition step without increasing the entropic cost (60).

#### HUMORAL RESPONSES TO CMPs

While a variety of CMPs have been developed with the ability to induce immune responses of desired specificities and functionality (61) they are perhaps most appealing as a probe to understand the



**FIGURE 2 | Examples depicting similarity of epitopes in dissimilar carbohydrate antigens.** Epitopes (hydroxyls) are represented by red-spheres. **(A)** Relationship between Lewis Y antigen on left side of panel with MCP on

right side. **(B)** Relationship between Lewis Y antigen on left side of panel with  $\alpha$ -4 Glucose on right side of panel. Interestingly the epitope defined on the glucose moiety defines a three-dimensional epitope on the Lewis Y antigen.

immunological response to carbohydrate antigens. An important feature of CMPs is in their ability to mediate contact-dependent T-cell help as an obligatory role in humoral immune responses to T-cell dependent antigens. Cognate B-cell/T-cell interactions during the immune response to protein antigens depend on T-cell co-stimulation. Details of how such interactions govern immune responses to carbohydrate-conjugate vaccines are few. We have shown that immunization with CMPs activate peptide-specific T helper type 1 (Th1) and type 2 (Th2) responses (62, 63). However, while behaving like a Th1 antigen (63), multivalent peptide mimetics still could induce a high carbohydrate-reactive IgM/IgG ratio with an endpoint titer of 1:2,000 (20). These results suggest that the multiple antigen peptide form might function like a Th2 independent immunogen in BALB/c mice. Furthermore, we observed that CMPs mediate cognate B and T-cell interactions as CMPs can induce antibodies in a host with deficiency in IgM production that typically do not respond to carbohydrate

antigens (62). In these studies apparently the B cells functioned as antigen-presenting cells. In addition, these studies suggest that B-cell subsets influence the interactions. More importantly, the type of TACA mimicked by the CMPs is expressed in mice (29, 64). Consequently, these studies are obtained in a toleragenic mouse model, further suggesting that tolerance is broken upon CMP immunization.

A characteristic of an effective mimotope based vaccine would be to prime for secondary responses upon boosting or challenge with native antigen (18, 27, 65–67). Peptide-mimotope anamnestic responses have been noted for mimotope-conjugates (65, 66). The identification of peptide mimetics relies upon the idea that antibody fine specificity epitope mapping patterns of carbohydrates and peptide mimetics might be used as a proxy for individual B-cell receptor specificity activated during a secondary antibody response. However, the idea of functional mimicry would suggest that immunization with a carbohydrate-mimic peptide might also

induce a specific subset or restricted anti-carbohydrate response. Our studies indicate that since peptide-conjugates elicit immune responses in xid mice (62), it is likely that antibodies to peptide and carbohydrate immunogens might be structurally unique and derived from different antibody subsets.

### POTENTIAL FOR CELLULAR IMMUNITY TARGETING CARBOHYDRATE ANTIGENS

Up until a few years ago, carbohydrate determinants were traditionally not considered as targets for Cytotoxic T-Lymphocytes (CTL) despite a variety of immunogenicity and specificity studies for the glycan moiety of synthetic *O*-glycosylated MHC-binding peptides suggest otherwise (68–70). GD2 was also implicated as a target upon CTL activation early on (71). Crystal structure analyses indeed show that T cells can recognize glycopeptides bound by MHC molecules on the surface of antigen-presenting cells (72,73). T cells, therefore, have the potential to react with the carbohydrate moiety of neoglycopeptide antigens, suggesting that T cells can target carbohydrate antigens expressed on tumor cells. However, it is also possible to generate carbohydrate-specific unrestricted CTL responses with MHC class-I-binding carrier peptides (74) that might explain the GD2 response (71). Nonetheless, how such T-cell responses are generated is presently unclear. From a vaccine perspective, the construction of glycopeptide/protein immunogens is problematic.

Rather than simple molecular mimicry, unpredictable arrays of common and differential contacts on class-I complexes can be used for their recognition by the same TCR. For example, bacterial polysaccharides with a distinct charge-motif can be emulated by peptides that can activate T cells (75). Lysine–aspartic acid (KD) peptides with repeating units are able to stimulate CD4<sup>+</sup> T cells *in vitro* and confer protection against abscesses induced by bacteria such as *Bacteroides fragilis* and *Staphylococcus aureus* (75). CMPs can induce a Th1 response in mice using a DNA platform (76). We have observed an augmented induction of CTL activity against Meth A tumor cells upon peptide-mimotope immunization (63, 77). The induction of carbohydrate-reactive T-lymphocytes with peptide mimics is based upon a functional definition of T-cell mimotopes. One possible explanation is that the peptide-mimotope activates CTLs, which bind to *O*-linked GlcNAc or GalNAc glycopeptides associated with MHC Class-I. Based upon crystal structure analysis of MHC complexes with glycopeptides, it appears that the central region of the putative T-cell-receptor-binding site is dominated by the extensive exposure of the tethered carbohydrate (72,73). Our modeling of CMPs in the MHC Class-I groove suggests that amino acids and glycans attached to a glycopeptides overlap in 3D space, providing an array of contacts for TCR recognition (12).

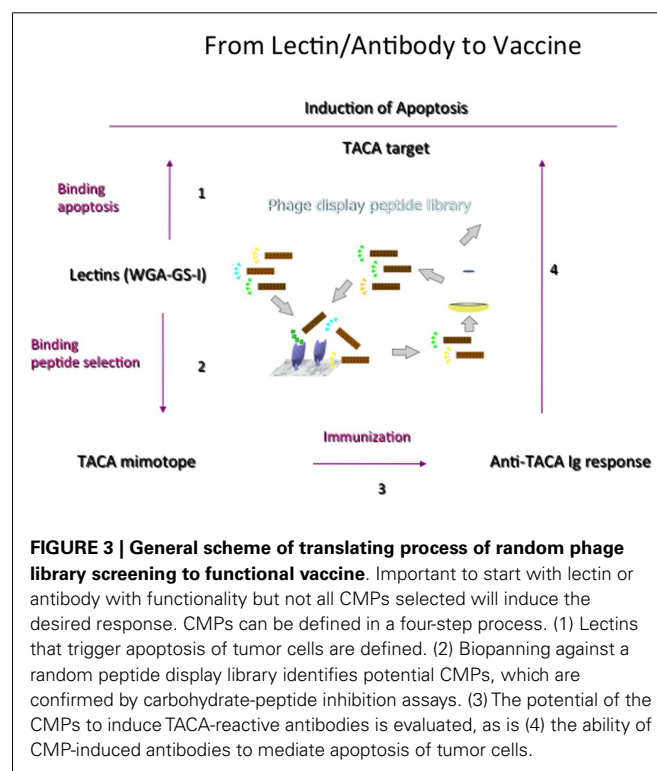
### FIDELITY IN MIMICRY

The ability to augment or enhance TACA-reactive antibodies using CMPs would be noteworthy. Much like anti-idiotypes, CMPs may elicit anti-saccharide responses, but fail to elicit the idiotypes and isotypes observed in the protective response to the microbial antigen (78). Functional antibodies depend not only on the host's ability to mount an immune response, but also on its ability to mount the appropriate immune response. Whether an antibody

response is protective or not may depend on both the fine antigenic specificity that may be associated with particular idiotypes and epitope binding characteristics, and the isotype, determining antibody effector function. Often times studies of peptide mimics selected by lectins or antibodies and then analyzed by structural approaches come to the conclusion that mimicry at structural level is minimal at best (13–15). The same conclusions are drawn in considering anti-idiotypic antibodies (79). Rather, mimics as peptides or anti-idiotypes serve as imprints of the structural characteristics of the nominal carbohydrate antigen and, consequently, give rise to antibodies with carbohydrate-like properties upon immunization. The question remains how to enhance the ability of TACA-mimetic peptides to induce TACA-reactive antibodies with higher titers and association constants. Herein lies the problem with mimics; the immune response is only assayed after a choice is made as to which mimic is to be followed. So what lessons can be learned about choosing the “true” mimic?

### From lectins to vaccines

While lectins have been generally used to identify CMPs and to understand the general features of recognition phenomena, **Figure 3** outlines the general development of CMPs in vaccine design using lectins as a template to induce antibodies that would emulate the actions of lectins. We have shown that this concept can be brought into practice (80). Plant lectins like *Griffonia simplicifolia* lectin I (GS-I) and wheat germ agglutinin (WGA) mediate the apoptosis of tumor cells. We have investigated the possibility of using these lectins as templates to select peptide-mimotopes of TACAs as immunogens to generate cross-reactive antibodies capable of mediating apoptosis of tumor cells (80).



**FIGURE 3 | General scheme of translating process of random phage library screening to functional vaccine.** Important to start with lectin or antibody with functionality but not all CMPs selected will induce the desired response. CMPs can be defined in a four-step process. (1) Lectins that trigger apoptosis of tumor cells are defined. (2) Biopanning against a random peptide display library identifies potential CMPs, which are confirmed by carbohydrate-peptide inhibition assays. (3) The potential of the CMPs to induce TACA-reactive antibodies is evaluated, as is (4) the ability of CMP-induced antibodies to mediate apoptosis of tumor cells.

Vaccine-induced anti-carbohydrate antibodies to both 106 and 107 (**Table 2**) reduced the outgrowth of micrometastases in the 4T1 spontaneous tumor model, significantly increasing survival time of tumor-bearing animals. This finding parallels suggestions that carbohydrate-reactive IgM with cytotoxic activity may have merit in the adjuvant setting if the right carbohydrate-associated targets are identified (81, 82). Interestingly, while both CMPs 106 and 107 are reactive with lectins only 107 induced responses that were directly cytotoxic to tumor cells. Both CMPs induced antibodies that mediated CDC, however, only CMP 107 induced serum IgM antibodies in mice that mediated the apoptosis of murine 4T1 and human MCF7 cell lines *in vitro*, paralleling the apoptotic activity of the lectins (80). This finding again highlights that selection of CMPs based upon antigenic mimicry does not automatically translate into inducing antibodies with a desired functionality.

Fundamental feature of these CMPs was their hydrophobic nature being built on motifs containing aromatic residues. Early on, peptides that mimic carbohydrate antigens were identified by analysis of reactivity of random peptide libraries on phage with the lectin Concanavalin A (ConA) (16, 17). These early peptides contain aromatic side chains, representing a generalized Trp/Tyr/X/Tyr (where X is a number of different residues) motif. Subsequent to these seminal studies other aromatic peptides displaying similarities to ConA-reactive ones were described (18, 83). Aromatic residues, hydrophobic, and hydrogen bonding amino acids seem favored but with the possibility that the W/YXY motif functionally mimic elements of Core 1 and 2 structures shared among otherwise dissimilar carbohydrate structures (**Figure 2**). Consequently, this motif type has been observed in peptides isolated by a number of anti-carbohydrate antibodies and lectins and might represent low-complexity surfaces (**Figure 1**) perhaps because of the bias in the amino acid composition of the mimetics. Such biased sequences do not necessarily converge on a canonical set of patterns although some motifs stand out. It is important to note how those peptides reactive with ConA were identified based upon a conception of antigen mimicry, as the work of Westerink et al. (18) were based upon immunological studies starting with an anti-Id that displayed immunological functionality.

### Structure-based reverse engineering to discover peptide mimics

The caveats associated with screening libraries with either lectins or with antibodies often lead to identifying mimics that fail to mimic critical contacts that the carbohydrate makes with the protein, and there is a possibility that such peptides may bind to alternative sites on protein to the carbohydrate-binding site,

making optimization of a true structural mimic from such a peptide impossible (84). The structural approaches to define the basis of mimicry have been previously discussed (13, 14). As mentioned above high-affinity peptides *per se* may not necessarily mimic critical contacts required for the function. In addition, the judicious choice of peptides for testing antibody responses against should be based on the peptide interaction with both the heavy and light chain in order to induce antibodies with similar antigen specific properties (28); as the combination of heavy and light chains will influence specificity. Thus, both the variable and the constant region of the antibodies induced by a peptide mimic or mimotope must be considered when assessing the success of any immunization.

To overcome the limitation of high-affinity peptides' lack of immunological mimicry, we adopted a "reverse engineering approach" sometime ago, which places emphasis on the maintaining critical contacts between carbohydrates and its protein partner (28, 29). This method is similar to fragment-based drug discovery (28). We have previously reviewed the structural concepts and approaches used in vaccine design applications that illustrate the value and limitations of using chemical (peptide libraries which are mimics of a ligand) and immunological information to define novel peptide immunogens that function as mimotopes to generate immune responses targeting TACA (85) and glycans on the human immunodeficiency virus (86). In this context, we showed that concepts associated with pharmacophore design (now considered reverse engineering) could be used to define CMPs applied to vaccine design (21, 28). We demonstrated a structure-assisted vaccine design approach, whereby small molecules, defined in crystallographic databases, could be used in principle to define peptide mimetics emulating the three-dimensional interaction scheme of a native carbohydrate antigen (21, 28). More importantly, it was shown that virtual screening led to motifs being observed experimentally and that they could display binding energetics similar to the nominal carbohydrate antigen (28).

We have also shown that by using this approach, an immunogenic peptide (911 **Table 2**) can be designed *de novo* using ConA as a template inducing antibodies with the same functionality as ConA in neutralizing HIV isolates (21). In addition, we showed that peptides could adopt structures that are similar to carbohydrate conformations that include extended beta strand type and helical structures (21). Using reactivity patterns of glycan binding to ConA coupled with structural design concepts we identified a peptide (referred to as 911) (**Table 2**) that when rendered as a multiple antigenic peptide (MAP) was reactive with ConA at

**Table 2 | Selected CMPs that we have studied.**

Peptide	Sequence	Lectin	Functionality
911	YRYRYGRYRSGSYRYRYGRYRSGS	Con A	Neutralizes HIV Lab isolates
912	RYRYGRYRSGS	Con A	
106	GGIYWRYDIYWRYDIYWRYD	GS-1, WGA	Mediates CDC
107	GGIYYRYDIYYRYDIYYRYD	GS-1, WGA	CDC, Apoptosis
P10	GVVWRYTAPVHLGDG	GS-1, WGA	Tumor growth inhibition
P10s	WRYTAPVHLGDG	GS-1, WGA	Tumor growth inhibition in mice, apoptosis in humans

lower concentrations than those required for reaction of some native oligosaccharide ligands of ConA (21). The 911-MAP displayed competitive inhibition with carbohydrate ligands of ConA, indicating that it binds at an overlapping carbohydrate-binding site on ConA. Isothermal Calorimetric analyses and immunoprecipitation experiments suggest that a shorter monovalent putative peptide 912 (Table 2) exhibited a weak affinity comparable to that of Me $\alpha$ Man (21). The 911-MAP exhibited a higher association constant and free energy of association with ConA compared with that found upon binding of the putative 912 peptide and the  $K_a$  and  $\Delta G$  values of 911-MAP are comparable to those of ConA-reactive trimannoside and pentasaccharide (21). Most importantly, the 911-MAP induce antibodies in mice that are capable of neutralizing HIV-1 III-B as assessed by p24 ELISA (21). This is work perhaps for the first time demonstrated that design-principles associated with CMPs could be useful to induce functional antibodies. Similar approaches have since been applied to investigate peptide recognition by anti- $\alpha$ -Gal antibodies (87) and in developing CMPs of gangliosides (88). As in our studies, it was found that peptides could interact with the same residues as those involved in carbohydrate recognition. In this context, CMPs are envisioned to be further enhanced as either inhibitors much like that in mainstream pharmacophore development or as in our case to develop vaccines targeting glycans.

To further emphasize the design principles to enhance the fidelity of mimicry, we tested the hypothesis that improving the hydrogen bond pattern through amino acid substitutions in a CMP, to be coincident with that for the carbohydrate ligand, will enhance the ability of CMPs to elicit anti-TACA antibodies with high titers and association constants (29). Based on anti-Id/Id crystal structures, highly directional bonds represent an important set of interactions to establish a basis for mimicry because they mainly confer the specificity in binding of the peptide and the carbohydrate antigen. In this exercise, we developed the CMP P10s (Table 2) (29). This CMP was identified from a random peptide library screen using the anti-GD2/GD3 antibody ME36.1 (89). P10 was shown to generate immune responses in mice that inhibited tumor growth *in vivo* (90).

In the development of P10s, we made use of the crystal structure of the anti-ganglioside antibody ME36.1 (29). Briefly, the crystal structure of ME36.1 was analyzed in the context of comparing GD2 binding and CMP binding using a molecular docking approach (29). Based on the hydrogen bonds interaction between GD2 and CDRs of ME36.1, P10s was designed. Conformational and docking calculations suggested that P10s would form an increased number of hydrogen bonds with ME36.1 that are in common with the GD2 hydrogen bond interaction pattern with ME36.1 [see Table 1 in Ref. (29)]. This increased level of mimicry would suggest that the immune response to GD2 upon immunization with P10s would be better. We observed that P10s did indeed induce higher titer antibodies to the target antigen and antigen expressing tumor cells than the parent CMP, P10. These studies suggest that for carbohydrate mimics, pharmacophore based design is superior over the conformational approach undertaken for other peptide mimics.

## PRECLINICAL ASSESSMENTS OF CMPs

Tumor-associated carbohydrate antigen are rather varied in their expression profiles on tumor cells and on normal tissue. TACAs are upregulated in many types of tumors, and therefore represent a potential vaccination target with widespread application. Cancer vaccines functionally resemble the process of autoimmune-mediated tissue damage (91).

Since tissue rejection is the goal of cancer immunotherapies, broad-spectrum, pan-antigens like TACA are plausible effective targets once the problem of their low immunogenicity is solved. This is the hope of CMP and anti-idiotypic vaccine research.

The basis of TACA mediate tumor rejection is akin to the observation that anti-Gal IgM and IgG mediate rejection of xenograft expressing  $\alpha$ -gal glycoconjugates with terminal Galalpha1-3Galbeta1-4GlcNAc sequences (alpha-galactosyl epitopes, natural xenoreactive antigens) that are present on various tissues in pigs and are recognized by human anti-alpha-galactosyl (alpha-Gal) antibodies (92). The tissue-rejection mediated by  $\alpha$ -Gal-reactive antibodies demonstrates the feasibility of targeting TACAs for tumor therapy because tumor-induced antibody responses resemble autoimmune responses (93).

The generation of tissue-rejection represents an important conceptual approach to cancer immunotherapy. Alpha-galactosylated xenoantigens (Galalpha1-3Galbeta1-4GlcNAcbeta1 and Galalpha1-3Galbeta1-4GlcNAcbeta1-3Galbeta1-4Glc) are often detected with the alpha-Gal-reactive lectin GS-1. However, this lectin exhibits a broad and variable specificity for carbohydrates terminating in alpha-Gal (94). The blood group reactive lectin GS-I, which recognizes alpha-galactosyl moieties is recognized as a surrogate marker to identify tumor expressed antigens reactive with anti-Gal antibodies and GS-I is of utility to interrogate terminal  $\alpha$ -GalNAc/Gal expression on human tissues (95). Some of these antigens are also expressed on normal cells at low levels, potentially creating a state of immune tolerance.

We have previously demonstrated that vaccination with the CMPs 106 and 107 (Table 2) can induce antibody responses leading to cell-mediated cytotoxicity and apoptosis, respectively, in murine models of cancer (80). In preclinical studies, we observe that immunization in mice with these CMPs do not induce significant immunopathology, organs including liver, kidney, heart, lungs, intestines, stomach, lymph nodes, spleen, brain, spinal cord, and eyes were examined in H&E stained sections. These organs are reactive with GS-1 and the CMPs induce antibodies reactive with GS-1 antigens (29, 64). No significant cellular infiltrates were identified in any organ, including brain and spinal cord, from any animal, and there was no evidence of necrosis or extensive apoptosis in these sections (29, 64). It is likely that the level or pattern of expression of these molecules on the surface of tumor cells differs significantly from that on normal cells mediated by antibody avidity and the clustering of glycan epitopes (96). This difference in expression may account for the relative specificity of immunologic injury for tumor cells over normal cells.

Antibodies induced by CMPs are thought to have low affinities for TACA that might compensate for the low-affinity of the carbohydrate cross-reactive antibodies, minimizing the destruction of normal tissue. Such results demonstrate that repeated injections of CMPs do not necessarily lead to immune mediated injury



and support the development of CMPs for clinical testing (29, 64). Bringing such vaccines to the treatment armamentarium may significantly improve outcomes for patients.

### CLINICAL ASPECTS OF MIMICRY

The potential benefits of inducing TACA-reactive antibodies in patients with cancer are demonstrated by observations that patient survival significantly correlates with ganglioside-reactive IgM levels (97). The fact that survival rates of cancer patients are correlated with low-titer, and presumably low-affinity, TACA-reactive antibodies argues that more robust antibody responses may not be necessary. Cross-reactions are important issues in vaccine development field. As self-antigens induce tolerance, vaccination with non-self-antigens that molecularly mimic self-antigens may avoid tolerance and lead to generation of anti-tumor immune responses. In this context, little attention has been paid to the fact that the tumor-associated antigen MUC1 might be a natural CMP. In a series of studies from McKenzie's group, it was noted that anti- $\alpha$ -Gal antibodies reacted with MUC1 antigens and that anti-MUC1 antibodies reacted with the  $\alpha$ -Gal sugar (56). In mice, MUC1 peptide immunization resulted in cellular responses with reported little humoral response. In contrast, the MUC1 peptide induced a strong antibody response in human immunization. It was argued that pre-existing anti-Gal antibodies in human was the basis for the differential response as the Gal-epitope is a natural antigen in mice (56).

The mimicking of MUC1 with the Gal-epitope might have important consequences. The natural cross-reactivity of anti-Gal antibodies against MUC1 might lend to confusion making it difficult to ascertain the relative contributions of antibodies in binding to MUC1 upon MUC1 immunization; e.g., whether one can dissect if anti-MUC1 antibodies are not anti-TF and whether anti-MUC1 antibodies are not anti-Gal antibodies (98). On the other hand, the cross-reactivity of anti-Gal antibodies with MUC1 might lend to anti-MUC cellular immunity. Dendritic cells (DCs) play an important role in the induction of T-cell responses. Fc gammaRs (Fc $\gamma$ R), expressed on DCs, facilitate the uptake of complexed antigen, resulting in efficient MHC class-I and MHC class-II Ag presentation and DC maturation (99, 100). IgG-complexed MUC1 internalized through Fc $\gamma$ R on DCs might be efficiently presented to CTLs through the MHC class-I pathway as observed in other systems (99, 100). However, these mechanisms might also be responsible for antibody-mediated enhancement *in vivo* as suggested by the McKenzie work in humans and in animal models where antigen-IgG and IgE complexes exacerbated Th2 cells rather than Th1 cells (101). Therefore, mimicry of the Gal-epitope by MUC1 might skew Th2 type responses to MUC1 vaccines, which is contradictory to the present paradigm that stresses Th1 responses as being beneficial to MUC1 and other tumor-associated antigens.

While CMPs of TACA have been described that include the ganglioside GD2 (89, 95, 102–105), the ganglioside GD3 (106), sialylated Lewis a/x (107), and LeY (89), none of these CMPs have made it to the clinic except for our P10s. In contrast, several anti-idiotypic antibodies that mimic different GD2 (108–111), GD3 (112), and N-glycolyl (NGc) gangliosides (40) have made it to clinical trials. The most advanced is the GD3 mimicking antibody Bec2, which has been tested in a Phase III trial

(112). Unfortunately, there was no improvement in survival or progression-free survival in the vaccination arm with Bec2. Each of these anti-idiotypes seems to have a different mechanism of action against cancer cells but parallel mechanisms observed with CMPs. In the case of the anti-idiotypic that mimics NGc gangliosides it generates a humoral response that triggers cell death but differently than typical apoptosis (113). Patients that developed IgG and/or IgM Abs against NeuGcGM3 showed longer median survival times (114). Immunizations with the GD2/GD3 surrogates are less mechanism based. Bec2 induces antibody responses in about 25% of subjects (115). Consequently, different strategies using Bec2 have been considered including priming (38) and in combination therapy with adjuvant (116). For GD2, the anti-idiotypes induce GD2 reactive antibodies, which mediate ADCC activity. This type of data suggests that the anti-idiotypes generally generate IgG1 type antibodies are efficient at ADCC (108) while IgG2, which are considered carbohydrate reactive is minimal at mediating ADCC.

### FUTURE DEVELOPMENT OF STRUCTURE-BASED VACCINES

Glycans or TACAs are important targets for cancer immunotherapy as suggested by immune surveillance mechanisms. TACAs display important biological effects in tumor biology and tumor immunology. Most importantly, the recognition properties of glycans by immune effector cells have suggested translational strategies in immune therapy. In this review, we elaborated on achievements that facilitate rational vaccine design using CMPs. In nature, immunogenic parts of pathogens and cancer cells that provide antigens for B-cell receptors and antigenic peptides that are presentable by MHC molecules to TCR have to be identified. There is much to learn from the B/TCR that see carbohydrates as antigens and then as immunogens. Carbohydrates define recognition patterns, which activate the innate immune system to induce an appropriate adaptive immune response. Regular considerations in using CMPs that are selected upon binding to these receptors have not been pursued in the clinic with much fervor. This is partly due to the perception of utility and the idea that we need "specific" responses to singular carbohydrate antigens. More thought needs to be directed toward rational design approaches, which we have shown can be successfully implemented and not indiscriminate studies of co-crystallization or NMR studies with CMPs derived from random phage screening that are selected biased toward high-affinity binders (13–15).

While there are no universally accepted strategies and tools to rationally design vaccines to elicit antibody responses, vaccines should include B-cell receptor epitopes, but these might be more of a clustered type as we have shown using MAP platforms. Nanoparticle concepts could play a role here if they can be manufactured under GMP clinical grade. MHC Class-I/II molecules process Glycopeptides and so it is thought that these could be incorporated as well. The choice of carbohydrate might also impact on inducing Th1 or Th2 responses. We have shown that naked peptides can do the same however (63, 77). These glycopeptides or naked peptides should display sequences that allow T-cell epitope formation in a complex with MHC molecules but with the realization that there are hundreds of alleles that are differentially combined between individuals. Choosing immunogenic peptides presented by MHC

faces the challenge of not only predicting sequences appropriate for complex with a particular MHC allele, but also finding peptides that can reliably build epitopes in the diverse genetic background within a human population.

The diversity of regulatory mechanisms involving glycans expands the range of possible effects of TACA targeting immunotherapeutic approaches (117). Anti-TACA antibodies, thus, may be involved in more than direct tumor cytotoxicity even though this mechanism is exciting. Although, the exact mechanism may represent a cascade of steps that are still to be established, TACA targeting has the potential to yield anti-tumor effects mediated by Natural Killer cells, which has not been thoroughly investigated in humans even though there is some evidence of therapeutic benefit (118, 119) or through neutralization of tumor immunosuppressive factors in the form of soluble gangliosides (120–122). Future work should clarify the points of involvement of antibody/carbohydrate interactions in modulating tumor growth and facilitating innate surveillance mechanisms.

The number of manuscripts published on CMPs has certainly diminished in recent years. The promise of CMPs 'to be functional in animal models of bacterial infection has been impressive starting with our own (18), yet no CMP for bacterial antigens has made it to the clinic. The same is said for anti-idiotypes of bacterial carbohydrate antigens. In the cancer area only our P10s CMP has made it to the clinic. Perhaps the problem with diminished work on CMPs is more about the perception of mimicry rather than outcomes. The same has been suggested for anti-idiotypes (34). In mouse models, CMPs are functionally relevant much like anti-idiotypes. But clinically, all vaccine types, mimic or not, display less than optimal activity. The focus in immunotherapy has therefore been centered on checkpoints that mediate the immune response. Nevertheless, the promise of CMPs and other surrogates of carbohydrates is to better understand the structural implications of the antibody-mediated interactions that has the potential for innovation in terms of rational design of reagents with biological, chemical, and pharmaceutical applications that underlies concepts of reverse immunology which is highlighted herein.

## REFERENCES

- Hakomori S. Tumor-associated carbohydrate antigens defining tumor malignancy: basis for development of anti-cancer vaccines. *Adv Exp Med Biol* (2001) **491**:369–402. doi:10.1007/978-1-4615-1267-7\_24
- Dong H, Dai H, Hu X, Xiong SD, Gao XM. The (1→6)-beta-glucan moiety represents a cross-reactive epitope of infection-induced malignancy surveillance. *J Immunol* (2014) **192**:1302–12. doi:10.4049/jimmunol.1301495
- Collins PM, Bum-Erdene K, Yu X, Blanchard H. Galectin-3 interactions with glycosphingolipids. *J Mol Biol* (2014) **426**:1439–51. doi:10.1016/j.jmb.2013.12.004
- Ragupathi G. Carbohydrate antigens as targets for active specific immunotherapy. *Cancer Immunol Immunother* (1996) **43**:152–7. doi:10.1007/s002620050316
- Tsai CM. Molecular mimicry of host structures by lipooligosaccharides of *Neisseria meningitidis*: characterization of sialylated and nonsialylated lacto-N-neotetraose (Galbeta1-4GlcNAcbeta1-3Galbeta1-4Glc) structures in lipooligosaccharides using monoclonal antibodies and specific lectins. *Adv Exp Med Biol* (2001) **491**:525–42.
- Mandrell RE. Further antigenic similarities of *Neisseria gonorrhoeae* lipooligosaccharides and human glycosphingolipids. *Infect Immun* (1992) **60**:3017–20.
- Braun JM, Beuth J, Blackwell CC, Giersen S, Higgins PG, Tzanakaki G, et al. *Neisseria meningitidis*, *Neisseria lactamica* and *Moraxella catarrhalis* share cross-reactive carbohydrate antigens. *Vaccine* (2004) **22**:898–908. doi:10.1016/j.vaccine.2003.11.036
- Ragupathi G, Liu NX, Musselli C, Powell S, Lloyd K, Livingston PO. Antibodies against tumor cell glycolipids and proteins, but not mucins, mediate complement-dependent cytotoxicity. *J Immunol* (2005) **174**:5706–12. doi:10.4049/jimmunol.174.9.5706
- Farhan H, Schuster C, Klinger M, Weisz E, Waxenecker G, Schuster M, et al. Inhibition of xenograft tumor growth and down-regulation of ErbB receptors by an antibody directed against Lewis Y antigen. *J Pharmacol Exp Ther* (2006) **319**:1459–66. doi:10.1124/jpet.106.107318
- Doronin II, Vishnyakova PA, Kholodenko IV, Ponomarev ED, Ryazantsev DY, Molotkovskaya IM, et al. Ganglioside GD2 in reception and transduction of cell death signal in tumor cells. *BMC Cancer* (2014) **14**:295. doi:10.1186/1471-2407-14-295
- Guo Z, Wang Q. Recent development in carbohydrate-based cancer vaccines. *Curr Opin Chem Biol* (2009) **13**:608–17. doi:10.1016/j.cbpa.2009.08.010
- Monzavi-Karbassi B, Cunto-Amesty G, Luo P, Kieber-Emmons T. Peptide mimotopes as surrogate antigens of carbohydrates in vaccine discovery. *Trends Biotechnol* (2002) **20**:207–14. doi:10.1016/S0167-7799(02)01940-6
- Clement MJ, Fortune A, Phalipon A, Marcel-Peyre V, Simenel C, Imberty A, et al. Toward a better understanding of the basis of the molecular mimicry of polysaccharide antigens by peptides: the example of *Shigella flexneri* 5a. *J Biol Chem* (2006) **281**:2317–32. doi:10.1074/jbc.M510172200
- Johnson MA, Pinto BM. Structural and functional studies of peptide-carbohydrate mimicry. *Top Curr Chem* (2008) **273**:55–116. doi:10.1007/1128\_2007\_1018
- Agostino M, Sandrin MS, Thompson PE, Farrugia W, Ramsland PA, Yuriev E. Carbohydrate-mimetic peptides: structural aspects of mimicry and therapeutic implications. *Expert Opin Biol Ther* (2011) **11**:211–24. doi:10.1517/14712598.2011.542140
- Oldenburg KR, Loganathan D, Goldstein IJ, Schultz PG, Gallop MA. Peptide ligands for a sugar-binding protein isolated from a random peptide library. *Proc Natl Acad Sci U S A* (1992) **89**:5393–7. doi:10.1073/pnas.89.12.5393
- Scott JK, Loganathan D, Easley RB, Gong X, Goldstein IJ. A family of concanavalin A-binding peptides from a hexapeptide epitope library. *Proc Natl Acad Sci U S A* (1992) **89**:5398–402. doi:10.1073/pnas.89.12.5398
- Westerink MA, Giardina PC, Apicella MA, Kieber-Emmons T. Peptide mimicry of the meningococcal group C capsular polysaccharide. *Proc Natl Acad Sci U S A* (1995) **92**:4021–5. doi:10.1073/pnas.92.9.4021
- Agadjanyan M, Luo P, Westerink MA, Carey LA, Hutchins W, Steplewski Z, et al. Peptide mimicry of carbohydrate epitopes on human immunodeficiency virus. *Nat Biotechnol* (1997) **15**:547–51. doi:10.1038/nbt0697-547
- Kieber-Emmons T, Luo P, Qiu J, Chang TY, O I, Blaszczyk-Thurin M, et al. Vaccination with carbohydrate peptide mimotopes promotes anti-tumor responses. *Nat Biotechnol* (1999) **17**:660–5. doi:10.1038/10870
- Cunto-Amesty G, Dam TK, Luo P, Monzavi-Karbassi B, Brewer CF, Van Cott TC, et al. Directing the immune response to carbohydrate antigens. *J Biol Chem* (2001) **276**:30490–8. doi:10.1074/jbc.M103257200
- Melzer H, Baier K, Felici F, Von Specht BU, Wiedermann G, Kollaritsch H, et al. Humoral immune response against proteophosphoglycan surface antigens of *Entamoeba histolytica* elicited by immunization with synthetic mimotope peptides. *FEMS Immunol Med Microbiol* (2003) **37**:179–83. doi:10.1016/S0928-8244(03)00074-9
- Maita RW, Datta K, Lees A, Belouski SS, Pirofski LA. Immunogenicity and efficacy of *Cryptococcus neoformans* capsular polysaccharide glucuronoxylomannan peptide mimotope-protein conjugates in human immunoglobulin transgenic mice. *Infect Immun* (2004) **72**:196–208. doi:10.1128/IAI.72.1.196-208.2004
- Park I, Choi IH, Kim SJ, Shin JS. Peptide mimotopes of *Neisseria meningitidis* group B capsular polysaccharide. *Yonsei Med J* (2004) **45**:755–8.
- Prinz DM, Smithson SL, Westerink MA. Two different methods result in the selection of peptides that induce a protective antibody response to *Neisseria meningitidis* serogroup C. *J Immunol Methods* (2004) **285**:1–14. doi:10.1016/j.jim.2003.08.005
- Buchwald UK, Lees A, Steinitz M, Pirofski LA. A peptide mimotope of type 8 pneumococcal capsular polysaccharide induces a protective immune response in mice. *Infect Immun* (2005) **73**:325–33. doi:10.1128/IAI.73.1.325-333.2005



27. Monzavi-Karbassi B, Shamloo S, Kieber-Emmons M, Jousheghany F, Luo P, Lin KY, et al. Priming characteristics of peptide mimotopes of carbohydrate antigens. *Vaccine* (2003) **21**:753–60. doi:10.1016/S0264-410X(02)00703-X
28. Luo P, Canziani G, Cunto-Amesty G, Kieber-Emmons T. A molecular basis for functional peptide mimicry of a carbohydrate antigen. *J Biol Chem* (2000) **275**:16146–54. doi:10.1074/jbc.M909121199
29. Monzavi-Karbassi B, Hennings LJ, Artaud C, Liu T, Jousheghany F, Pashov A, et al. Preclinical studies of carbohydrate mimetic peptide vaccines for breast cancer and melanoma. *Vaccine* (2007) **25**:3022–31. doi:10.1016/j.vaccine.2007.01.072
30. Van Regenmortel MH. Transcending the structuralist paradigm in immunology-affinity and biological activity rather than purely structural considerations should guide the design of synthetic peptide epitopes. *Biomed Pept Proteins Nucleic Acids* (1995) **1**:109–16.
31. Van Regenmortel MH. Molecular design versus empirical discovery in peptide-based vaccines. Coming to terms with fuzzy recognition sites and ill-defined structure-function relationships in immunology. *Vaccine* (1999) **18**:216–21. doi:10.1016/S0264-410X(99)00192-9
32. Van Regenmortel MH. Antigenicity and immunogenicity of synthetic peptides. *Biologicals* (2001) **29**:209–13. doi:10.1006/biol.2001.0308
33. Oldstone MB. Molecular mimicry: its evolution from concept to mechanism as a cause of autoimmune diseases. *Mol Immunol* (2014) **2**:2. doi:10.1089/mab.2013.0090
34. Kieber-Emmons T, Monzavi-Karbassi B, Pashov A, Saha S, Murali R, Kohler H. The promise of the anti-idiotypic concept. *Front Oncol* (2012) **2**:196. doi:10.3389/fonc.2012.00196
35. Jacobsen JT, Lunde E, Sundvold-Gjerstad V, Munthe LA, Bogen B. The cellular mechanism by which complementary Id+ and anti-Id antibodies communicate: T cells integrated into idiotype regulation. *Immunol Cell Biol* (2010) **88**:515–22. doi:10.1038/icb.2009.118
36. Foon KA, Lutzky J, Baral RN, Yannelli JR, Hutchins L, Teitelbaum A, et al. Clinical and immune responses in advanced melanoma patients immunized with an anti-idiotypic antibody mimicking disialoganglioside GD2. *J Clin Oncol* (2000) **18**:376–84.
37. Chapman PB. Vaccinating against GD3 ganglioside using BEC2 anti-idiotypic monoclonal antibody. *Curr Opin Invest Drugs* (2003) **4**:710–5.
38. Chapman PB, Wu D, Ragupathi G, Lu S, Williams L, Hwu WJ, et al. Sequential immunization of melanoma patients with GD3 ganglioside vaccine and anti-idiotypic monoclonal antibody that mimics GD3 ganglioside. *Clin Cancer Res* (2004) **10**:4717–23. doi:10.1158/1078-0432.CCR-04-0345
39. Alfonso S, Diaz RM, De La Torre A, Santiesteban E, Aguirre F, Perez K, et al. IE10 anti-idiotypic vaccine in non-small cell lung cancer: experience in stage IIIB/IV patients. *Cancer Biol Ther* (2007) **6**:1847–52. doi:10.4161/cbt.6.12.5000
40. Fernandez LE, Gabri MR, Guthmann MD, Gomez RE, Gold S, Fainboim L, et al. NGGGM3 ganglioside: a privileged target for cancer vaccines. *Clin Dev Immunol* (2010) **2010**:814397. doi:10.1155/2010/814397
41. Hoffmuller U, Knaute T, Hahn M, Hohne W, Schneider-Mergener J, Kramer A. Evolutionary transition pathways for changing peptide ligand specificity and structure. *EMBO J* (2000) **19**:4866–74. doi:10.1093/emboj/19.18.4866
42. Coletta A, Pinney JW, Solis DY, Marsh J, Pettifer SR, Attwood TK. Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst Biol* (2010) **4**:43. doi:10.1186/1752-0509-1184-1143
43. Bonsor DA, Grishkovskaya I, Dodson EJ, Kleanthous C. Molecular mimicry enables competitive recruitment by a natively disordered protein. *J Am Chem Soc* (2007) **129**:4800–7. doi:10.1021/ja070153n
44. Notkins AL. Polyreactivity of antibody molecules. *Trends Immunol* (2004) **25**:174–9. doi:10.1016/j.it.2004.02.004
45. Atlan H, Cohen IR. Immune information, self-organization and meaning. *Int Immunol* (1998) **10**:711–7. doi:10.1093/intimm/10.6.711
46. Mason D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol Today* (1998) **19**:395–404. doi:10.1016/S0167-5699(98)01299-7
47. Dimitrov JD, Ivanovska ND, Lacroix-Desmazes S, Doltchinkova VR, Kaveri SV, Vassilev TL. Ferrous ions and reactive oxygen species increase antigen-binding and anti-inflammatory activities of immunoglobulin G. *J Biol Chem* (2006) **281**:439–46. doi:10.1074/jbc.M509190200
48. McIntyre JA, Faulk WP. Redox-reactive autoantibodies: biochemistry, characterization, and specificities. *Clin Rev Allergy Immunol* (2009) **37**:49–54. doi:10.1007/s12016-008-8093-y
49. Dimitrov JD, Pashov A, Vassilev T. Antibody polyspecificity: what does it matter? In: Lutz HU, editor. *Naturally Occurring Antibodies (NABs)*. Austin, TX: Springer (2012). 268 p.
50. Matzinger P. The danger model: a renewed sense of self. *Science* (2002) **296**:301–5. doi:10.1126/science.1071059
51. Medzhitov R, Janeway CA Jr. Decoding the patterns of self and nonself by the innate immune system. *Science* (2002) **296**:298–300. doi:10.1126/science.1068883
52. Forrest S, Javornik B, Smith RE, Perelson AS. Using genetic algorithms to explore pattern recognition in the immune system. *Evol Comput* (1993) **1**:191–211. doi:10.1162/evco.1993.1.3.191
53. Kearney JF. Immune recognition of OxLDL in atherosclerosis. *J Clin Invest* (2000) **105**:1683–5. doi:10.1172/JCI10426
54. Roozendaal R, Carroll MC. Emerging patterns in complement-mediated pathogen recognition. *Cell* (2006) **125**:29–32. doi:10.1016/j.cell.2006.03.018
55. Quinn A, Adderson EE, Shackelford PG, Carroll WL, Cunningham MW. Autoantibody germ-line gene segment encodes VH and VL regions of a human anti-streptococcal monoclonal antibody recognizing streptococcal M protein and human cardiac myosin epitopes. *J Immunol* (1995) **154**:4203–12.
56. Apostolopoulos V, Osinski C, McKenzie IF. MUC1 cross-reactive Gal alpha(1,3)Gal antibodies in humans switch immune responses from cellular to humoral. *Nat Med* (1998) **4**:315–20. doi:10.1038/nm0398-315
57. Pashov A, Canziani G, Macleod S, Plaxco J, Monzavi-Karbassi B, Kieber-Emmons T. Targeting carbohydrate antigens in HIV vaccine development. *Vaccine* (2005) **23**:2168–75. doi:10.1016/j.vaccine.2005.01.045
58. Moore CJ, Auzanneau FI. Understanding the recognition of Lewis X by anti-Le(x) monoclonal antibodies. *J Med Chem* (2013) **56**:8183–90. doi:10.1021/jm401304h
59. Yang L, Adam C, Nichol GS, Cockroft SL. How much do van der Waals dispersion forces contribute to molecular recognition in solution? *Nat Chem* (2013) **5**:1006–10. doi:10.1038/nchem.1779
60. Bostrom J, Haber L, Koenig P, Kelley RF, Fuh G. High affinity antigen recognition of the dual specific variants of herceptin is entropy-driven in spite of structural plasticity. *PLoS One* (2011) **6**:e17887. doi:10.1371/journal.pone.0017887
61. Riemer AB, Jensen-Jarolim E. Mimotope vaccines: epitope mimics induce anti-cancer antibodies. *Immunol Lett* (2007) **113**:1–5. doi:10.1016/j.imlet.2007.07.008
62. Cunto-Amesty G, Luo P, Monzavi-Karbassi B, Lees A, Kieber-Emmons T. Exploiting molecular mimicry to broaden the immune response to carbohydrate antigens for vaccine development. *Vaccine* (2001) **19**:2361–8. doi:10.1016/S0264-410X(00)00527-2
63. Monzavi-Karbassi B, Cunto-Amesty G, Luo P, Shamloo S, Blaszczyk-Thurin M, Kieber-Emmons T. Immunization with a carbohydrate mimicking peptide augments tumor-specific cellular responses. *Int Immunol* (2001) **13**:1361–71. doi:10.1093/intimm/13.11.1361
64. Hennings L, Artaud C, Jousheghany F, Monzavi-Karbassi B, Pashov A, Kieber-Emmons T. Carbohydrate mimetic peptides augment carbohydrate-reactive immune responses in the absence of immune pathology. *Cancers (Basel)* (2011) **3**:4151–69. doi:10.3390/cancers3044151
65. Grothaus MC, Srivastava N, Smithson SL, Kieber-Emmons T, Williams DB, Carlone GM, et al. Selection of an immunogenic peptide mimic of the capsular polysaccharide of *Neisseria meningitidis* serogroup A using a peptide display library. *Vaccine* (2000) **18**:1253–63. doi:10.1016/S0264-410X(99)00390-4
66. Fleuridor R, Lees A, Pirofski L. A cryptococcal capsular polysaccharide mimotope prolongs the survival of mice with *Cryptococcus neoformans* infection. *J Immunol* (2001) **166**:1087–96. doi:10.4049/jimmunol.166.2.1087
67. Cunto-Amesty G, Luo P, Monzavi-Karbassi B, Lees A, Alexander J, Del Guercio MF, et al. Peptide mimotopes as prototypic templates of broad-spectrum surrogates of carbohydrate antigens. *Cell Mol Biol (Noisy-le-grand)* (2003) **49**:245–54.
68. Haurum JS, Arsequell G, Lellouch AC, Wong SY, Dwek RA, McMichael AJ, et al. Recognition of carbohydrate by major histocompatibility complex class I-restricted, glycopeptide-specific cytotoxic T lymphocytes. *J Exp Med* (1994) **180**:739–44. doi:10.1084/jem.180.2.739
69. Jensen T, Galli-Stampino L, Mouritsen S, Frische K, Peters S, Meldal M, et al. T cell recognition of Tn-glycosylated peptide antigens. *Eur J Immunol* (1996) **26**:1342–9. doi:10.1002/eji.1830260625

70. Galli-Stampino L, Meinjohanns E, Frische K, Meldal M, Jensen T, Werdelin O, et al. T-cell recognition of tumor-associated carbohydrates: the nature of the glycan moiety plays a decisive role in determining glycopeptide immunogenicity. *Cancer Res* (1997) **57**:3214–22.
71. Zhao XJ, Cheung NK. GD2 oligosaccharide: target for cytotoxic T lymphocytes. *J Exp Med* (1995) **182**:67–74. doi:10.1084/jem.182.1.67
72. Glithero A, Tormo J, Haurum JS, Arseque G, Valencia G, Edwards J, et al. Crystal structures of two H-2Db/glycopeptide complexes suggest a molecular basis for CTL cross-reactivity. *Immunity* (1999) **10**:63–74. doi:10.1016/S1074-7613(00)80007-2
73. Speir JA, Abdel-Motal UM, Jondal M, Wilson IA. Crystal structure of an MHC class I presented glycopeptide that generates carbohydrate-specific CTL. *Immunity* (1999) **10**:51–61. doi:10.1016/S1074-7613(00)80006-0
74. Abdel-Motal UM, Berg L, Rosen A, Bengtsson M, Thorpe CJ, Kihlberg J, et al. Immunization with glycosylated Kb-binding peptides generates carbohydrate-specific, unrestricted cytotoxic T cells. *Eur J Immunol* (1996) **26**:544–51. doi:10.1002/eji.1830260307
75. Tzianabos AO, Finberg RW, Wang Y, Chan M, Onderdonk AB, Jennings HJ, et al. T cells activated by zwitterionic molecules prevent abscesses induced by pathogenic bacteria. *J Biol Chem* (2000) **275**:6733–40. doi:10.1074/jbc.275.10.6733
76. Kieber-Emmons T, Monzavi-Karbassi B, Wang B, Luo P, Weiner DB. Cutting edge: DNA immunization with minigenes of carbohydrate mimotopes induce functional anti-carbohydrate antibody response. *J Immunol* (2000) **165**:623–7. doi:10.1049/jimmunol.165.2.623
77. Monzavi-Karbassi B, Luo P, Jousheghany F, Torres-Quinones M, Cunto-Amesty G, Artaud C, et al. A mimic of tumor rejection antigen-associated carbohydrates mediates an antitumor cellular response. *Cancer Res* (2004) **64**:2162–6. doi:10.1158/0008-5472.CAN-03-1532
78. Harris SL, Dagtas AS, Diamond B. Regulating the isotypic and idiotypic profile of an anti-PC antibody response: lessons from peptide mimics. *Mol Immunol* (2002) **39**:263–72. doi:10.1016/S0161-5890(02)00116-5
79. Talavera A, Eriksson A, Okvist M, Lopez-Requena A, Fernandez-Marrero Y, Perez R, et al. Crystal structure of an anti-ganglioside antibody, and modelling of the functional mimicry of its NeuGc-GM3 antigen by an anti-idiotypic antibody. *Mol Immunol* (2009) **46**:3466–75. doi:10.1016/j.molimm.2009.07.032
80. Monzavi-Karbassi B, Artaud C, Jousheghany F, Hennings L, Carcel-Trullols J, Shaaf S, et al. Reduction of spontaneous metastases through induction of carbohydrate cross-reactive apoptotic antibodies. *J Immunol* (2005) **174**:7057–65. doi:10.1049/jimmunol.174.11.7057
81. Vollmers HP, Brandlein S. Death by stress: natural IgM-induced apoptosis. *Methods Find Exp Clin Pharmacol* (2005) **27**:185–91. doi:10.1358/mf.2005.27.3.890876
82. Vollmers HP, Brandlein S. Natural antibodies and cancer. *N Biotechnol* (2009) **25**:294–8. doi:10.1016/j.nbt.2009.03.016
83. Hoess R, Brinkmann U, Handel T, Pastan I. Identification of a peptide which binds to the carbohydrate-specific monoclonal antibody B3. *Gene* (1993) **128**:43–9. doi:10.1016/0378-1119(93)90151-R
84. Jain D, Kaur KJ, Salunke DM. Enhanced binding of a rationally designed peptide ligand of concanavalin A arises from improved geometrical complementarity. *Biochemistry* (2001) **40**:12059–66. doi:10.1021/bi011254f
85. Pashov A, Perry M, Dyar M, Chow M, Kieber-Emmons T. Carbohydrate mimotopes in the rational design of cancer vaccines. *Curr Top Med Chem* (2005) **5**:1171–85. doi:10.2174/156802605774370928
86. Pashov A, Perry M, Dyar M, Chow M, Kieber-Emmons T. Defining carbohydrate antigens as HIV vaccine candidates. *Curr Pharm Des* (2007) **13**:185–201. doi:10.2174/13816120779313678
87. Agostino M, Sandrin MS, Thompson PE, Ramsland PA, Yuriev E. Peptide inhibitors of xenoreactive antibodies mimic the interaction profile of the native carbohydrate antigens. *Biopolymers* (2011) **96**:193–206. doi:10.1002/bip.21427
88. Agostino M, Yuriev E, Ramsland PA. Antibody recognition of cancer-related gangliosides and their mimics investigated using in silico site mapping. *PLoS One* (2012) **7**:e35457. doi:10.1371/journal.pone.0035457
89. Qiu J, Luo P, Wasmund K, Steplewski Z, Kieber-Emmons T. Towards the development of peptide mimotopes of carbohydrate antigens as cancer vaccines. *Hybridoma* (1999) **18**:103–12. doi:10.1089/hyb.1999.18.103
90. Wondimu A, Zhang T, Kieber-Emmons T, Gimotty P, Sproesser K, Somasundaram R, et al. Peptides mimicking GD2 ganglioside elicit cellular, humoral and tumor-protective immune responses in mice. *Cancer Immunol Immunother* (2008) **57**:1079–89. doi:10.1007/s00262-007-0439-4
91. Hurwitz AA, Ji Q. Autoimmune depigmentation following sensitization to melanoma antigens. *Methods Mol Med* (2004) **102**:421–7.
92. Galili U. Anti-Gal: an abundant human natural antibody of multiple pathogenesis and clinical benefits. *Immunology* (2013) **140**:1–11. doi:10.1111/imm.12110
93. Preiss S, Kammertoens T, Lampert C, Willmsky G, Blankenstein T. Tumor-induced antibodies resemble the response to tissue damage. *Int J Cancer* (2005) **115**:456–62. doi:10.1002/ijc.20914
94. Kirkeby S, Moe D. Lectin interactions with alpha-galactosylated xenoantigens. *Xenotransplantation* (2002) **9**:260–7. doi:10.1034/j.1399-3089.2002.01078.x
95. Korourian S, Siegel E, Kieber-Emmons T, Monzavi-Karbassi B. Expression analysis of carbohydrate antigens in ductal carcinoma in situ of the breast by lectin histochemistry. *BMC Cancer* (2008) **8**:136. doi:10.1186/1471-2407-1188-1136
96. Zuckier LS, Berkowitz EZ, Sattenberg RJ, Zhao QH, Deng HF, Scharff MD. Influence of affinity and antigen density on antibody localization in a modifiable tumor targeting model. *Cancer Res* (2000) **60**:7008–13.
97. Takahashi T, Johnson TD, Nishinaka Y, Morton DL, Irie RE. IgM anti-ganglioside antibodies induced by melanoma cell vaccine correlate with survival of melanoma patients. *J Invest Dermatol* (1999) **112**:205–9. doi:10.1046/j.1523-1747.1999.00493.x
98. Hermesen BB, Verheijen RH, Menko FH, Gille JJ, Van Uffelen K, Blankenstein MA, et al. Humoral immune responses to MUC1 in women with a BRCA1 or BRCA2 mutation. *Eur J Cancer* (2007) **43**:1556–63. doi:10.1016/j.ejca.2007.04.007
99. Amigorena S. Fc gamma receptors and cross-presentation in dendritic cells. *J Exp Med* (2002) **195**:F1–3. doi:10.1084/jem.20011925
100. Dhodapkar KM, Krasovsky J, Williamson B, Dhodapkar MV. Antitumor monoclonal antibodies enhance cross-presentation of cellular antigens and the generation of myeloma-specific killer T cells by dendritic cells. *J Exp Med* (2002) **195**:125–33. doi:10.1084/jem.20011097
101. Wernersson S, Kleinau S, Heyman B. Immune complex-mediated enhancement of antibody responses without induction of delayed-type hypersensitivity. *Scand J Immunol* (2000) **52**:563–9. doi:10.1046/j.1365-3083.2000.00813.x
102. Forster-Waldl E, Riemer AB, Dehof AK, Neumann D, Bramswig K, Boltz-Nitulescu G, et al. Isolation and structural analysis of peptide mimotopes for the disialoganglioside GD2, a neuroblastoma tumor antigen. *Mol Immunol* (2005) **42**:319–25. doi:10.1016/j.molimm.2004.07.011
103. Riemer AB, Forster-Waldl E, Bramswig KH, Pollak A, Zielinski CC, Pehamberger H, et al. Induction of IgG antibodies against the GD2 carbohydrate tumor antigen by vaccination with peptide mimotopes. *Eur J Immunol* (2006) **36**:1267–74. doi:10.1002/eji.200535279
104. Horwacik I, Czaplicki D, Talarek K, Kowalczyk A, Bolesta E, Kozbor D, et al. Selection of novel peptide mimics of the GD2 ganglioside from a constrained phage-displayed peptide library. *Int J Mol Med* (2007) **19**:829–39.
105. Horwacik I, Kurcinski M, Bzowska M, Kowalczyk AK, Czaplicki D, Kolinski A, et al. Analysis and optimization of interactions between peptides mimicking the GD2 ganglioside and the monoclonal antibody 14G2a. *Int J Mol Med* (2011) **28**:47–57. doi:10.3892/ijmm.2011.3655
106. Willers J, Lucchese A, Kanduc D, Ferrone S. Molecular mimicry of phage displayed peptides mimicking GD3 ganglioside. *Peptides* (1999) **20**:1021–6. doi:10.1016/S0196-9781(99)00095-9
107. O I, Kieber-Emmons T, Otvos L Jr, Blaszczyk-Thurin M. Peptides mimicking sialyl-Lewis X isolated from a random peptide library and peptide array. *Ann N Y Acad Sci* (1999) **886**:276–9. doi:10.1111/j.1749-6632.1999.tb09435.x
108. Foon KA, Sen G, Hutchins L, Kshala OL, Baral R, Banerjee M, et al. Antibody responses in melanoma patients immunized with an anti-idiotypic antibody mimicking disialoganglioside GD2. *Clin Cancer Res* (1998) **4**:1117–24.
109. Cheung NK, Guo HF, Heller G, Cheung IY. Induction of Ab3 and Ab3' antibody was associated with long-term survival after anti-G(D2) antibody therapy of stage 4 neuroblastoma. *Clin Cancer Res* (2000) **6**:2653–60.
110. Zeytin HE, Tripathi PK, Bhattacharya-Chatterjee M, Foon KA, Chatterjee SK. Construction and characterization of DNA vaccines encoding the single-chain variable fragment of the anti-idiotypic antibody 1A7 mimicking the tumor-associated antigen disialoganglioside GD2. *Cancer Gene Ther* (2000) **7**:1426–36. doi:10.1038/sj.cgt.7700240
111. Lode HN, Schmidt M, Seidel D, Huebener N, Brackrock D, Bleeke M, et al. Vaccination with anti-idiotypic antibody ganglidiomab mediates a GD(2)-specific

- anti-neuroblastoma immune response. *Cancer Immunol Immunother* (2013) **62**:999–1010. doi:10.1007/s00262-013-1413-y
112. Bottomley A, Debruyne C, Felip E, Millward M, Thiberville L, D'Addario G, et al. Symptom and quality of life results of an international randomised phase III study of adjuvant vaccination with Bec2/BCG in responding patients with limited disease small-cell lung cancer. *Eur J Cancer* (2008) **44**:2178–84. doi:10.1016/j.ejca.2008.06.036
  113. Hernandez AM, Rodriguez N, Gonzalez JE, Reyes E, Rondon T, Grinan T, et al. Anti-NeuGcGM3 antibodies, actively elicited by idiotype vaccination in nonsmall cell lung cancer patients, induce tumor cell death by an oncosis-like mechanism. *J Immunol* (2011) **186**:3735–44. doi:10.4049/jimmunol.1000609
  114. Hernandez AM, Toledo D, Martinez D, Grinan T, Brito V, Macias A, et al. Characterization of the antibody response against NeuGcGM3 ganglioside elicited in non-small cell lung cancer patients immunized with an anti-idiotypic antibody. *J Immunol* (2008) **181**:6625–34. doi:10.4049/jimmunol.181.9.6625
  115. Chapman PB, Williams L, Salibi N, Hwu WJ, Krown SE, Livingston PO. A phase II trial comparing five dose levels of BEC2 anti-idiotypic monoclonal antibody vaccine that mimics GD3 ganglioside. *Vaccine* (2004) **22**:2904–9. doi:10.1016/j.vaccine.2003.12.028
  116. McCaffery M, Yao TJ, Williams L, Livingston PO, Houghton AN, Chapman PB. Immunization of melanoma patients with BEC2 anti-idiotypic monoclonal antibody that mimics GD3 ganglioside: enhanced immunogenicity when combined with adjuvant. *Clin Cancer Res* (1996) **2**:679–86.
  117. Vazquez AM, Rodreguez-Zhurbenko N, Lopez AM. Anti-ganglioside anti-idiotypic vaccination: more than molecular mimicry. *Front Oncol* (2012) **2**:170. doi:10.3389/fonc.2012.00170
  118. Koehn TA, Trimble LL, Alderson KL, Erbe AK, McDowell KA, Grzywacz B, et al. Increasing the clinical efficacy of NK and antibody-mediated cancer immunotherapy: potential predictors of successful clinical outcome based on observations in high-risk neuroblastoma. *Front Pharmacol* (2012) **3**:91. doi:10.3389/fphar.2012.00091
  119. Tarek N, Le Ludec JB, Gallagher MM, Zheng J, Venstrom JM, Chamberlain E, et al. Unlicensed NK cells target neuroblastoma following anti-GD2 antibody treatment. *J Clin Invest* (2012) **122**:3260–70. doi:10.1172/JCI62749
  120. Perez CA, Ravindranath MH, Soh D, Gonzales A, Ye W, Morton DL. Serum anti-ganglioside IgM antibodies in soft tissue sarcoma: clinical prognostic implications. *Cancer J* (2002) **8**:384–94. doi:10.1097/00130404-200209000-00009
  121. Ravindranath MH, Muthugounder S, Presser N, Ye X, Brosman S, Morton DL. Endogenous immune response to gangliosides in patients with confined prostate cancer. *Int J Cancer* (2005) **116**:368–77. doi:10.1002/ijc.21023
  122. Ravindranath MH, Muthugounder S, Hannah MR, Morton DL. Significance of endogenous augmentation of antiganglioside IgM in cancer patients: potential tool for early detection and management of cancer therapy. *Ann N Y Acad Sci* (2007) **1107**:212–22. doi:10.1196/annals.1381.023

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 April 2014; accepted: 17 June 2014; published online: 30 June 2014.

Citation: Kieber-Emmons T, Saha S, Pashov A, Monzavi-Karbassi B and Murali R (2014) Carbohydrate-mimetic peptides for pan anti-tumor responses. *Front. Immunol.* **5**:308. doi: 10.3389/fimmu.2014.00308

This article was submitted to Immunotherapies and Vaccines, a section of the journal *Frontiers in Immunology*.

Copyright © 2014 Kieber-Emmons, Saha, Pashov, Monzavi-Karbassi and Murali. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Predicting the origins of anti-blood group antibody specificity: a case study of the ABO A- and B-antigens

Spandana Makeneni<sup>1</sup>, Ye Ji<sup>1</sup>, David C. Watson<sup>2</sup>, N. Martin Young<sup>2</sup> and Robert J. Woods<sup>1,3\*</sup>

<sup>1</sup> Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA

<sup>2</sup> Human Health Therapeutics, National Research Council Canada, Ottawa, ON, Canada

<sup>3</sup> School of Chemistry, National University of Ireland, Galway, Ireland

## Edited by:

Elizabeth Yuriev, Monash University, Australia

## Reviewed by:

Jessica Kate Holien, St. Vincents Institute of Medical Research, Australia

Ricardo Mancera, Curtin University, Australia

## \*Correspondence:

Robert J. Woods, Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA 30602, USA  
e-mail: rwoods@ccrc.uga.edu

The ABO blood group system is the most important blood type system in human transfusion medicine. Here, we explore the specificity of antibody recognition toward ABO blood group antigens using computational modeling and biolayer interferometry. Automated docking and molecular dynamics simulations were used to explore the origin of the specificity of an anti-blood group A antibody variable fragment (Fv AC1001). The analysis predicts a number of Fv-antigen interactions that contribute to affinity, including a hydrogen bond between a His<sup>L49</sup> and the carbonyl moiety of the GalNAc in antigen A. This interaction was consistent with the dependence of affinity on pH, as measured experimentally; at lower pH there is an increase in binding affinity. Binding energy calculations provide unique insight into the origin of interaction energies at a per-residue level in both the scFv and the trisaccharide antigen. The calculations indicate that while the antibody can accommodate both blood group A and B antigens in its combining site, the A antigen is preferred by 4 kcal/mol, consistent with the lack of binding observed for the B antigen.

**Keywords:** molecular docking, MD simulations, blood group antigens, antibody specificity, GLYCAM, AMBER

## INTRODUCTION

Since its discovery in 1900 (1), the ABO blood group system has played a crucial role in defining human blood and tissue compatibility. The blood type of an individual indicates the presence or absence of relevant antigens and antibodies. The three blood types share a core oligosaccharide antigen (H), and based on the glycosyl transferases inherited, different antigens are synthesized (2–4); type A transferase adds a terminal non-reducing *N*-acetylgalactosamine (GalNAc) residue; type B transferase adds galactose (Gal), whereas individuals with blood group O retain the unmodified H antigen. During the first years of life, the immune system forms antibodies upon exposure to non-self antigens from various exogenous factors. Thus an A-type individual will have circulating antibodies specific for the B-antigen, and vice-versa. The high degree of specificity is notable given that the only difference between the structures of the A- and B-antigens is the replacement of an acetamido moiety (in A) with a hydroxyl group (in B). Because of the presence of circulating antibodies, a mismatched blood transfusion or organ transplant can lead to hyperacute immune response and death (5, 6). Additionally, under certain circumstances, incompatibilities in blood groups between mother and child can trigger the mother's immune system to produce antibodies against the fetus, causing hemolytic disease (7).

Alterations in the structures of the ABO antigens often occur during carcinogenesis and therefore they have also been considered tumor markers (8, 9). Recently, strong correlations have been established between the presence of particular ABO and Lewis antigens and susceptibility to infectious diseases, such as *Helicobacter pylori*, norovirus, and cholera (10), wherein the blood group antigens can be exploited as receptors for bacterial and

viral adhesion. Conversely, it has been suggested that endogenous anti-blood group antibodies can recognize blood-group-like carbohydrate antigens on pathogen surfaces, conferring protection against infection (11).

Despite their clinical importance, relatively little is known about the structural basis for these highly specific antibodies–antigen interactions. Although X-ray crystallography has been used to characterize antibody–carbohydrate complexes, the generally enhanced flexibility and conformational heterogeneity of oligosaccharides detracts from the ability to generate co-crystals (12). Additionally, anti-carbohydrate antibodies bind to their antigens with an affinity that is 3–5 orders of magnitude lower than typical antibodies that bind to protein or peptide antigens. Difficulties in generating 3D structures for carbohydrate–antibody complexes have led to the increasing use of theoretical structure prediction methods (13, 14), which, while convenient, are prone to predicting false positives due to inaccuracies in pose scoring functions (15) and to the omission of carbohydrate conformational preferences (16).

In this study, we examined the structural origin of the antigenicity (the specificity and affinity) of a monoclonal antibody raised against blood group A (BGA) antigen, for which an apo structure of the single-chain variable fragment (scFv AC1001) has been reported (17). The specificity data from screening two independent glycan arrays [Consortium for Functional Glycomics (v4.0, request ID: 1808) and from the group of Jeff Gildersleeve] confirmed that the scFv displayed no detectable binding to any B-antigens and only bound to BGA-containing glycans. To provide a structural interpretation for the specificity of AC1001 for BGA over blood groups H (BGH) and B (BGB), we generated a

3D model of the immune complex using molecular docking and refined it by molecular dynamics (MD) simulation. Despite its limitations, molecular docking, with or without additional experimental constraints, such as from NMR data, is often the only approach that may be employed to generate the structure of a ligand–protein complex, in the absence of direct crystallographic data. To enhance the success rate, a recent carbohydrate conformational energy function (16) was employed with AutoDock VINA (18), which quantifies the conformational preferences of oligosaccharides based on their glycosidic torsion angles. MD simulations (50 ns) were subsequently performed to ensure that the docked complexes were stable under physically realistic conditions, and in that event, the MD data were employed in binding free energy calculations. A particular advantage of MD-based energy calculations is that they provide statistically converged values that may be partitioned into contributions from individual residues in the protein and ligand (19).

## MATERIALS AND METHODS

### CLONING, EXPRESSION, AND PURIFICATION OF scFv

An scFv gene containing a short linker (RADAA) and the Leu 103H Val mutation (17), with a His<sub>6</sub> tag, was assembled by PCR and cloned into the phagemid pSK4. The construct was maintained in *Escherichia coli* TG1 cells. Cells from positive clones, as judged by DNA sequence analysis, were grown in minimal media, induced, and subjected to periplasmic extraction. The scFv dimer was purified from the extract by Ni<sup>2+</sup> immobilized metal affinity chromatography, by elution with an imidazole gradient.

### BIOLAYER INTERFEROMETRY

Affinity measurements were performed on a biolayer interferometer (Octet Red96, ForteBio). Data were processed using the Data Acquisition and Analysis 8.0 software (ForteBio), and kinetic binding constants were determined from a 1:1 binding model using the OriginPro software (OriginLab). The scFv was immobilized on an amine reactive second-generation (AR2G) biosensor (Lot No. 1311212, ForteBio). The BGA trisaccharide was analyzed as the conjugate to bovine serum albumin (BSA–BGA) and was dissolved in an analysis buffer containing 10 mM HEPES, 150 mM NaCl, 3.4 mM EDTA, and 0.005% Tween 20 at a range of pH values (5, 5.5, 6, 6.5, and 7). A BSA–Le<sup>x</sup> trisaccharide conjugate (Prod. No. NGP0302, V-Labs, Inc.) and BSA (Prod. No. 23209, Pierce Thermo Scientific, Rockford, IL, USA) were used as negative controls. Details of the biolayer interferometry (BLI) conditions are provided in Supplemental Material.

### AUTOMATED DOCKING

Docking was performed using AutoDock VINA (18) with 20 docked poses generated for each experiment. The protein and the ligand files were prepared using Autodock tools (ADT) (20) with Gassteiger (21) partial atomic charges assigned to both the protein and ligand residues. The crystal structure of the scFv (PDB ID: 1JV5) was employed, together with a 3D structure of BGA obtained from the GLYCAM-Web server ([www.glycam.org](http://www.glycam.org)). Crystal waters were removed prior to docking and hydrogen atoms were added to the protein using ADT, whereas hydrogen atoms in the ligand were assigned from the GLYCAM residue templates. The glycosidic  $\phi$  and  $\psi$  torsion angles were allowed to

be flexible during docking, as were all the hydroxyl groups. The protein was maintained rigid. The docking grid box (dimensions: 26.25 Å × 26.25 Å × 37.5 Å) was centered relative to the complementarity determining regions (CDRs) of the antibody as described previously (16). For the mutational-docking approach, Trp<sup>H100</sup> was mutated to Ala by deleting the side-chain atoms of the Trp residue in the crystal structure, followed by processing with the tleap module in AMBER (22). Ala<sup>H100</sup> was reverted back to Trp by restoring the crystal coordinates of the side chain of Trp<sup>H100</sup>. The docked poses from the mutational approach were filtered based on the clashes with the reverted Trp. Poses in which the clashes could not be eliminated by implicit energy minimization (details are in the “MD simulations” section) were rejected. Ligand conformations of all the docked poses from both the flexible and mutational-docking approaches were scored using the recently reported carbohydrate intrinsic (CHI) energy scoring function (16). Any conformations with total CHI-energies >5 kcal/mol were rejected. The BGB complex was generated directly from that generated for BGA by simple replacement of the NAc group by an OH group.

### MD SIMULATIONS

All the MD simulations were performed with the GPU implementation of the pmed code, pmed.cud\_SPDP (23), from AMBER12 (22). The calculations employed the ff99SSB (24) parameters for the protein and the GLYCAM06h (25) parameters for the carbohydrate. For the BGA, BGB–scFv complex simulations, an implicit solvent energy minimization (5000 steps of steepest descent followed by 5000 steps of conjugate gradient), were performed to optimize the side-chain positions of the reverted Trp residue. During this minimization, the backbone atoms of the framework regions were restrained with a 5 kcal/mol Å<sup>2</sup> while the CDRs and the ligand were allowed to be flexible. The systems were then solvated in a cubic water box [120 Å per side, with a TIP3P water (26)]. Each system was energy minimized using explicit solvent (10,000 steps of steepest descent, 10,000 steps of conjugate gradient). During this energy minimization, the protein residues were restrained with a force constant of 100 kcal/mol Å<sup>2</sup> allowing only the solvent and ligand to relax. This minimization was followed by heating from 5 to 300 K over the course of 50 ps at constant volume. Production MD simulations were performed for 50 ns at constant pressure (NPT ensemble) with the temperature held constant at 300 K using a Langevin thermostat. During the heating and the production MD, the backbone atoms of the protein were restrained with a force constant of 5 kcal/mol Å<sup>2</sup>, with the protein side chains and ligand atoms allowed to be flexible. The backbone atoms were restrained in order to ensure that the protein fold remained stable during the course of the simulation. For the BGA trisaccharide MD simulation, the system was solvated in a cubic water box (120 Å per side, with a TIP3P water) and energy minimized using explicit solvent (5000 steps of steepest descent, 5000 steps of conjugate gradient). This was followed by heating from 5 to 300 K for a period of 50 ps at constant volume. Production MD simulations were performed for 50 ns at constant pressure (NPT). During the minimization, heating, and production MD simulations, there were no restraints placed on the trisaccharide. For both BGA, BGB–scFv complexes and BGA



trisaccharide simulations, all covalent bonds involving hydrogen atoms were constrained using the SHAKE (27) algorithm, allowing a time step of 2 fs. A non-bonded cut-off of 8 Å was used and long-range electrostatics were employed using the particle mesh Ewald (PME) method (28). Snapshots were collected at 1 ps intervals for subsequent analysis.

## ANALYSIS

The stability of the complexes was assessed by monitoring the root-mean-squared-displacement (RMSD) of the ligand position, the glycosidic torsion angles, the ring conformations, and the protein–ligand hydrogen bonds. All these values except for the ring conformation analysis were generated using the ptraj module of AMBERTOOLS 12 (29). Ligand RMSD values were calculated for the ring atoms, relative to the first time step of the simulation. Hydrogen bond interactions between the protein and the ligand were measured with distance and angle cut-off values of 3.5 Å and 120°, respectively. The ring conformations of each individual residue in the ligand during the course of simulation were analyzed using the recently reported BFMP method (Makeneni et al., submitted). Binding free energies were calculated with the MM-GBSA (30, 31) module in AMBERTOOLS12. All the water molecules were removed prior to the MM-GBSA calculation, and desolvation free energies were approximated using the generalized born implicit solvation model (igb = 2) (32).

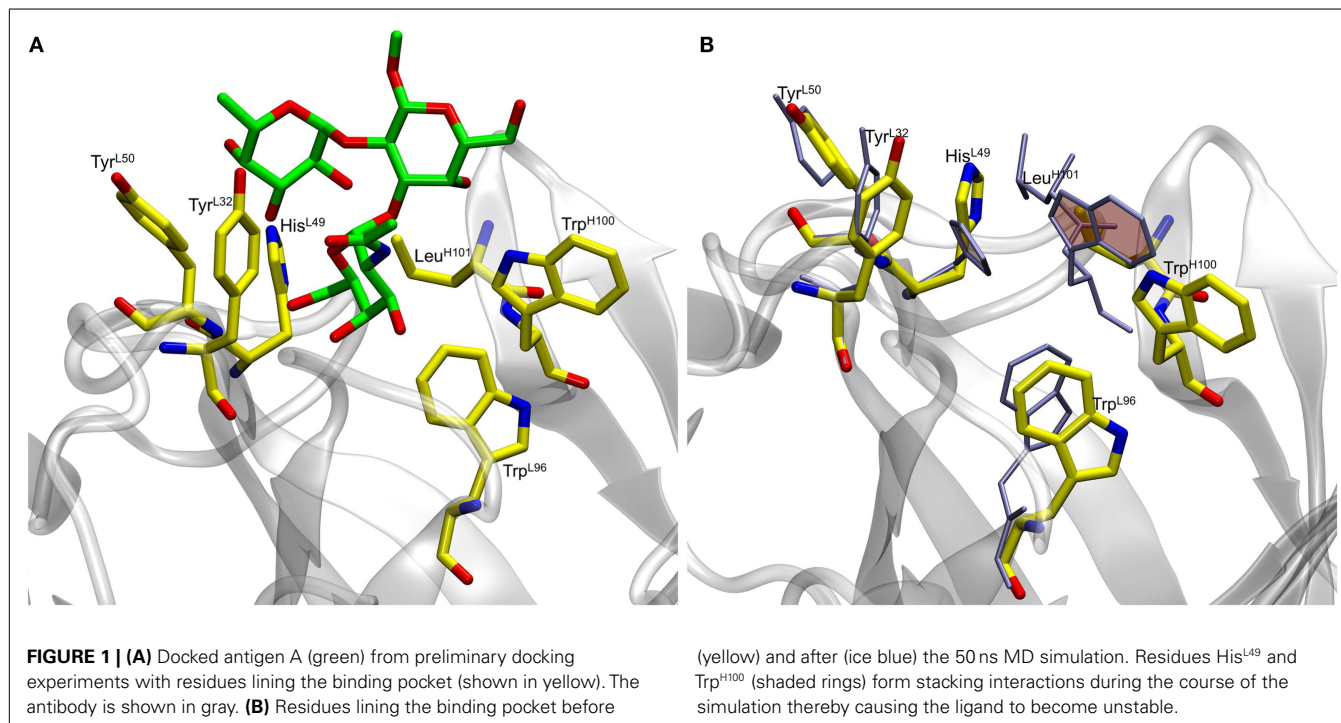
## RESULTS AND DISCUSSION

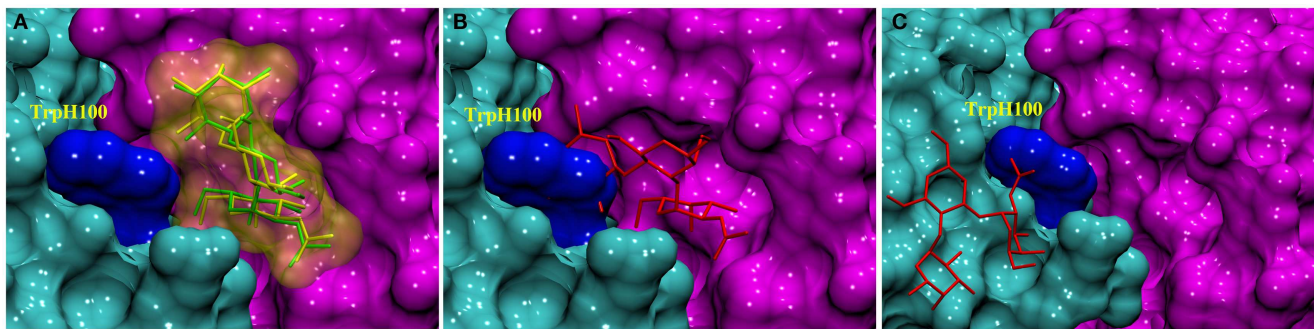
### DOCKING ANALYSIS

In preliminary experiments, docking to the rigid scFv structure yielded complexes that failed to remain stable during subsequent 10 ns MD simulations (Table S1 in Supplementary Material). The spontaneous dissociation of the complex during MD

simulation suggested that the docking had failed to detect the correct, high affinity, pose (33). Upon inspection of the MD data, it was observed that light chain residue His<sup>L49</sup> forms a stacking interaction with heavy chain residue Trp<sup>H100</sup>, which occupies a large volume of the presumed binding site, potentially preventing deeper penetration of the ligand (Figure 1).

As Trp residues can also form stacking interactions with the apolar face of monosaccharides in antibody complexes (34), we hypothesized that the trisaccharide ligand might compete for formation of such an interaction with Trp<sup>H100</sup>. For example, the galactose (Gal) residue in a *Salmonella* trisaccharide antigen stacks against Trp<sup>L93</sup> in the complex with Fab Se155-4 (34). In addition, in the same complex, Trp<sup>H33</sup> stacks against the C-6 position in the 6-deoxy sugar Abequose. The BGA antigen contains GalNAc and a 6-deoxy monosaccharide (fucose, Fuc), thus a revised docking experiment was sought that would permit the formation of such interactions with the aromatic residues in the binding pocket. Thus, two alternative docking experiments were designed: in the first, the side-chain torsion angles of Trp<sup>H100</sup> were allowed to be flexible during docking (termed flexible residue docking); while in the second, Trp<sup>H100</sup> was mutated to Ala prior to docking, and then reverted back to Trp after docking (mutational residue docking). The docked poses were filtered based on three criteria. First, poses in which the GalNAc was not located within the binding pocket were eliminated (Figure 2C). This criterion was adopted based on the results from two array screenings, which indicated that the antibody interacts exclusively with the BGA antigens (Tables S2 and S3 in Supplementary Material) and because the only structural difference between BGA and BGB is the presence of the NAC moiety in the former. Therefore it was hypothesized that the ability of the antibody to discriminate between these two antigens





**FIGURE 2 | Docked complexes of BGA (stick structure) in the scFv binding site (heavy and light chains shown as solvent accessible surfaces in cyan and pink, respectively, the Trp<sup>H100</sup> surface is shown in dark blue). (A) The stick structures in green and yellow represent the best-docked poses from the Trp<sup>H100</sup>-mutagenesis and the flexible**

residue docking approaches, respectively. **(B)** An example of a docked pose (red) that was eliminated on the basis of clashes ensuing from the Ala<sup>H100</sup>Trp mutation. **(C)** An example of a docked pose (red) that was eliminated on the basis of the orientation of the ligand in the binding pocket.

**Table 1 | Comparison of glycosidic torsion angles between experimentally observed values and average values obtained from the MD simulations.**

	$(\phi_1, \varphi_1)^a$		$(\phi_2, \varphi_2)^b$	
	Experimental	Theoretical	Experimental	Theoretical
BGA	$62^\circ < \phi_1 < 82^\circ$	$-68 \pm 14^\circ$	$-77^\circ < \phi_2 < -67^\circ$	$-69 \pm 11^\circ$
trisaccharide	$61^\circ < \phi_1 < 74^\circ$	$51 \pm 25^\circ$	$-109^\circ < \phi_2 < -86^\circ$	$-101 \pm 26^\circ$
BGA–scFv complex	$68^\circ, 77^\circ$	$82 \pm 11^\circ, 68 \pm 7^\circ$	$-68^\circ, -90^\circ$	$-69 \pm 8^\circ, -113 \pm 10^\circ$

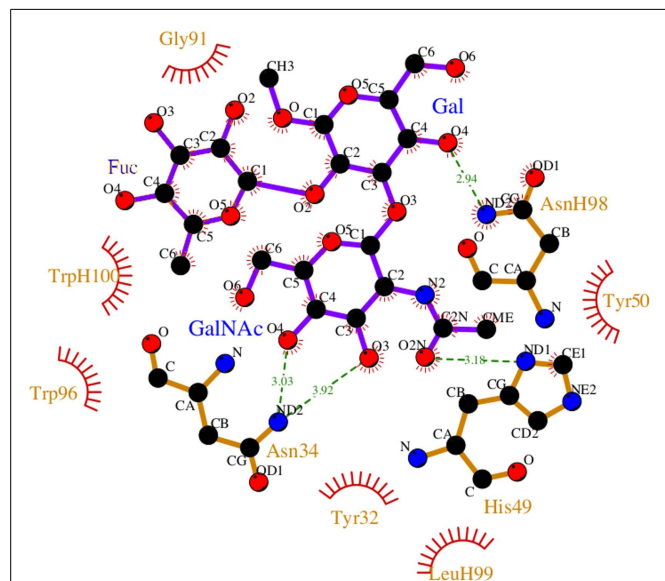
<sup>a</sup>Glycosidic torsion angles for the GalNAc $\alpha$ (1,3)Gal ( $\phi_1, \varphi_1$ ).

<sup>b</sup>Torsion angles for Fuc $\alpha$ (1,2)Gal ( $\phi_2, \varphi_2$ ).

would be dependent on interactions with this residue. Second, in the case of the mutational approach, poses were rejected if the Ala-Trp mutation led to irreconcilable steric clashes with the antigen (**Figure 2B**). All the docked poses obtained from each of these approaches were then scored using a CHI scoring function. After applying these criteria, both docking approaches identified essentially equivalent antigen poses ( $0.48 \text{ \AA}$  RMSD between ligand positions) (**Figure 2A**), in which the C-6 atom of the GalNAc forms a CH/ $\pi$  stacking interaction with the Trp<sup>H96</sup>. This complex was selected for further analysis by MD simulation.

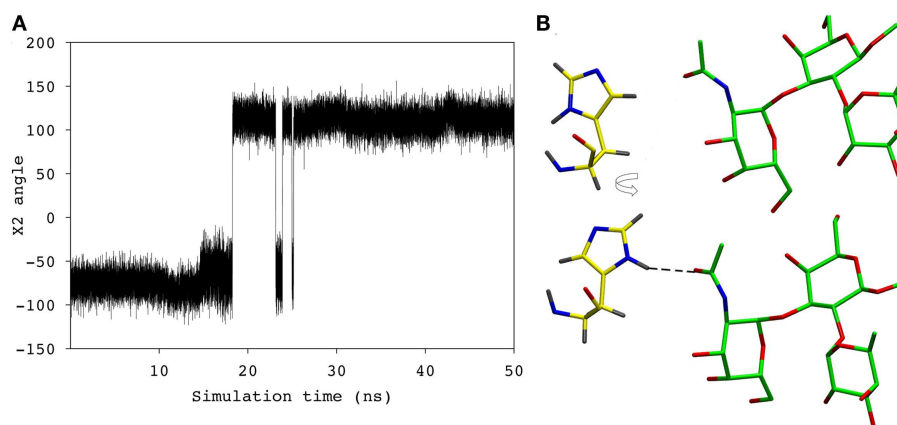
### STRUCTURAL STABILITY OF THE IMMUNE COMPLEXES Blood group A

The final docked model of the blood group antigen A bound to the antibody remained stable during the course of a 50 ns simulation based on the RMSD of the ring atoms of the ligand, which remained between 2 and  $4 \text{ \AA}$  over the course of the simulation (**Figure 5**). An analysis of the ring conformational preferences showed that all three residues in the trisaccharide remained in the  $^4C_1$  chair conformations. The  $\phi$ - and  $\psi$ -glycosidic torsion angles for the GalNAc $\alpha$ (1,3)Gal ( $\phi_1, \varphi_1$ ) and Fuc $\alpha$ (1,2)Gal ( $\phi_2, \varphi_2$ ),



**FIGURE 3 | Non-bonded interactions between the BGA and Fab AC1001 [prepared using LigPlot (38)].** The structure represents a single frame of the MD simulation that is closest to the average RMSD of the structure during the simulation. Unless shown with an H, all residues are from VL.

$\psi_2$ ) linkages were monitored throughout both the simulations (BGA–scFv complex and BGA trisaccharide in solution) and the average values were found to be in agreement with the values observed for the same trisaccharide in the complex with *Dolichos biflorus* lectin as well as the conformations of the trisaccharide in solution (35) (**Table 1**). The stacking interactions between the GalNAc and Trp<sup>H96</sup> interactions were characterized by the angle ( $\theta$ ) between the normals to the ring planes, and the distance ( $R$ ) between their centroids (36). For an ideal stacking conformation,  $\theta$  should be around  $180^\circ$  or  $0^\circ$ , and for CH/ $\pi$ , it should be around  $90^\circ$ . The average  $\theta$  value was close to the latter at  $108^\circ$  (with a standard deviation of  $9^\circ$ ) at a distance of  $6.5 \text{ \AA}$ .



**FIGURE 4 | (A)** The  $\chi_2$  angle of the His<sup>L49</sup> during the course of the simulation. **(B)** His<sup>L49</sup> (shown in yellow) during the first 18 ns of the simulation (top) and the remainder of the simulation (bottom).

**Table 2 | Hydrogen bonds between BGA and the scFv during the MD simulation.**

Donor		Acceptor		MD period: 0–18 ns		MD period: 18–50 ns	
				Distance <sup>a</sup> (Å)	Occupancy (%)	Distance (Å)	Occupancy (%)
GalNAc	O3	Asn <sup>L34</sup>	H81	3.1 (0.18) <sup>c</sup>	67	> 3.5	–
	O4	Asn <sup>L34</sup>	H81	3.1(0.22)	32	3.0 (0.17)	77
	O2N	His <sup>L49</sup>	H8	>3.5	–	2.9 (0.16)	91
Gal	O4	GalNAc	H2N	3.2 (0.17)	65	3.2 (0.17)	31
	O4	Asn <sup>H98</sup>	H81	3.1 (0.18)	45	3.1 (0.17)	41

<sup>a</sup>Standard deviations in parentheses.

During the course of the MD simulation, the side chain of His<sup>L49</sup> was observed to flip from its initial orientation ( $\chi_2 = (-73^\circ)$ ) to ( $115^\circ$ ) in which it could form a hydrogen bond with the N-acetyl group of the GalNAc residue (**Figures 3 and 4; Table 2**). This interaction remained stable for the remainder of the 50 ns simulation. This side-chain flip may represent an example of induced fit during ligand binding, however, at the resolution of the present X-ray data (2.2 Å), it is not possible to reliably discriminate between histidine  $\chi_2$  rotamers (37).

### Blood group B

To probe the specificity of the antibody for antigen B, the scFv was screened experimentally against an array of neoglycoconjugates including ABO and related blood group antigens. The screening confirmed the exclusive specificity of the antibody for BGA-related antigens (Tables S2 and S3 in Supplementary Material). Computational carbohydrate grafting (39) of the relevant glycans from the array onto the bound BGA trisaccharide in the scFv complex confirmed that all of the BGA- and BGB-related glycans could be accommodated in the binding pocket (Table S3 in Supplementary Material). Therefore, the lack of binding of the BGB-glycans does not appear to be due to steric collisions, but rather to the loss of affinity arising from the absence of the NAC group in the BGA congeners. MD simulation of the BGB–scFv

complex was employed to examine the effect of the loss of the NAC moiety on the stability and affinity of the structural difference in the antigens on the stability and affinity of the putative immune complex. Despite the fact that the MD simulations of the two complexes (BGA and BGB) were started with the antigens aligned in identical binding modes, the BGB antigen dissociated from the antibody after a relatively short simulation period of 10 ns. In order to eliminate the possibility that this instability arose due to artifacts from the conversion of the BGA to BGB antigen, two additional simulations were performed with independent initial atomic velocities. In both cases, the ligand appeared to dissociate from the antibody after approximately 10 ns (**Figure 5**). To enable comparison with the BGA complex, only the data from the initial stable 10 ns period of the BGB complex were chosen for analysis.

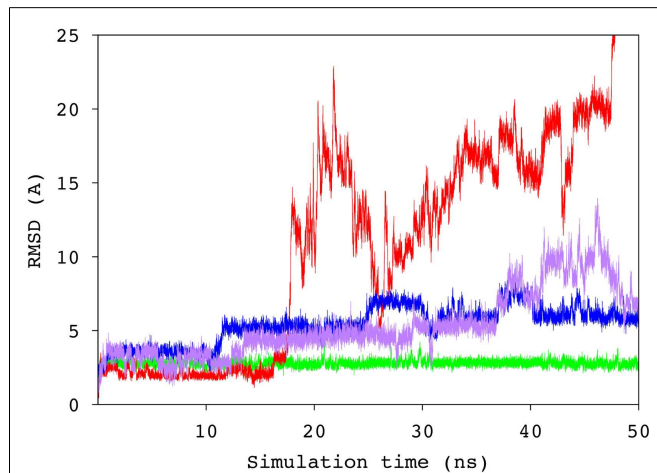
In antigen–scFv complexes, the Gal or GalNAc residues are flanked by residues Tyr<sup>L50</sup>, Asn<sup>L34</sup> and His<sup>L49</sup> on one side of the antigen (Group 1) and residues Trp<sup>H100</sup> and Trp<sup>L96</sup> (Group 2) on the other; the Fuc interacts with Gly<sup>L91</sup> and Asn<sup>L92</sup> (Group 3) (**Figure 6**). In contrast to the case of the BGA antigen, in the BGB–scFv simulation His<sup>L49</sup> does not form a stabilizing interaction with the terminal Gal residue. Additionally, the Gal and Fuc residues display enhanced flexibility owing to the loss of stabilizing interactions with residues from Groups 2 and 3.



### INVOLVEMENT OF HIS<sup>L49</sup> IN BINDING AFFINITY

All histidines in the scFv were protonated by default for modeling with a hydrogen atom at the  $\delta$  nitrogen position. During the MD simulation of the BGA–scFv complex, the  $\chi_2$  angle of His<sup>L49</sup> flips ( $-73^\circ$  to  $115^\circ$ ) enabling a hydrogen bond to form with the carbonyl moiety of the NAc group in the GalNAc residue in BGA, which would be expected to be significant for enhancing the stability of the BGA–scFv complex. In the BGB complex, the same His<sup>L49</sup> forms an interaction with the non-terminal Gal residue. The interaction with His<sup>L49</sup> suggests that there might

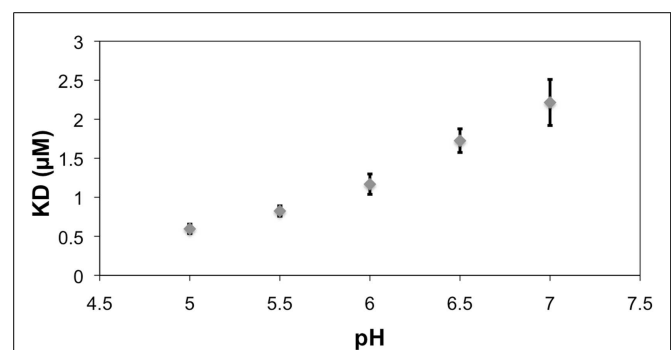
also be a pH dependence on binding; at lower pH all histidines would be positively charged, potentially enhancing the strength of the His<sup>L49</sup>–BGA hydrogen bond, leading to higher binding affinity. This prediction was confirmed by BLI measurements, which showed a marked decrease in the apparent  $K_D$  as the pH dropped below the  $pK_a$  of histidine (Figure 7). It should be noted that this protonation would not be localized to His<sup>L49</sup> nevertheless, no enhanced non-specific binding was observed at low pH for either BSA or BGA–Le<sup>x</sup> (Figures S1–S3 in Supplementary Material), supporting a role for a direct interaction between His<sup>L49</sup> and the BGA antigen.



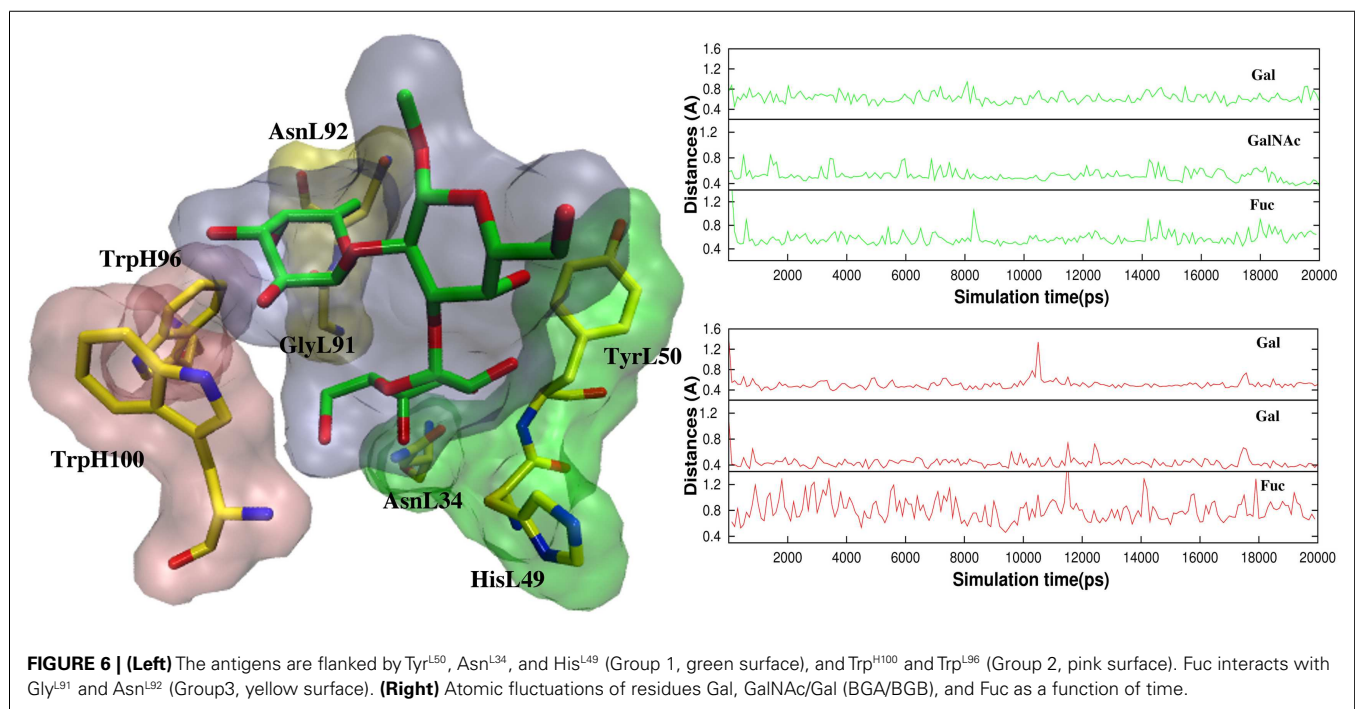
**FIGURE 5 |** Time series of the RMSD values for the ring atoms of the BGA (green) and BGB (from three independent simulations, blue, purple, and red) antigens, relative to the starting conformation of the complex.

### BINDING ENERGY ANALYSIS

A per-residue decomposition of the interaction energies in the immune complexes indicated that, in the case of BGA, the



**FIGURE 7 |** The reference (BSA)-subtracted pH dependence of the apparent  $K_D$  for the interaction between scFv AC1001 and the BSA–BGA conjugate. Error bars are derived from replicates of five measurements. Note, the  $pK_a$  of histidine is 6.04 (40).



**FIGURE 6 |** (Left) The antigens are flanked by Tyr<sup>L50</sup>, Asn<sup>L34</sup>, and His<sup>L49</sup> (Group 1, green surface), and Trp<sup>H100</sup> and Trp<sup>L96</sup> (Group 2, pink surface). Fuc interacts with Gly<sup>L91</sup> and Asn<sup>L92</sup> (Group 3, yellow surface). (Right) Atomic fluctuations of residues Gal, GalNAc/Gal (BGA/BGB), and Fuc as a function of time.

GalNAc residue contributed 25% (-8.1 kcal/mol) toward the binding energy, compared to a reduced (-4.3 kcal/mol) contribution from the corresponding Gal residue in BGB (Table 3). This loss

of 4 kcal/mol of interaction energy is the predominant difference between the two antigens, and would be enough to reduce the affinity by nearly 800-fold, consistent with the lack of apparent binding of the BGB analogs in the glycan array screening. In addition, this analysis identified the residues that contributed significantly toward antigen binding.

In the BGA–scFv complex, residues from CDR L3 make the maximum contributions to binding (Gly<sup>L91</sup> + Trp<sup>L96</sup> + Asn<sup>L92</sup> + Thr<sup>L93</sup> = -7.2 kcal/mol) followed by H3 (Asn<sup>H98</sup> + Trp<sup>H100</sup> + Leu<sup>H99</sup> = -5.5 kcal/mol), L1 (Tyr<sup>L32</sup> + Asn<sup>L34</sup> = -4.5 kcal/mol), and L2 (Tyr<sup>L50</sup> = -1.02 kcal/mol). In contrast, in the case of BGB, the same residues from L3 contribute less than a total of 1 kcal/mol to the interaction energies. The most significant single residues are Tyr<sup>L32</sup>, Gly<sup>L91</sup>, Trp<sup>H100</sup>, and Trp<sup>L96</sup>, which each contributes more than 2 kcal/mol and together account for 50% of the total affinity. Residues Gly<sup>L91</sup> and Asn<sup>L92</sup> that form hydrogen bonds with the Fuc residue together contribute -4.0 kcal/mol to the binding of BGA, but fail to make any stable interactions in the BGB simulation and therefore contribute negligibly to the affinity. It is these interactions that provide the predominant contributions to the preferential binding of the BGA antigen. While in the BGB complex, His<sup>L49</sup> does not form any stable hydrogen bonds with the terminal Gal, it is able to form new, albeit transient, interactions with the non-terminal Gal for 30% of the stable simulation period. Therefore, while the per-residue contribution values indicate that His<sup>L49</sup> makes a contribution greater than -1.5 kcal/mol in both cases, the interactions it forms in BGA are more stable when compared to the interactions in BGB.

## CONCLUSION

In this study, 3D models of the BGA and BGB trisaccharides in complex with scFv AC1001 were generated that provided a detailed atomic level rationalization of the interactions and dynamics responsible for antigen specificity. Quantification of the binding affinities identified key residues in the binding site that are predicted to contribute to specific and non-specific interactions with each antigen and led to the confirmed prediction of enhanced binding at lower pH. The spontaneous dissociation of antigen B from the scFv–BGB complexes (in three different simulations) indicated that MD simulations confirm the known preference of this antibody for the A antigen, and support a role for MD simulations in overcoming limitations associated with ligand docking. The present study illustrates that integration of multiple experimental (affinity measurements, glycan array screening, and crystallography) and theoretical (ligand docking, MD simulation, and energy decomposition) methods provides a powerful platform for predicting the origin of antibody–carbohydrate specificity.

## AUTHOR CONTRIBUTIONS

Spandana Makeneni, N. Martin Young, and Robert J. Woods conceived and designed the experiments. Spandana Makeneni, Ye Ji, and David C. Watson performed the experiments. Spandana Makeneni, Ye Ji, and Robert J. Woods analyzed the data. Spandana Makeneni, Ye Ji, N. Martin Young, David C. Watson, and Robert J. Woods contributed reagents/materials/analysis tools. Spandana Makeneni, N. Martin Young, and Robert J. Woods wrote the paper.

**Table 3 | Key<sup>a</sup> per-residue contributions toward the energy for the BGA and BGB–scFv complexes.**

Residue	vdW	Electrostatic	Polar desolvation	Non-polar desolvation	Total
<b>Antibody</b>					
Tyr <sup>L32</sup>	-4.2 <sup>b</sup>	-1.4	2.1	-0.5	-4.0
	-3.0 <sup>b</sup>	-1.7	2.0	-0.5	-3.2
Gly <sup>L91</sup>	-0.8	-4.8	3.2	-0.2	-2.5
	–	–	–	–	–
Trp <sup>H100</sup>	-2.5	-1.2	1.7	-0.3	-2.2
	-2.8	-1.0	1.2	-0.4	-2.9
Trp <sup>L96</sup>	-1.4	-1.1	0.7	-0.2	-2.0
	-0.7	-0.3	0.2	-0.1	-0.9
Asn <sup>L34</sup>	-0.8	-1.6	0.9	-0.1	-1.6
	-0.6	-3.7	1.5	-0.1	-2.9
Asn <sup>H98</sup>	-2.2	-2.9	4.0	-0.5	-1.5
	-2.8	-1.6	3.8	-0.5	-1.1
Asn <sup>L92</sup>	-1	-2.9	2.6	-0.2	-1.5
	–	–	–	–	–
Tyr <sup>L50</sup>	-1.6	-0.9	1.1	-0.1	-1.5
	-1.7	-1.0	1.4	-0.1	-1.5
Leu <sup>H99</sup>	-1.4	-1	1	-0.1	-1.4
	-1.4	-0.3	0.8	-0.1	-1
His <sup>L49</sup>	-0.6	-3.3	2.5	-0.1	-1.4
	-1.3	-2.8	2.5	-0.2	-1.8
Thr <sup>L93</sup>	-0.5	-0.6	0.6	0.0	-0.5
	–	–	–	–	–
<b>Subtotal</b>	-17.1	-20.5	20.5	-2.2	-19.3
	-11.3	-12.4	13.4	-2	-15.3
<b>Antigen</b>					
Gal	-3.1	-0.4	1.2	-0.4	-2.7
	-6.2	-3.7	5.2	-0.9	-5.6
GalNAc/Gal	-13.1	-12.5	19.6	-2.1	-8.1
	-10.1	-8.2	15.8	-1.7	-4.3
Fuc	-4.0	-9.7	12.2	-0.8	-2.3
	-2.5	-1.2	4.5	-0.4	0.4
<b>Ligand total</b>	-20.2	-22.6	32.9	-3.2	-13.2
	-18.9	-13.2	25.4	-1.4	-9.6

<sup>a</sup> Key residues defined as those that contribute >0.5 kcal/mol to the total interaction energy for either the BGA or BGB in the complexes. Only the initial stable 10 ns period of the BGB simulation was employed, whereas the entire 50 ns trajectory for BGA was analyzed.

<sup>b</sup> Upper row, values for BGA; lower, BGB.



## ACKNOWLEDGMENTS

We would like to thank Dr. Jeffery C. Gildersleeve and W. Shea Wright for screening the scFv against the carbohydrate array. We wish to acknowledge the Consortium for Functional Glycomics grant number GM62116 for screening the antibody against their array. We thank the National Institutes for Health [R01 GM094919 (EUREKA) and P41 GM103390], as well as the Science Foundation of Ireland (08/IN.1/B2070) and the European Research Development Fund for support.

## SUPPLEMENTARY MATERIAL

Included in the supplemental information are details and raw data from the BLI experiments. Results from screening of the scFv against both CFG and carbohydrate arrays (Dr. Jeffery Gildersleeve's group) are also presented in the supplemental material. The Supplementary Material for this article can be found online at <http://www.frontiersin.org/Journal/10.3389/fimmu.2014.00397>

## REFERENCES

- Landsteiner K. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralbl Bakteriol* (1900) 27(10):357–62.
- Yamamoto F. Review: ABO blood group system – ABH oligosaccharide antigens, anti-A and anti-B, A and B glycosyltransferases, and ABO genes. *Immunohematology* (2004) 20(1):3–22.
- Chester MA, Olsson ML. The ABO blood group gene: a locus of considerable genetic diversity. *Transfus Med Rev* (2001) 15(3):177–200. doi:10.1053/tmrv.2001.24591
- Marionneau S, Cailleau-Thomas A, Rocher J, Le Moullac-Vaidye B, Ruvoën N, Clément M, et al. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie* (2001) 83(7):565–73. doi:10.1016/S0300-9084(01)01321-9
- Williamson LM, Lowe S, Love EM, Cohen H, Soldan K, McClelland DB, et al. Serious hazards of transfusion (SHOT) initiative: analysis of the first two annual reports. *BMJ* (1999) 319(7201):16–9. doi:10.1136/bmj.319.7201.16
- Sazama K. Reports of 355 transfusion-associated deaths: 1976 through 1985. *Transfusion* (1990) 30(7):583–90. doi:10.1046/j.1537-2995.1990.30790385515.x
- Ozolek JA, Watchko JF, Mimouni F. Prevalence and lack of clinical significance of blood group incompatibility in mothers with blood type A or B. *J Pediatr* (1994) 125(1):87–91. doi:10.1016/S0022-3476(94)70131-8
- Le Pendu J, Marionneau S, Cailleau-Thomas A, Rocher J, Le Moullac-Vaidye B, Clément M. ABH and Lewis histo-blood group antigens in cancer. *APMIS* (2001) 109(1):9–31. doi:10.1111/j.1600-0463.2001.tb00011.x
- Dabelsteen E, Gao S. ABO blood-group antigens in oral cancer. *J Dent Res* (2005) 84(1):21–8. doi:10.1177/154405910508400103
- Anstee DJ. The relationship between blood groups and disease. *Blood* (2010) 115(23):4635–43. doi:10.1182/blood-2010-01-261859
- Ceravolo IP, Sanchez BAM, Sousa TN, Guerra BM, Soares IS, Braga EM, et al. Naturally acquired inhibitory antibodies to *Plasmodium vivax* Duffy binding protein are short-lived and allele-specific following a single malaria infection. *Clin Exp Immunol* (2009) 156(3):502–10. doi:10.1111/j.1365-2249.2009.03931.x
- DeMarco ML, Woods RJ. Structural glycobiology: a game of snakes and ladders. *Glycobiology* (2008) 18(6):426–40. doi:10.1093/glycob/cwn026
- Woods R, Yongye A. Computational techniques applied to defining carbohydrate antigenicity. In: Kosma P, Müller-Loennies S, editors. *Anticarbhydrate Antibodies*. Vienna: Springer (2012). p. 361–83.
- Paula S, Monson N, Ball WJ Jr. Molecular modeling of cardiac glycoside binding by the human sequence monoclonal antibody 1B3. *Proteins* (2005) 60(3):382–91. doi:10.1002/prot.20484
- Woods RJ, Tessier MB. Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan–protein complexes. *Curr Opin Struct Biol* (2010) 20(5):575–83. doi:10.1016/j.sbi.2010.07.005
- Nivedha AK, Makeneni S, Foley BL, Tessier MB, Woods RJ. Importance of ligand conformational energies in carbohydrate docking: sorting the wheat from the chaff. *J Comput Chem* (2014) 35(7):526–39. doi:10.1002/jcc.23517
- Thomas R, Patenaude SI, MacKenzie CR, To R, Hiram T, Young NM, et al. Structure of an anti-blood group A Fv and improvement of its binding affinity without loss of specificity. *J Biol Chem* (2002) 277(3):2059–64. doi:10.1074/jbc.M104364200
- Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* (2010) 31(2):455–61. doi:10.1002/jcc.21334
- Hadden JA, Tessier MB, Fadda E, Woods RJ. Calculating binding free energies for protein-carbohydrate complexes. *Methods in Molecular Biology: Glycoinformatics* (in press).
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* (2009) 30(16):2785–91. doi:10.1002/jcc.21256
- Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. *Tetrahedron* (1980) 36(22):3219–28. doi:10.1016/0040-4020(80)80168-2
- Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, et al. *AMBER 12*. San Francisco, CA: University of California (2012).
- Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J Chem Theory Comput* (2012) 8(5):1542–55. doi:10.1021/ct200909j
- Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* (2006) 65(3):712–25. doi:10.1002/prot.21123
- Kirschner KN, Yongye AB, Tschampel SM, González-Outeirino J, Daniels CR, Foley BL, et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* (2008) 29(4):622–55. doi:10.1002/jcc.20820
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* (1983) 79(2):926–35. doi:10.1063/1.445869
- Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* (1977) 23(3):327–41. doi:10.1016/0021-9991(77)90098-5
- Darden T, York D, Pedersen L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys* (1993) 98(12):10089–92. doi:10.1063/1.464397
- Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* (2013) 9(7):3084–95. doi:10.1021/ct400341p
- Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* (2000) 33(12):889–97. doi:10.1021/ar000033j
- Srinivasan J, Miller J, Kollman PA, Case DA. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J Biomol Struct Dyn* (1998) 16(3):671–82. doi:10.1080/07391102.1998.10508279
- Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* (2004) 55(2):383–94. doi:10.1002/prot.20033
- Proctor Elizabeth A, Yin S, Tropsha A, Dokholyan Nikolay V. Discrete molecular dynamics distinguishes native-like binding poses from decoys in difficult targets. *Biophys J* (2012) 102(1):144–51. doi:10.1016/j.bpj.2011.11.4008
- Cyglar M, Rose DR, Bundle DR. Recognition of a cell-surface oligosaccharide of pathogenic *Salmonella* by an antibody Fab fragment. *Science* (1991) 253(5018):442–5. doi:10.2307/2878892
- Casset F, Peters T, Etlzer M, Korchagina E, Nifant'ev N, Pérez S, et al. Conformational analysis of blood group A trisaccharide in solution and in the binding site of *Dolichos biflorus* lectin using transient and transferred nuclear Overhauser enhancement (NOE) and rotating-frame NOE experiments. *Eur J Biochem* (1996) 239(3):710–9. doi:10.1111/j.1432-1033.1996.0710u.x
- Ford MG, Weimar T, Kohli T, Woods RJ. Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity. *Proteins* (2003) 53(2):229–40. doi:10.1002/prot.10428
- Glusker JPL, Mitchell L, Rossi M. Crystal structure analysis for chemists and biologists. *J Chem Educ* (1995) 72(3):A73. doi:10.1021/ed072pA73.9
- Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* (1995) 8(2):127–34. doi:10.1093/protein/8.2.127

39. Tessier MB, Grant OC, Heimburg-Molinaro J, Smith D, Jadey S, Gulick AM, et al. Computational screening of the human TF-glycome provides a structural definition for the specificity of anti-tumor antibody JAA-F11. *PLoS One* (2013) **8**(1):e54874. doi:10.1371/journal.pone.0054874
40. Wood EJ. Data for biochemical research (third edition) by R M C Dawson, D C Elliott, W H Elliott and K M Jones. *Biochem Educ* (1987) **15**(2):97. doi:10.1016/0307-4412(87)90110-5

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 July 2014; accepted: 05 August 2014; published online: 22 August 2014.

Citation: Makeneni S, Ji Y, Watson DC, Young NM and Woods RJ (2014) Predicting the origins of anti-blood group antibody specificity: a case study of the ABO A- and B-antigens. *Front. Immunol.* **5**:397. doi: 10.3389/fimmu.2014.00397

This article was submitted to *Immunotherapies and Vaccines*, a section of the journal *Frontiers in Immunology*.

Copyright © 2014 Makeneni, Ji, Watson, Young and Woods. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Differential site accessibility mechanistically explains subcellular-specific *N*-glycosylation determinants

Ling Yen Lee, Chi-Hung Lin, Susan Fanayan, Nicolle H. Packer and Morten Thaysen-Andersen\*

Department of Chemistry and Biomolecular Sciences, Biomolecular Frontiers Research Centre, Macquarie University, Sydney, NSW, Australia

## Edited by:

Elizabeth Yuriev, Monash University, Australia

## Reviewed by:

Jamie Heimbarg-Molinaro, Emory University, USA

Tony Velkov, Monash University, Australia

## \*Correspondence:

Morten Thaysen-Andersen,  
Department of Chemistry and  
Biomolecular Sciences, Biomolecular  
Frontiers Research Centre, Macquarie  
University, Sydney, NSW 2109,  
Australia  
e-mail: morten.andersen@mq.edu.au

Glycoproteins perform extra- and intracellular functions in innate and adaptive immunity by lectin-based interactions to exposed glyco-determinants. Herein, we document and mechanistically explain the formation of subcellular-specific *N*-glycosylation determinants on glycoproteins trafficking through the shared biosynthetic machinery of human cells. LC-MS/MS-based quantitative glycomics showed that the secreted glycoproteins of eight human breast epithelial cells displaying diverse geno- and phenotypes consistently displayed more processed, primarily complex type, *N*-glycans than the high-mannose-rich microsomal glycoproteins. Detailed subcellular glycome profiling of proteins derived from three breast cell lines (MCF7/MDA468/MCF10A) demonstrated that secreted glycoproteins displayed significantly more  $\alpha$ -sialylation and  $\alpha$ 1,6-fucosylation, but less  $\alpha$ -mannosylation, than both the intermediately glycan-processed cell-surface glycoproteomes and the under-processed microsomal glycoproteomes. Subcellular proteomics and gene ontology revealed substantial presence of endoplasmic reticulum resident glycoproteins in the microsomes and confirmed significant enrichment of secreted and cell-surface glycoproteins in the respective subcellular fractions. The solvent accessibility of the glycosylation sites on maturely folded proteins of the 100 most abundant putative *N*-glycoproteins observed uniquely in the three subcellular glycoproteomes correlated with the glycan type processing thereby mechanistically explaining the formation of subcellular-specific *N*-glycosylation. In conclusion, human cells have developed mechanisms to simultaneously and reproducibly generate subcellular-specific *N*-glycosylation using a shared biosynthetic machinery. This aspect of protein-specific glycosylation is important for structural and functional glycobiology and discussed here in the context of the spatio-temporal interaction of glyco-determinants with lectins central to infection and immunity.

**Keywords:** *N*-glycosylation, solvent accessibility, *N*-glycome, subcellular location, glycoproteome, glycosylation site, *N*-glycan, glycoprotein

## INTRODUCTION

Significant parts of the human genome and cellular energy are dedicated to produce and regulate protein glycosylation (1). Hence, it is no surprise that this abundant post-translational modification is important in a wide spectrum of biological processes to maintain cellular homeostasis (2). Dysregulation of protein glycosylation is a cause and/or effect of numerous pathological conditions including, but not limited to, congenital disorder of glycosylation (3), cystic fibrosis (4), inflammation (5), auto-immunity (6), and cancer (7). The extracellular location of secreted and cell-surface-tethered proteins carrying *N*-linked glycosylation is ideal for facilitating molecular interactions with the surrounding environment (8). Intracellular functions of *N*-glycoproteins are also known (9, 10). The terminal determinants of host *N*-glycans (so-called “self” and “altered self” in disease) are recognized by endogenous and exogenous glycan-binding proteins commonly called lectins. Interactions between lectins and *N*-glycans are central in innate and adaptive immunity (11). Important examples include the C-type lectins, which may be crudely divided into lectins having affinity for  $\alpha$ -mannose/ $\alpha$ -fucose-terminated *N*-glycans including dendritic cell-specific intercellular adhesion molecule-3-grabbing

non-integrin (DC-SIGN), macrophage mannose receptors and Langerin (12), and lectins having affinity for galactose/GalNAc terminating glycans such as macrophage galactose lectin and DC-asialoglycoprotein receptor (13, 14). In addition, siglecs (I-type lectins) and galectins (S-type lectins) are important for facilitating a functional immune response (15).

The human *N*-glycosylation biosynthetic machinery is relatively well understood (16, 17). In brief, the synthesis is initiated by the transfer of common immature glycan precursors i.e., Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub> to conserved sequons (N<sub>x</sub>T/S, x ≠ P) on translocating polypeptide chains. The glycan precursor is then remodeled through sequential trimming and elongation by specific glycosidases and glycosyltransferases located in the endoplasmic reticulum (ER) and the *cis*-, medial, and *trans*-Golgi, respectively. This series of enzymatic processes first results in the trafficking *N*-glycoproteins being comprised of attached high-mannose-type *N*-glycans, which progresses to the hybrid- and complex-type stage if sufficient interactions with the processing enzymes occur (17). The Golgi-based *N*-glycan processing, including the formation of glycan types and the addition of terminal determinants such as  $\alpha$ -fucosylation and  $\alpha$ -sialylation, occurs on maturely folded

glycoproteins (18, 19). An extensive and reproducible repertoire of *N*-glycans is usually present on a given glycosylation site (20). This *N*-glycan microheterogeneity on proteins results from incomplete processing by the multiple competing enzymatic reactions that can be influenced by cellular factors including the availability of nucleotide sugars, glycosylation enzyme activity, and glycoprotein trafficking time through the biosynthetic machinery. Such cellular factors contribute to cell- and tissue-specific *N*-glycosylation (21). Importantly, the structures of the individual glycoproteins trafficking through the glycosylation machinery dramatically influence the degree of *N*-glycan processing creating protein- and site-specific *N*-glycosylation (22). By thorough literature-based curation of published site-specific glycoproteomics data of mammalian *N*-glycoproteins, we recently confirmed that several structural features including glycan type formation,  $\alpha$ 1,6-(core) fucosylation, and  $\beta$ 1,4/6-GlcNAc branching of *N*-glycans are strongly correlated with the solvent accessibility of the glycosylation sites of maturely folded glycoproteins (19). As such, extensive *N*-glycan processing was observed for proteins displaying solvent accessible glycosylation sites relative to spatially hidden sites. Thus, differential site accessibility can explain how glycoproteins produced simultaneously in the same cell, and even sequons on the same glycoproteins, can present widely different *N*-glycan structural repertoires.

Considering the importance of protein- and site-specific *N*-glycosylation in many aspects of glycobiology including glycoimmunology, we here seek to further explore this feature in the context of the multiple subcellular glycoproteomes that traffic through the shared glycosylation machinery in the secretory pathway of human cells, yet end up at different cellular locations. Due to the functional implications of both intra- and extracellular *N*-glycoproteins, we focus on the secreted, cell-surface, and intracellular glycoproteomes, the latter fraction largely represented by microsomal proteins (23). Understanding, how the subcellular glycoproteomes are generated and regulated under normal and altered physiological conditions of the cell is valuable to the understanding of their involvement in immune biology. Recent analytical developments in glycomics (24–27) and glycoproteomics (28–31) have, together with more conventional proteomics, enabled sensitive, and detailed system-wide investigations of the regulation of protein *N*-glycosylation in immunity (32).

Using LC-MS/MS-based glycomics and proteomics on multiple subcellular fractions from a panel of human cell lines displaying diverse cellular characteristics, we here document that human cells have developed a general mechanism to reproducibly generate vastly different *N*-glycan determinants on their differently located subcellular glycoproteomes that trafficked simultaneously through a shared biosynthetic machinery. We provide evidence that the subcellular-specific protein *N*-glycosylation arises from differential solvent accessibilities of the glycosylation sites of maturely folded glycoproteins that localize to different subcellular compartments following the glycan processing. This aspect of protein-specific glycosylation is discussed here in the context of immunity and infection due to the crucial role of endogenous and exogenous lectins recognizing exposed self, and altered self, glyco-determinants to facilitate the functional immune response.

## MATERIALS AND METHODS

### CELLULAR ORIGIN, CULTURE CONDITIONS, AND DOUBLING TIME

Multiple human cells showing diverse geno- and phenotypical characteristics were used to demonstrate the general nature of the cellular mechanisms observed in this study. Human mammary epithelial cells (HMEC) were purchased (product # CC-2551, Lonza). Human breast epithelial cell lines MCF10A, MCF7, SKBR3, MDA-MB-157 (MDA157), MDA-MB-231 (MDA231), and HS578T as well as a human colon cancer epithelial cell line SW480 were obtained from American Type Culture Collection (Manassas, VA, USA). HMEC was grown in HuMEC Ready Media (Invitrogen). MCF10A was cultured in DMEM/F12 with the addition of 5% horse serum (Invitrogen), 20 ng/mL epidermal growth factor (EGF) (Invitrogen), 0.5  $\mu$ g/mL hydrocortisone (Sigma), 100 ng/mL cholera toxin (Sigma), and 8  $\mu$ g/mL insulin (Invitrogen). Other cell lines were grown in RPMI (Sigma) supplemented in 5% fetal bovine serum (FBS) (Invitrogen), 10 mM glutamine (Invitrogen), and 10  $\mu$ g/mL insulin. Cells were maintained at 37°C in 5% CO<sub>2</sub> for all experiments. The breast cell lines were grown in triplicates to ~80% confluence and washed at least four times with ice-cold phosphate buffered saline (PBS) to remove traces of FBS and incubated in serum-free media at 37°C in 5% CO<sub>2</sub> for 48 h prior to subcellular fractionation.

To measure the cellular doubling times of the breast cell lines, cells were seeded at  $1.3 \times 10^4$  cells/mL/well in six-well plates and incubated overnight at 37°C in 5% CO<sub>2</sub>. Cells were counted every 24 h over a four-day period using a cell counter (Bio-Rad). The doubling time for each cell line was determined from their exponential growth phase. For overview of the investigated cells and associated data, see Table S1 in Supplementary Material.

### COLLECTION AND PREPARATION OF SUBCELLULAR GLYCOPROTEOMES FROM BREAST CELL LINES

The *secreted* subcellular glycoproteomes were collected by sampling 30 mL of serum-free culture media followed by centrifugation at  $2,000 \times g$  to pellet any floating cells. The supernatants were concentrated and buffer exchanged into PBS (1 $\times$ ) using 10,000 MWCO Amicon Ultra membranes (Millipore). Proteins were then precipitated with nine volumes of acetone overnight at –20°C. The pellets were stored at –80°C until further analysis.

The *cell-surface* subcellular glycoproteomes were isolated from MCF7, MDA468, and MCF10A breast epithelial cell lines using a commercial biotinylation kit (product # 89881, Pierce) to specifically biotinylate the cell-surface glycoproteins. The protocol supplied by the manufacturer was followed. Briefly, monolayers of cultured cells grown in 75 cm<sup>2</sup> culture flasks were washed three-times with PBS (1 $\times$ ) before incubation in EZ-Link sulfo-NHS-SS-biotin in ice-cold PBS (1 $\times$ ) for 30 min at 4°C on a rocking platform. The labeling reactions were terminated and the biotinylated cells were washed and collected by scraping in Tris-buffered saline (TBS) (1 $\times$ ), followed by centrifugation at  $500 \times g$  for 3 min. The supernatants were discarded and the cell pellets were disrupted in manufacturer-provided lysis buffer by ultra-sonication using five 1-s bursts with a Sonifier 450 (Branson Sonifier, Wilmington, NC, USA). The cell lysates were centrifuged at  $10,000 \times g$  for 2 min at 4°C. Solubilized biotinylated cell-surface proteins in the clarified supernatants were isolated using NeutrAvidin Agarose.

Cell-surface-bound proteins were eluted using 50 mM DTT and precipitated with acetone overnight at  $-20^{\circ}\text{C}$ . The pellets were stored at  $-80^{\circ}\text{C}$  until analysis.

The *microsome* (total membrane) subcellular glycoproteomes were obtained by first removing serum-free media, thoroughly washing cells with PBS (1 $\times$ ), and harvesting cells in 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA containing a protease inhibitor cocktail (Roche Diagnostics). The cells were ultrasonicated on ice for three rounds of 10-s bursts using a Sonifier 450 and centrifuged at  $2,000 \times g$  for 20 min at  $4^{\circ}\text{C}$  to remove intact cells and nuclei. The supernatants were ultra-centrifuged at  $120,000 \times g$  for 80 min after which the supernatants were discarded. The microsomal membrane pellets were washed twice with ice-cold 0.1 M sodium carbonate and resuspended in 25 mM Tris-HCl pH 7.4, 150 mM NaCl, and 1% (v/v) Triton X-114. Samples were phase partitioned by incubation at  $37^{\circ}\text{C}$  for 20 min, followed by  $1,000 \times g$  centrifugation for 10 min. The upper aqueous layer was carefully removed and nine volumes of ice-cold acetone were added to the lower detergent phase and incubated overnight at  $-20^{\circ}\text{C}$  to precipitate the proteins. The pellets were stored at  $-80^{\circ}\text{C}$  until further analysis.

The protein concentrations of the subcellular fractions were measured using Bradford reagents (Sigma). Equal protein amounts were precipitated in the three subcellular fractions and the resulting pellets were solubilized in 8 M urea for spotting on PVDF membranes for *N*-glycome profiling or in NuPAGE LDS sample buffer for gel electrophoresis prior to proteome profiling.

#### SUBCELLULAR FRACTIONATION OF HUMAN COLON CANCER CELL LINES

SW480 cells ( $5 \times 10^7$ ) were washed twice with homogenization buffer (20 mM HEPES, pH 7.5, and 0.25 M sucrose). Cell pellets were resuspended to a final volume of 2 mL in homogenization buffer and lysed using an Ultra-Turrax disperser (Ika). After a low speed centrifugation at  $1,000 \times g$  for 10 min, the supernatant was collected as the post-nuclear fraction (PNF). The PNF was subjected to ultracentrifugation at 30,000 rpm for 1 h in a SW41Ti rotor (Beckman Coulter) to pellet the microsome. ER and Golgi-enriched membranes were prepared as described (33). Briefly, 1 mL of PNF (usually 2.5–3 mg protein) was adjusted to 1.4 M sucrose by adding 2 mL of 2 M sucrose. A discontinuous sucrose gradient was made by sequentially loading 1.5 mL of 1.6 M sucrose, 3 mL PNF in 1.4 M sucrose, 3 mL of 1.2 M sucrose, and 3 mL of 0.8 M sucrose. All sucrose solutions contained 20 mM HEPES pH 7.5. Ultracentrifugation was conducted at 28,500 rpm for 2 h in a SW41Ti rotor. Enriched-Golgi membranes were harvested at the 0.8 M/1.2 M interface. Enriched ER membranes were harvested from the 1.4 M layer. The collected ER and Golgi membranes were diluted by homogenization buffer to reduce concentration of sucrose and subsequently pelleted by ultracentrifugation at 30,000 rpm for 1 h in a SW41Ti rotor. Pelleted ER- and Golgi-enriched membranes were resuspended in 8 M urea and protein concentrations were determined by BCA assays (Pierce).

#### RELEASE AND PREPARATION OF *N*-GLYCANS FROM SUBCELLULAR GLYCOPROTEOMES

*N*-glycans were released from  $\sim 20 \mu\text{g}$  secreted proteins,  $50 \mu\text{g}$  cell-surface proteins, and  $50 \mu\text{g}$  microsome membrane proteins

as previously described (27). Briefly, protein mixtures were immobilized on methanol-activated PVDF membranes (Millipore) and allowed to dry overnight. Membrane-bound proteins were incubated with 2.5 U PNGase F (*Flavobacterium meningospeticum*, Roche) for 16 h at  $37^{\circ}\text{C}$  to ensure complete release of *N*-glycans. Released *N*-glycans were incubated with 100 mM ammonium acetate (pH 5) for 1 h at RT and subsequently dried by vacuum centrifugation. Reduction of *N*-glycans was performed with  $20 \mu\text{L}$  1 M sodium borohydride (Sigma) in 50 mM potassium hydroxide (Sigma) for 3 h at  $50^{\circ}\text{C}$ . Reactions were quenched with  $2 \mu\text{L}$  glacial acetic acid. Dual desalting was performed in micro-SPE formats using strong cation exchange/ $\text{C}_{18}$  and carbon columns (27). Desalted *N*-glycans were eluted from the carbon columns with  $20 \mu\text{L}$  40% acetonitrile (ACN) containing 0.1% (v/v) trifluoroacetic acid and dried by vacuum centrifugation (34). Samples were stored at  $-80^{\circ}\text{C}$  if not analyzed immediately.

#### DIGESTION AND PREPARATION OF PEPTIDE MIXTURES FROM SUBCELLULAR GLYCOPROTEOMES

The subcellular glycoproteomes of the breast cells ( $\sim 50 \mu\text{g}$  protein/fraction) i.e., secreted, cell surface, and microsomes and of colon cells ( $\sim 10 \mu\text{g}$  protein/fraction) i.e., microsome and ER- and Golgi-enriched membrane fractions were reduced and alkylated and subsequently in-gel (breast cells) or in-solution (colon cells) digested. Prior to in-gel digestion, samples were loaded in  $10 \mu\text{L}$  NuPAGE LDS buffer and separated on 4–12% Bis-Tris PAGE gels (Invitrogen). Electrophoresis was performed at 200 V for 50 min. After separation of proteins, gels were fixed in 40% (v/v) ethanol and 10% (v/v) acetic acid for at least 2 h, stained overnight with Coomassie Blue G250 (Bio-Rad) and destained in ultra-pure water (Millipore). In-gel trypsin digestion of all samples was performed from eight equal sized gel fractions. Each fraction was sliced into 1 mm pieces and placed in a 96-well plate. The gel pieces were destained with 50% (v/v) ACN in 50 mM ammonium bicarbonate until clear, dehydrated in 100% (v/v) ACN, and dried. Sequence-grade porcine trypsin (Promega) (1:30 enzyme/substrate, w/w) was used to digest the proteins overnight at  $37^{\circ}\text{C}$ . Tryptic peptide mixtures were then collected and two rounds of gel extractions of peptides were performed with 2% (v/v) formic acid in 50% (v/v) ACN and 50 mM ammonium bicarbonate. The extracts were combined, peptide mixtures dried by vacuum centrifugation, redissolved in  $10 \mu\text{L}$  0.1% (v/v) formic acid, and desalted as described below. For in-solution digestion, samples were diluted to  $<1$  M urea (final concentration) and trypsinized (sequence-grade porcine trypsin, 1:40 enzyme/substrate, w/w) overnight at  $37^{\circ}\text{C}$ . Following proteolysis, the peptide mixtures were acidified by adding formic acid to a final concentration of 0.1% (v/v). Desalted of peptide mixtures were performed using self-packed  $\text{C}_{18}$  SPE tips. Briefly,  $\text{C}_{18}$  tips were washed three-times with  $20 \mu\text{L}$  100% ACN, three-times with  $20 \mu\text{L}$  50% (v/v) ACN in 0.1% formic acid, and equilibrated with  $50 \mu\text{L}$  0.1% (v/v) formic acid. After sample loading, tips were washed three-times with  $20 \mu\text{L}$  0.1% formic acid. Peptides were eluted with  $20 \mu\text{L}$  60% (v/v) ACN in 0.1% formic acid and  $20 \mu\text{L}$  90% (v/v) ACN in 0.1% formic acid and dried. The desalted fractions were dried and stored at  $-80^{\circ}\text{C}$  until LC-MS/MS.



### LC-MS/MS-BASED *N*-GLYCOMICS

*N*-glycans alditols were separated using a porous graphitized carbon (PGC) LC column [5  $\mu$ m (particle size) Hypercarb KAPPA, 100 mm (length)  $\times$  200  $\mu$ m (ID), 250 Å (pore size), Thermo Scientific] using an Ultimate 3000 HPLC system (Dionex) connected directly to an ESI-MS/MS HCT Ultra ion trap (Bruker Daltonics). Separation was performed using a binary gradient solvent system made up of solvent A (aqueous 10 mM  $\text{NH}_4\text{HCO}_3$ ) and solvent B (90% ACN/10 mM ammonium bicarbonate). The flow rate was 2  $\mu$ L/min and a total gradient of 100 min was programmed as follows: 0–2.5% solvent B for 0–13 min; 2.5–17.5% solvent B for 14–48 min; 17.5–50% solvent B for 48–65 min; 50–100% solvent B for 65–75 min; 100% solvent B for 75–80 min; back to 0% solvent B for 80–85 min, and 100% solvent A equilibration for 15 min. Settings for the MS/MS were as follows: drying gas flow: 6 L/min; drying gas temperature: 300°C; nebulizer gas: 12 p.s.i.; skimmer: –40.0 V; trap drive: –99.1 V; and capillary exit: –166 V. Smart fragmentation was used with start- and end-amplitude of 30 and 200%, respectively. Ions were detected in ion charge control set at 100,000 ions/scan and with maximum accumulation time of 200 ms. MS spectra were obtained in negative ion mode with three scan events: a full scan ( $m/z$  400–2,200) at a scan speed of 8,100  $m/z/s$  and data-dependent MS/MS scans after CID fragmentation of the top two most intense precursor ions with an absolute intensity threshold of 30,000 and a relative intensity threshold of 5% relative to the base peak. Dynamic inclusion was inactivated to ensure MS/MS generation of closely eluting *N*-glycan isomers. Precursors were observed mainly in charge states  $Z = -1$  and/or  $-2$ . Mass accuracy calibration of the mass spectrometer was performed using a well-defined tune mix (Agilent) prior to acquisition. *N*-glycans released from bovine fetuin served as positive controls for the sample preparation and the LC-MS/MS performance. Differences between observed and theoretical precursor and fragment masses were generally <0.2 Da. Three LC-MS/MS technical replicates were performed for the subcellular fractions.

### LC-MS/MS-BASED PROTEOMICS

Three LC-MS/MS technical replicates of the subcellular proteomes of the breast cells were analyzed using a Q-Exactive (Thermo Scientific). Peptide mixtures in 0.1% (v/v) formic acid were loaded onto a  $\text{C}_{18}$  reversed phase column packed in-house [2.7  $\mu$ m (particle size) HaloLink Resins, Promega, column dimensions: 100 mm (length)  $\times$  75  $\mu$ m (ID)]. Separation of peptides was performed over a 60 min gradient with the first 50 min of the linear gradient increasing from 0 to 50% in solvent B [0.1% (v/v) aqueous formic acid/100% (v/v) ACN] and then to 85% solvent B for the next 2 min and maintained at 85% for 8 min. The flow rate was constant at 300 nL/min. The Easy-nLC (Thermo Scientific) was connected directly to the nano-ESI source of the Q-Exactive. MS full scans were acquired with resolution of 35,000 in the positive ion mode over  $m/z$  350–2,000 range and an automatic gain control (AGC) target value of  $1 \times 10^6$ . The top 10 most intense precursor ions were then isolated for MS/MS using higher energy collisional dissociation fragmentation at 17,500 resolution with the following settings: collision energy: 30%; AGC target:  $2 \times 10^5$ ; isolation window:  $m/z$  3.0; and dynamic exclusion enabled. Precursors with

unassigned or  $Z = +1$  charge states were ignored for MS/MS selection.

The subcellular proteomes of the colon cells were LC-MS/MS analyzed using a Triple TOF 5600 (ABSciex). Peptides were separated by a nanoLC system (Eksigent) on a  $\text{C}_{18}$  reversed phase column [ProteCol 100 mm (length)  $\times$  150  $\mu$ m, (ID): 3  $\mu$ m (particle size), 300 Å (pore size); SGE Analytical Science] with a 90 min gradient from 5 to 40% solvent B [90% (v/v) ACN with 0.1% formic acid] at a constant flow rate of 600 nL/min. The top 10 most intense precursor ions with  $Z = +2$ ,  $+3$ , and  $+4$  were selected for MS/MS using CID fragmentation.

### ANALYSIS OF *N*-GLYCOME LC-MS/MS DATA

*N*-glycome raw data for all subcellular glycoproteomes were viewed and manually analyzed using DataAnalysis v4.0 (Bruker Daltonics). Monoisotopic masses were obtained and searched against GlycoMod<sup>1</sup> to obtain possible monosaccharide compositions, which were subsequently verified manually by *de novo* sequencing of corresponding MS/MS spectra and by taking account of PGC chromatographic retention time. The glycan type and the terminating monosaccharide determinants could unambiguously be identified using this method (27). The relative abundances of the observed *N*-glycans were determined using the ratio of the extracted ion chromatogram (EIC) peak area of each *N*-glycan species over the sum of EIC peak areas of all observed *N*-glycans in the sample. This has been shown to be a reasonably accurate method for relative *N*-glycan quantitation (35). The extent of *N*-glycan processing was measured by evaluating the relative molar proportion of the relative unprocessed species (i.e., immature mono-glucosylated glycans and high-mannose type *N*-glycans) and the processed species (i.e., hybrid, complex, and paucimannose type *N*-glycans) of the total *N*-glycome. In addition, the degree of monosaccharide determinants including  $\alpha$ 1,2/3/6-mannose,  $\beta$ 1,3/4-galactose,  $\alpha$ 1,3/4/6-fucose, and  $\alpha$ 2,3/6-sialic acid terminating *N*-glycans were calculated as a relative molar abundance of both the entire *N*-glycome and of the potentially modified *N*-glycan substrates (e.g., complex/hybrid-types). Since multiple determinants may be displayed by a given *N*-glycan, the total summed to more than 100%.

### ANALYSIS OF LC-MS/MS-BASED PROTEOMIC DATA AND GENE ONTOLOGY

For breast cell proteomes, raw spectra were converted to .mgf files using Proteome Discoverer Daemon v1.3 (Thermo Scientific) and searched against SwissProt protein database (*Homo sapiens*, 20,279 reviewed entries, November 2013 release) using the Global Proteome Machine (Cyclone). The following search criteria were used: carbamidomethylation was a fixed modification and oxidation and deamidation were variable modifications for methionine and asparagine/glutamine residues, respectively. Mass tolerances of 10 ppm and 0.02 Da were selected for precursor and product ions, respectively, with a maximum of two missed tryptic cleavages.

For colon cell proteomes, MS/MS spectra were extracted by ProteinPilot v4.2 (ABSciex) and searched using Mascot v2.4.0

<sup>1</sup><http://web.expasy.org/glycomod/>

(Matrix Science) against SwissProt protein database (*Homo sapiens*, 20,253 entries, April 2013 release) using trypsin as the digestion enzyme. Precursor and product ion tolerances were 20 ppm and 0.50 Da, respectively. Oxidation of methionine residues and carbamidomethylation of cysteine residues were used as variable modifications.

Scaffold v4.2.1 (Proteome Software) was used to validate MS/MS-based peptide and protein identifications. Peptides were accepted if they were confidently identified at  $\geq 95.0\%$  probability as evaluated by the local false discovery rate (FDR) algorithm. Proteins were included if they were confidently identified at  $\geq 99.0\%$  probability as assigned by the Protein Prophet algorithm incorporated in the software. Proteins containing shared or similar peptides, and which could not be differentiated based on MS/MS analysis alone, were grouped to satisfy the principles of parsimony. Proteins, which confidently shared identified peptides were grouped into clusters. Proteins were annotated using gene ontology (GO) terms from NCBI. The protein identifications were stringently filtered based on the presence of a minimum of two peptides in all replicates. The relative abundances of proteins were determined by conventional spectral counting and adjusted by taking the polypeptide length into account. Putative *N*-glycoproteins in the proteome of the subcellular fractions were predicted *in silico* based on the presence of one or more sequons (NxT/S,  $x \neq P$ ) and a signal peptides (for secreted proteins) and/or transmembrane regions (for cell-surface and microsome proteins) using prediction tools including SignalP (v4.1) (36), Transmembrane Hidden Markov Model (TMHMM v2.0) (37), PrediSi (38), and Phobius (39). Mitochondrial and nuclear membrane proteins were excluded as these are unlikely to enter the ER–Golgi glycosylation pathway. Ambiguous assignments were manually checked (validated or discarded) with information from Uniprot. Potential sequons were obtained using NetNGlyc (40). These *in silico* prediction tools generated lists of experimentally validated and putative glycoproteins. The 100 most abundant glycoproteins in each subcellular fraction were used to assess glycosylation site accessibility. The contribution of these glycoproteins to the total glycoproteome in each sample was estimated by multiplying the normalized spectral count of the individual glycoproteins with their potential glycosylation sites, a measure termed “sequon-weighted normalized spectral count.”

#### SELECTION OF PDB 3D STRUCTURE FOR GLYCOSYLATION SITE ACCESSIBILITY DETERMINATION

Three-dimensional protein structures were obtained from the protein data bank (PDB) database<sup>2</sup>. If multiple structures were available for a glycoprotein, the best match to the naturally occurring variant was chosen by considering the following parameters in a prioritized order: (1) high protein sequence coverage and resolution of the 3D structure, (2) source of protein (purified from organism/tissue over artificial expression system), (3) known site-specific mutations, (4) presence of artificial/natural ligands, and (5) oligomerization of the solved 3D structure. The experimentally obtained PDB structures used in this study were all based

on X-ray crystallography, Table S2 in Supplementary Material. Where no experimentally determined structures were available (43%), structure homologs were obtained from ProteinModel-Portal<sup>3</sup>, Swiss-model repository<sup>4</sup>, or ModBase<sup>5</sup>. High sequence homology was used as a selection criterion when choosing homology model. The average sequence homology for all structures was 67%, which is considered very reliable for homology modeling (41), Table S1 in Supplementary Material. 3D protein structures were viewed with RasMol v2.7.5 (RasWin Molecular Graphics) for visual inspection.

#### GLYCOSYLATION SITE ACCESSIBILITY DETERMINATION FROM MATURELY FOLDED GLYCOPROTEINS

The glycosylation site solvent accessibility was determined by measuring the accessibility to the individual asparagine residues forming the glycosylation sites using NACCESS<sup>6</sup> (42), an accurate and frequently used solvent accessibility determination program (19, 43–45). NACCESS calculates the atomic accessible area by predicting van der Waal's interactions when a probe is rolled around on the protein surface (46, 47). The maximum probe size offered by the program (5 Å radius) was used as a default in this study to simulate as closely as possible the accessibility of the glycosylation enzymes to the glycosylation sites. NACCESS produces unit-less and absolute accessibility values as the output format (denoted “arb. units”), which are comparable between glycosylation sites of different glycoproteins (19). Prior to the measurements of site accessibility, any water molecules, sugars, ligands, and other hetero-atoms/molecules, not part of the core polypeptide chain, were removed from the protein surface. Negligible accessibility differences were observed for the “native” and the monomeric form of glycoproteins with quaternary structures (data not shown). Hence, in the case of multimers, glycosylation site solvent accessibilities derived from the monomeric structures were not considered in the analysis.

#### STATISTICAL ANALYSIS

All relative abundances of *N*-glycans were presented as a percentage out of 100% as mean  $\pm$  SD. Glycosylation site accessibilities were presented as mean  $\pm$  SEM to illustrate the potential spread of mean instead of the individual data points, which can be hugely influenced by the (local) accuracy and quality of the PDB structures. To overcome this potential issue of PDB “noise,” relative large numbers of data points ( $n$ ) were needed. Data were analyzed using Prism v6 (GraphPad). One-way ANOVA analysis was performed for statistical comparison between the three subcellular fractions followed by *post hoc* Tukey's tests. All  $p$  values were adjusted taking into account the multiple comparisons made and reported as multiplicity adjusted  $p$  values.  $p < 0.05$  was regarded as statistically significant and indicated with “\*.” Stronger statistical significance was indicated as follows: \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; \*\*\*\* $p < 0.0001$ . Simple linear regression and corresponding correlation coefficients ( $R^2$ ) were obtained to evaluate the relationship between

<sup>2</sup><http://www.rcsb.org/pdb>

<sup>3</sup><http://www.proteinmodelportal.org>

<sup>4</sup><http://swissmodel.expasy.org>

<sup>5</sup><http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi>

<sup>6</sup><http://wolf.bms.umist.ac.uk/naccess/>

the degree of *N*-glycan processing in terms of glycan type and expression of terminal glycan determinants and the glycosylation site solvent accessibility.

## RESULTS

### SUBCELLULAR-SPECIFIC *N*-GLYCOSYLATION OF HUMAN BREAST EPITHELIAL CELLS

Label-free quantitative *N*-glycome mapping of the secreted and microsome (total membrane) subcellular glycoproteomes of a panel of eight cultured human breast cells (i.e., MCF7, SKBR3, MDA157, MDA231, MDA468, HS578T, HMEC, and MCF10A) displaying diverse cellular features showed differential *N*-glycan processing of the two fractions, **Figure 1A**. The glycoproteins secreted into the cultured media consistently displayed a significantly higher proportion of processed *N*-glycan types (i.e., hybrid, complex, and paucimannose) (74.2–95.0% mol/mol of total *N*-glycome) than the high-mannose-rich microsomal subcellular glycoproteomes (22.1–55.6%,  $p < 0.0001$ –0.05). Little, if any, correlation between the *N*-glycan processing stage and the cellular doubling time ( $R^2 = 0.13$ ) or the protein secretion rate ( $R^2 = 0.35$ ), respectively, was detected of the secreted glycoproteomes across the cell line panel, Figure S1 in Supplementary Material. No correlation was detected between the *N*-glycan processing stage of the microsomal glycoproteins and the cellular doubling time ( $R^2 = 0.04$ ) or the protein secretion rate ( $R^2 = 0.01$ ).

In-depth, *N*-glycan profiling of the secreted, microsomal, and cell-surface enriched glycoproteomes was carried out for MCF7, MDA468, and MCF10A cells as representative cells for the breast cell line panel. Differential *N*-glycan processing was evident as exemplified by the clear differences seen in the *N*-glycome *m/z* profiles of the three subcellular fractions of MCF7 cells, **Figure 1B**. The cell-surface glycoproteins derived from MCF7 and MDA468 (but not MCF10A) cells were subjected to more *N*-glycan processing than microsomal proteins ( $p < 0.01$ –0.05) and all the three cell lines showed significantly increased abundance of the more processed *N*-glycans on the secreted proteins ( $p < 0.0001$ –0.01), **Figure 1C**.

### SUBCELLULAR-SPECIFIC DISTRIBUTION OF *N*-GLYCAN DETERMINANTS

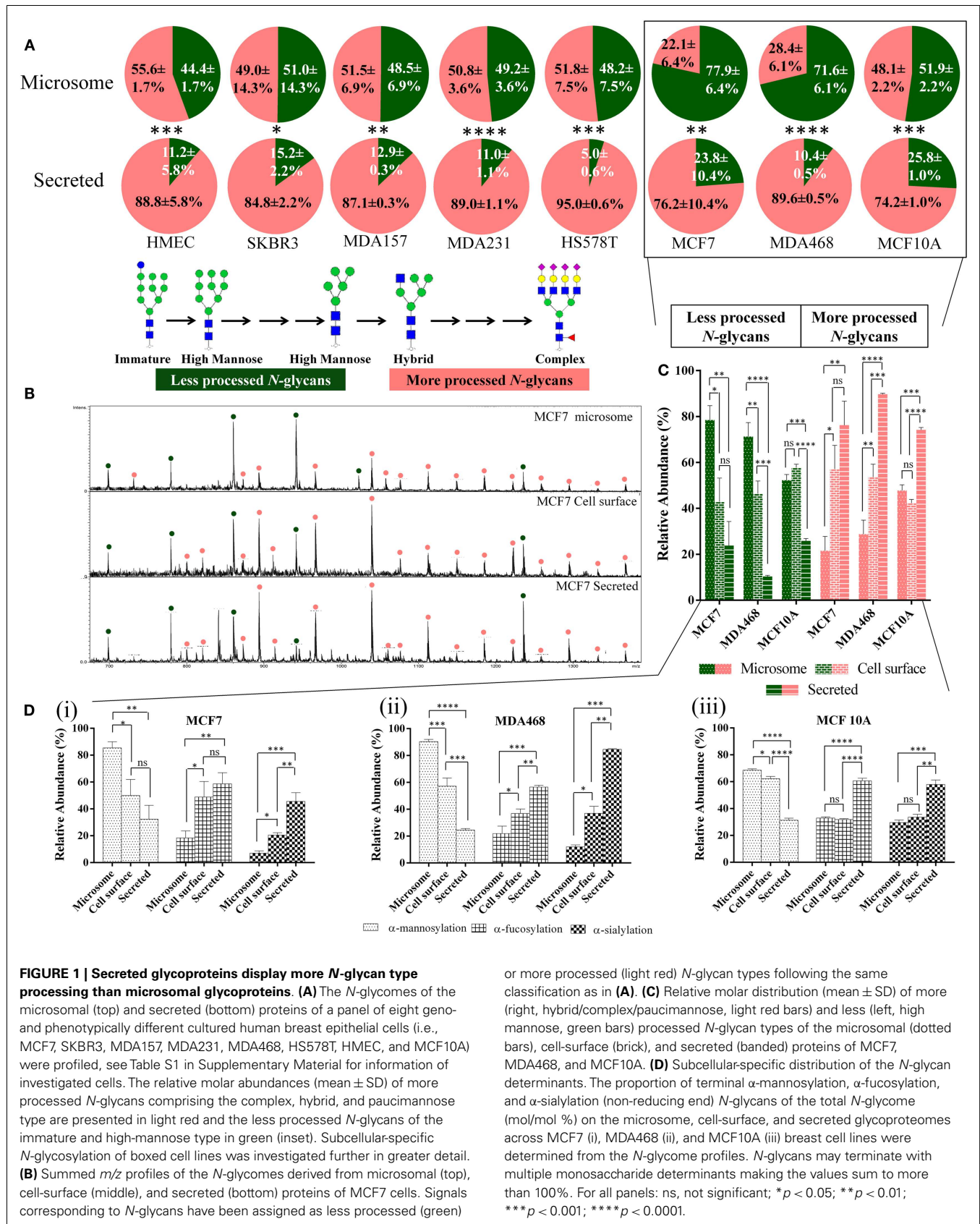
To further evaluate the subcellular-specific distribution of common *N*-glycosylation determinants, which may be recognized by different immuno-lectins, terminal  $\alpha$ -mannose,  $\alpha$ -fucose, and  $\alpha$ -sialic acid residues were mapped based on the obtained *N*-glycome profiles, **Figure 1D**. As expected from the glycan type distribution, terminating  $\alpha$ -mannosylation was found to be significantly reduced on the secreted and cell-surface proteins relative to the microsomal proteins. The  $\alpha$ -fucosylation, primarily of the  $\alpha 1,6$ -(core) type, and  $\alpha 2,3/6$ -sialylation were concomitantly significantly higher in the secreted fractions than in the cell-surface-enriched fraction (with the exception of fucosylation of MCF7) and in the microsomal fraction of all three cell lines. Taking the incomplete subcellular fractionation into account (see “Proteomics- and GO-Based Assessment of Subcellular Fractionation”), we estimate that very little terminal  $\alpha$ -mannosylation is present on protein *N*-glycans in contact with the extracellular environment in the investigated cells and that

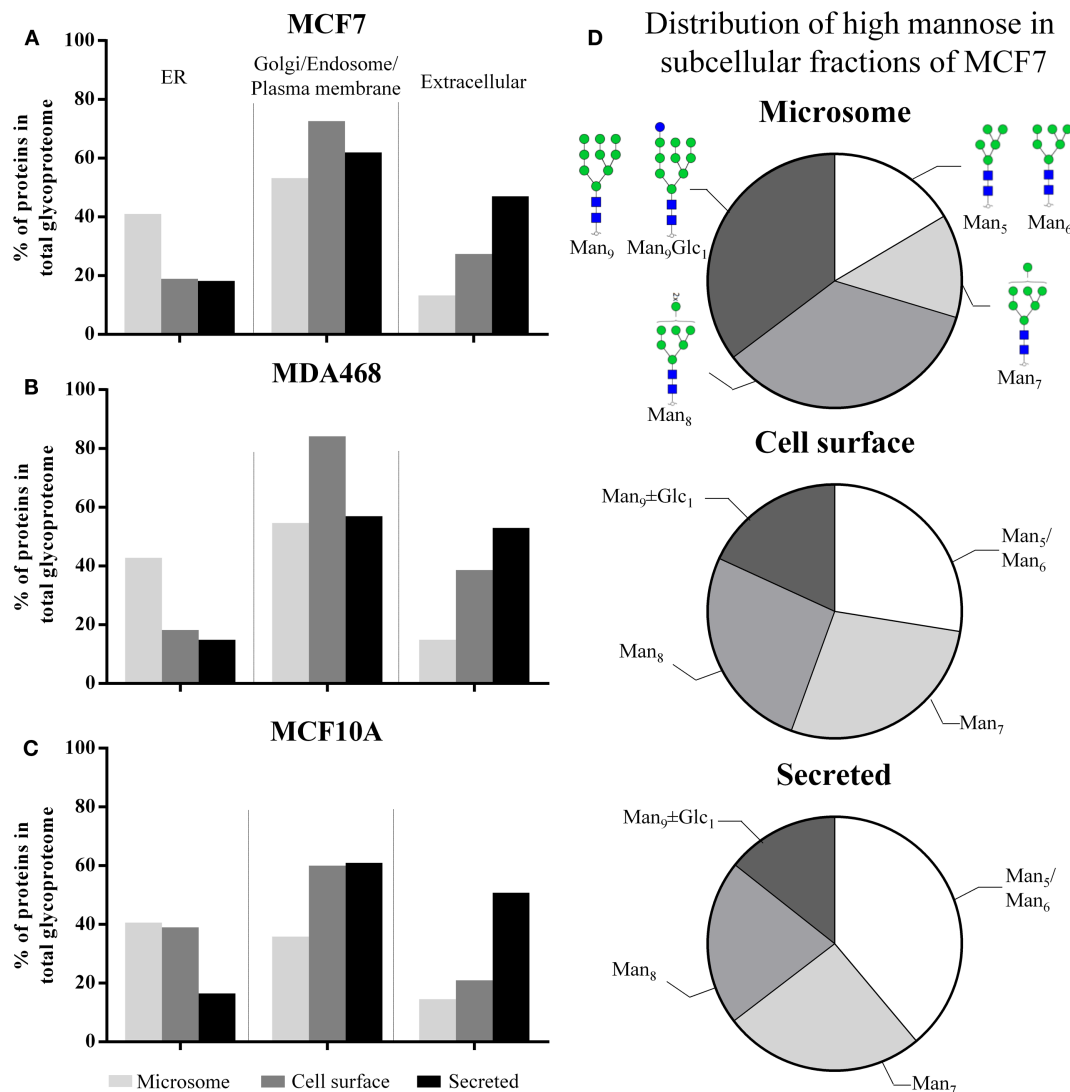
little  $\alpha$ -sialylation and  $\alpha$ -fucosylation are carried by intracellular (microsomal) *N*-glycoproteins.

### PROTEOMICS- AND GO-BASED ASSESSMENT OF SUBCELLULAR FRACTIONATION

In total, 2,297, 2,636, and 2,042 human proteins were identified across the three subcellular fractions in MCF7, MDA468, and MCF10A, respectively. Putative *N*-glycoproteins fulfilling our strict prediction criteria i.e., presence of the following: one or more sequons (NxT/S, x  $\neq$  P); and signal peptides (for secreted proteins); and/or transmembrane regions (for membrane-tethered proteins) comprised significant proportions of the subcellular proteomes (15.7–31.0%), Table S3A in Supplementary Material. The GO terms “ER”, “Golgi/endosome/plasma membrane”, and “extracellular” were used to evaluate the localization/origin of the glycoproteins identified in the subcellular fractions. In agreement with a previous study (23), the GO annotation of the identified proteins showed that the microsomes in general contained a high proportion of ER-residing proteins, **Figures 2A–C**. Although the proteins are only broadly, and possibly somewhat inaccurately, classified on the basis of GO terms, the trends clearly indicated significant enrichment, although not complete isolation, of the desired proteins in the respective subcellular fractions. The ER-based contribution to the microsome was supported by the fact that a significant proportion of the high-mannose *N*-glycans identified in this fraction were of the immature type i.e., Man<sub>9</sub>  $\pm$  Glc<sub>1</sub> (MCF7: 35.3  $\pm$  0.9%, MDA468: 40.2  $\pm$  2.0%, and MCF10A: 31.8  $\pm$  0.4%, mol/mol of the total high-mannose *N*-glycans), **Figure 2D** (MCF7 data) and Figure S2 in Supplementary Material (MDA468 and MCF10A data).

To further investigate the intracellular *N*-glycosylation and confirm the presence of ER-rich microsomes, the *N*-glycome and proteome of ER- and Golgi-enriched fractions of human colon epithelial cancer cells (SW480) as prepared by the method of sucrose density gradient centrifugation, were mapped and compared to the microsome profiles derived from the same cells, Figure S3A in Supplementary Material. Quantitative analysis of four reliable and representative markers of the ER (i.e., 78 kDa glucose-regulated protein, protein disulfide bond isomerase, calreticulin, and protein transport protein Sec61 alpha isoform 1) and Golgi (i.e., polypeptide *N*-acetylgalactosaminyltransferase 2,  $\beta$ -1,4-galactosyltransferase 1, Golgi apparatus protein 1, and Golgi membrane protein 1) compartments revealed a high abundance of ER-specific proteins in the ER-enriched fraction, Figure S3B in Supplementary Material. However, there was still a significant presence of ER proteins in the Golgi-enriched and microsome fractions. In contrast, the ER-enriched and microsome fractions were essentially free of Golgi proteins, Figure S3C in Supplementary Material. In line with our breast epithelial cell data, the proteins in the ER-enriched fraction contained a significantly higher degree of high-mannose (Glc<sub>0-1</sub>Man<sub>5-9</sub>GlcNAc<sub>2</sub>) (92%) *N*-glycans than the proteins in the microsome (75%) and the Golgi-enriched fraction (51%). Taken together, the data confirm that the microsomes of human breast and colon epithelial cells predominantly contain ER proteins and that such intracellular proteins mostly carry high-mannose type *N*-glycosylation. Since the Golgi fraction contains few, if any, ER proteins, it becomes clear that the majority of post-ER *N*-glycans are of the complex type.





**FIGURE 2 | (A–C)** The subcellular proteomes of MCF7, MDA468, and MCF10A breast epithelial cell lines were mapped according to GO terms into ER, Golgi/endosome/plasma membrane, and extracellular region classifiers. This confirmed enrichment, but not isolation, of cell-surface and secreted proteins in the respective subcellular fractions. In addition, the classification confirmed that the microsomes contained a significant proportion of ER-residing

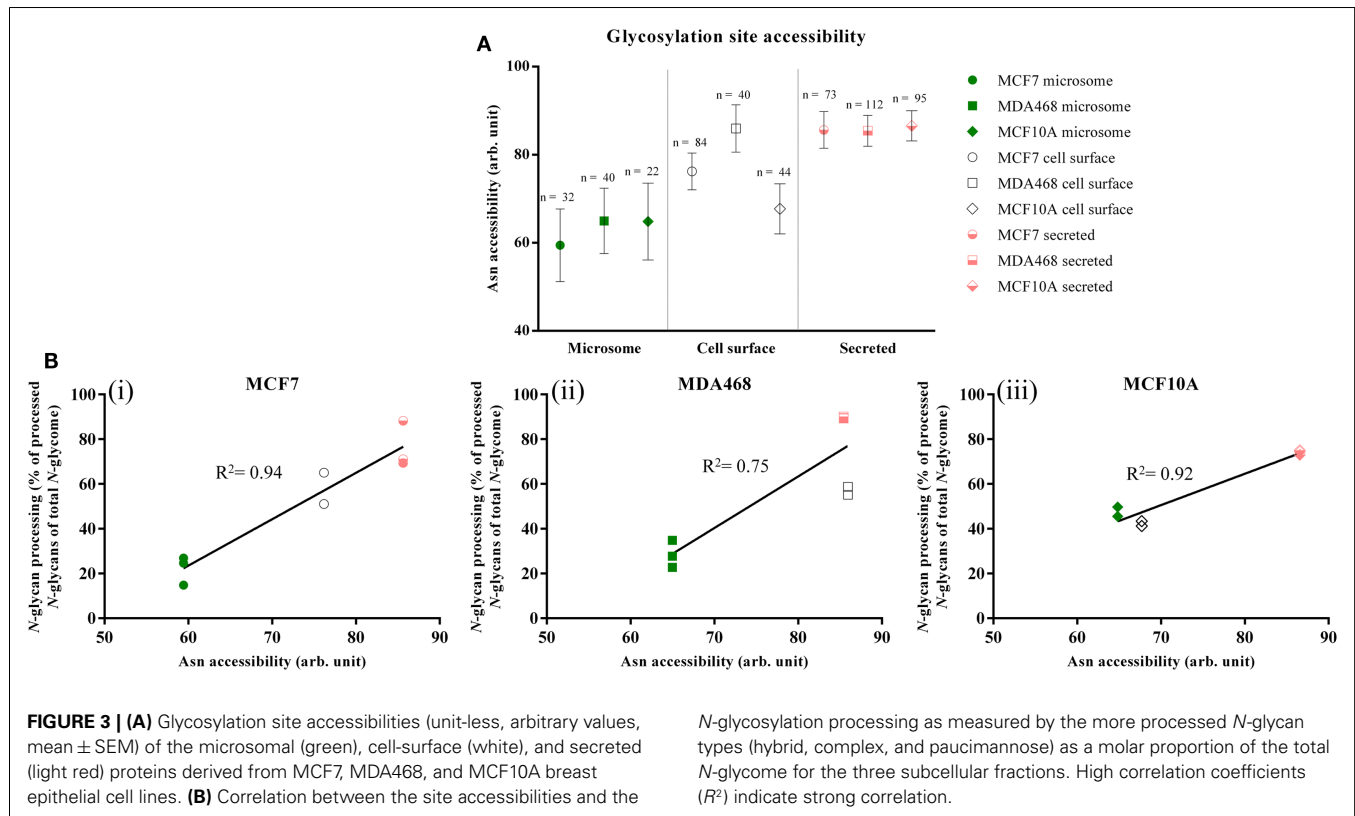
proteins. **(D)** The subcellular distribution of the high-mannose glycan type series on proteins derived from MCF7 into Man<sub>5</sub>, Man<sub>6</sub>, Man<sub>7</sub>, Man<sub>8</sub>, Man<sub>9</sub>±Glc<sub>1</sub>, the latter representing immature *N*-glycans normally only associated with intracellular ER *N*-glycosylation. See Figure S2 in Supplementary Material for the subcellular distribution of the high-mannose glycan type series of MDA468 and MCF10A.

### DIFFERENTIAL ASN SITE ACCESSIBILITIES EXPLAIN SUBCELLULAR-SPECIFIC *N*-GLYCOSYLATION

To investigate a possible link between the observed subcellular-specific *N*-glycosylation and protein *N*-glycosylation site accessibility, *in silico* assessment of site accessibility was performed of the identified proteins predicted to be *N*-glycosylated. Due to the laborious and time-consuming approach of determining glycoprotein site accessibility (19), only the most abundant subset of the putative *N*-glycoproteins observed in the subcellular fractions were included in the accessibility assessment. The relative abundances of the individual putative glycoproteins were calculated by a conventional normalized spectral counting strategy; however, the number of sequons of the individual proteins was factored

into the calculation to ensure a fair representation of heavily and lightly *N*-glycosylated proteins. We call this term “sequon-weighted normalized spectral counts.” Based on sequon-weighted normalized spectral counts, the 100 most abundant glycoproteins uniquely present in the three subcellular fractions, which, by weight, comprised 70–100% of the individual subcellular glycoproteomes, were used to assess glycosylation site accessibility, Table S3B in Supplementary Material. The solvent site accessibilities were determined using an established approach based on van der Waal interactions of the asparagine residue of the glycosylation sites to solvent (19). 3D-glycoprotein structures (experimental or homology modeled) were available for approximately one-third of the 189, 89, and 183 putative *N*-glycoproteins identified





uniquely in the microsome, cell-surface, and secreted fraction, respectively, Figure S4 in Supplementary Material. This yielded site-accessibility datasets covering in total 161 (microsome), 189 (cell-surface), and 236 (secreted) *N*-glycosylation sites from the three cell types.

Differential site accessibilities were observed for the three subcellular glycoproteomes for all three investigated breast cell lines, **Figure 3A** (see also Figures S5A–C in Supplementary Material for an alternative representation showing 95% confidence intervals). Glycosylation sites of secreted glycoproteins were significantly more accessible [MCF7:  $85.63 \pm 35.47$ ,  $n = 73$ ; MD468:  $85.44 \pm 36.85$ ,  $n = 112$ ; MCF10A:  $86.56 \pm 33.54$  (all unit-less arbitrary values),  $n = 95$ ] than sites on microsomal proteins (MCF7:  $59.44 \pm 46.58$ ,  $n = 32$ ; MD468:  $64.98 \pm 46.99$ ,  $n = 40$ ; MCF10A:  $64.84 \pm 40.97$ ,  $n = 22$ ,  $p < 0.01$ ). In agreement with the *N*-glycomes that carried a mixture of less processed high-mannose and more processed *N*-glycan types, the sites of cell-surface proteins were intermediately accessible: cell-surface sites were either statistically similar in accessibility to the microsomal protein sites (MCF10A:  $67.70 \pm 37.66$ ,  $n = 44$ ) or similar to the secreted protein sites (MCF7:  $76.20 \pm 38.13$ ,  $n = 84$ ; MD468:  $85.95 \pm 34.08$ ,  $n = 40$ ). For all three breast cell lines, the glycosylation site accessibilities were strongly correlated with the *N*-glycan processing as measured by their glycan type (MCF7:  $R^2 = 0.94$ ; MD468:  $R^2 = 0.75$ ; MCF10A:  $R^2 = 0.92$ ), **Figure 3B**. Higher average glycosylation site accessibility of the secreted and partly also the cell-surface glycoproteins resulted, as such, in more *N*-glycan processing in terms of glycan type formation.

Other subcellular-specific *N*-glycosylation signatures including core fucosylation,  $\beta$ -galactosylation, and  $\alpha$ -sialylation were found to correlate only weakly or not at all with glycosylation site accessibility upon search for consistent trends across the three different cell lines, Table S4 in Supplementary Material.

## DISCUSSION

### SUBCELLULAR-SPECIFIC PROTEIN *N*-GLYCOSYLATION OF HUMAN CELLS

All *N*-linked glycoproteins synthesized by a given cell are processed by a common glycosylation machinery. Despite this shared biosynthetic machinery, we observed that a panel of human breast epithelial cells of different geno- and phenotypes, reproducibly produced subcellular glycoproteomes with distinct *N*-glycosylation signatures. The *N*-glycans attached to proteins enriched from the cell-surface, and in particular the secreted glycoproteins, were significantly more processed with respect to their glycan type (i.e., hybrid/complex/paucimannose) than the predominantly high-mannose type microsomal proteins for all investigated cells. As such, subcellular-specific *N*-glycosylation can be predicted to be a general cellular feature not restricted to the investigated breast epithelial cells. Deeper dissection of the intracellular organelle-specificity of colon cell *N*-glycosylation supported this concept. The capacity of human cells to generate multiple subcellular glycoproteomes displaying specific *N*-glycosylation profiles has, to the best of our knowledge, not been systematically investigated.

The importance of cell-surface *N*-glycosylation for cell–cell and cell–protein interactions has prompted several investigations

of the cell-surface (alternatively termed plasma membrane) *N*-glycosylation. High-mannose type *N*-glycans, in particular Man<sub>8-9</sub> structures, were previously reported to be the dominating features of the plasma membrane of human embryonic stem cells (48) and of cancer cells (49, 50). However, cell lysates and total membrane fractions similar to our microsome preparations were used in these studies suggesting significant contributions from intracellular high-mannose-rich ER-residing *N*-glycoproteins (23). Hence, the actual cell-surface *N*-glycomes in the previous work may not have been accurately captured. Specific cell-surface enrichment methods such as biotinylation labeling strategies used in this study or adhesion-based isolation methods (23) indicate that human cell-surfaces instead are generally decorated with more processed *N*-glycan types.

Of the six cancerous breast cells investigated in this study, only MCF7 and MDA468 displayed predominantly (>70%) high-mannose *N*-glycans of the microsomal proteins. Approximately equal distribution of high-mannose and the more processed *N*-glycan types of microsomal proteins were detected in the remaining four cancerous (SKBR3, MDA157, MDA231, and HS578T) and the two non-cancerous cells (HMEC and MCF10A). In addition, no consistent over-representation of high-mannose *N*-glycans were detected for the secreted proteins derived from the cancerous cell lines relative to the non-cancerous cell lines. Together this indicates that high-mannose *N*-glycosylation is not linked directly to tumorigenesis. Others have associated serum-derived high-mannose *N*-glycoproteins to pathogenesis including cancer and inflammation (5, 51); however, whether these under-processed species are a result of leakage of intracellular glycoproteins as a consequence of cell death or active cellular secretion from intact cells remains to be described. Based on in-depth comparative analysis of the *N*-glycomes derived from secreted proteins of breast and colon epithelial cells of non-cancerous and cancerous nature, we have recently identified several tumor- and sub-type specific *N*-glycosylation signatures amongst the complex *N*-glycans including alterations of sialylation,  $\alpha$ 1,6-fucosylation, and bisecting  $\beta$ 1,4-GlcNAcylation (submitted) (52).

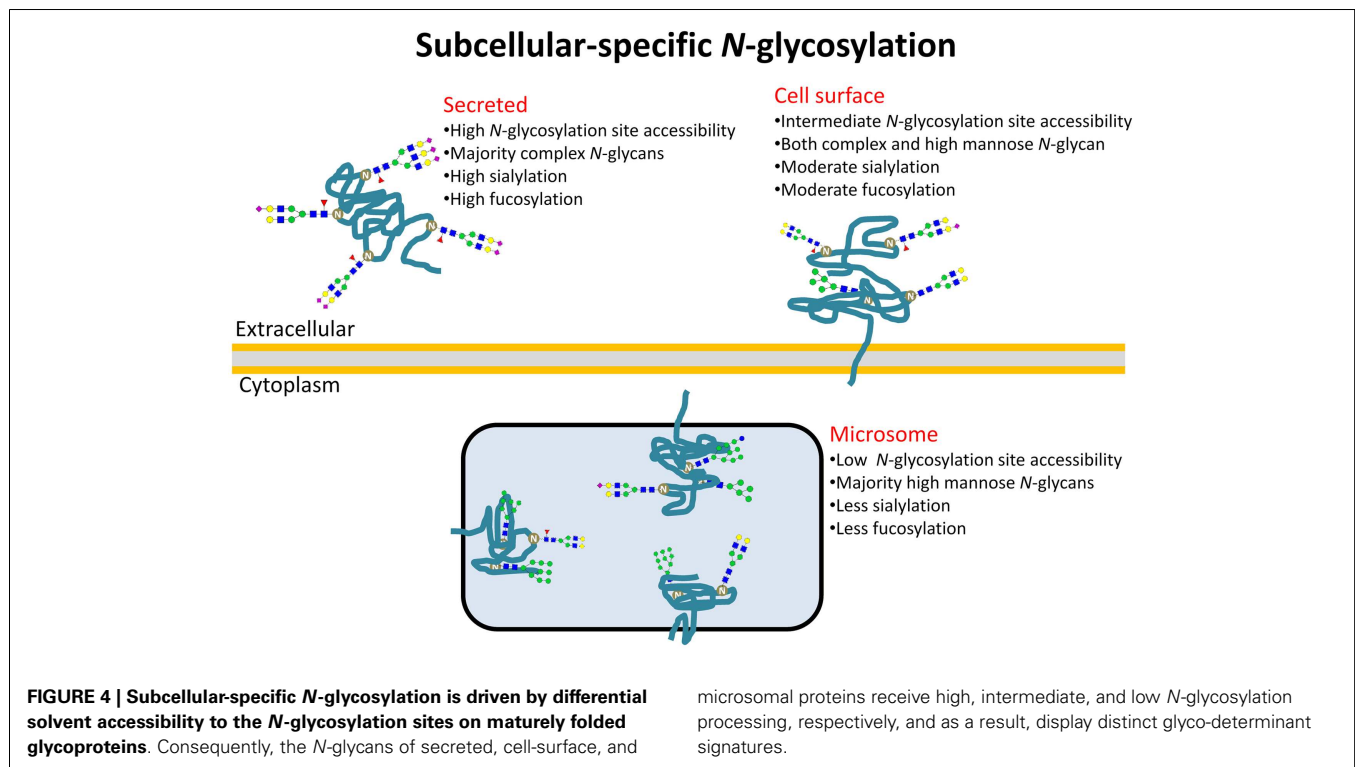
#### **SITE ACCESSIBILITIES MECHANISTICALLY EXPLAIN SUBCELLULAR-SPECIFIC *N*-GLYCOSYLATION**

We have previously shown that solvent accessibility of the glycosylation site of *N*-glycoproteins is an important factor in generating protein- and site-specific *N*-glycosylation (19). We used literature-based glycoproteomics of more than 100 mammalian glycoproteins produced under different cellular and physiological conditions to establish that site accessibility of maturely folded glycoproteins correlates with *N*-glycan processing features including glycan type,  $\alpha$ 1,6-fucosylation and  $\beta$ 1,4/6-GlcNAc-branching. We emphasized in that study that relatively large datasets were required to compensate for the potential inaccuracy of the individual PDB structures and the relative simplistic solvent accessibility assessment simulating the accessibility of the processing glycosylation enzymes to the protein glycosylation sites.

Herein, we used a similar approach using our own *N*-glycosylation data acquired from eight cell lines fractionated into subcellular glycoproteomes to further explore the determining features of site-specific *N*-glycosylation in the context of subcellular

localization of proteins. Homogenous cell cultures were an essential tool to ensure that the isolated subcellular glycoproteomes were produced simultaneously under the same physiological conditions of the glycosylation machinery. Although the *N*-glycomes, as expected, varied considerably between the different cell lines, our experimental data not only validated the strong correlation of the *N*-glycan type and the glycosylation site accessibility of maturely folded glycoproteins in agreement with our previously report (19), but also mechanistically explained that subcellular-specific *N*-glycosylation is driven by differences in site accessibilities of the individual glycoproteins ending up at different subcellular destinations, **Figure 4**. Intracellular (microsome) *N*-glycoproteins receive little glycan processing of the high-mannose intermediates as a result of limited site accessibility, whereas the secreted *N*-glycoproteins are modified almost entirely to more processed *N*-glycan types due to high site accessibilities. As such, *N*-glycan processing may be a targeting signal or a requirement for intracellular (ER–Golgi-residing) glycoproteins to translocate to the surface for cell-surface integration/secretion via vesicles. Keeping in mind there may be many exceptions to the molecular trends presented here, it is tempting to view the glycosylation site accessibility, and, thus, the *N*-glycan type, as a crude predictor of subcellular location of human glycoproteins.

We have previously linked core fucosylation to glycosylation site accessibility (19). Interestingly, glycosylation site accessibility alone could not explain the differential core fucosylation of the subcellular fractionated proteins in our data: the secreted proteins did not have a higher degree of core fucosylation of complex/hybrid-type *N*-glycans than the cell-surface proteins although the secreted proteins had significantly higher accessibilities. This surprising observation may be explained by a possible advantage of the membrane-embedded cell-surface glycoproteins to achieve preferential interaction with the membrane-bound fucosyltransferase 8 (FUT8) facilitating the addition of  $\alpha$ 1,6-fucose residues to the chitobiose cores of *N*-glycans. Soluble (luminal) glycoproteins may be less likely to interact with FUT8. This explanation is congruent with our previous observation describing FUT8 discrimination of soluble *N*-glycoproteins over membrane *N*-glycoproteins (19). Similar processing preference was not observed for the multiple processing enzymes responsible for the formation of the glycan type. As expected, the glycan modification more distal from the protein surface i.e.,  $\beta$ -1,3/4-galactosylation and  $\alpha$ 2,3/6-sialylation were not found to be correlated with glycosylation site accessibility since the glycosyltransferases most likely have unhindered access to the substrates relatively far from the protein surface. By the same token, we cannot rule out that a more refined accessibility determination approach, which not only takes into account the glycosylation site solvent accessibility, but also the conjugated *N*-glycans (53–56), may expose that other subcellular-specific *N*-glycan features correlate with site accessibility. New developments in glycoproteomics may also support and strengthen these observations by giving more accurate insight into the connectivity of glycosylation of the individual protein carriers (31). Finally, it should be emphasized that although the subcellular glycoproteomes share a common biosynthetic machinery, slightly different trafficking rates and/or routes to their final destinations are factors that may



contribute to yield distinct subcellular *N*-glycosylation. Other cellular factors including the glycosylation enzyme activity or the availability of nucleotide donors may also indirectly contribute to subcellular-specific *N*-glycosylation by having differential effects on the individual subcellular glycoproteomes.

#### SUBCELLULAR-SPECIFIC GLYCO-DETERMINANTS IN IMMUNITY

The distinct *N*-glycosylation signatures carried by the subcellular glycoproteomes may be functionally important in immunity if we consider the key role of *N*-glycans as mediators for an effective innate and adaptive immune response through their specific interaction with endogenous lectins. In addition, opportunistic pathogens often use exposed *N*-glycan determinants as receptors for adhesion using exogenous lectins (11). The observed subcellular-specific glycosylation is here briefly discussed in the context of glyco-immunity and infection; it is stressed that further empirical evidence is required to validate these proposed relationships.

We found that  $\alpha$ -sialylation was a more abundant feature of the secreted *N*-glycoproteins than cell-surface proteins. High sialylation of secreted glycoproteins is essential to mask penultimate galactose residues from being exposed and recognized by asialoglycoprotein receptors, a C-type lectin (12). Thus, the high sialylation of secreted glycoproteins may be a requirement to ensure prolonged circulation half-life. In addition, high sialylation of secreted glycoproteins can act as a strong decoy for the less sialylated cell-surface proteins, to which opportunistic pathogens are known to adhere through sialic acid-recognizing I-type lectins (alternatively termed siglecs) (57, 58). Displaying less-than-complete sialylation of the cell-surface proteins also ensures that a gradient of

biological activity toward endogenous siglecs for cellular signaling and endocytosis (59) is maintained through structural diversity, which may confer an immunological advantage to the host cells (60).

The secreted *N*-glycoproteins were over-represented in  $\alpha$ 1,6-core fucosylation relative to the cell-surface proteins. In line with our previous observations, the higher degree of core fucosylation may serve to either mask hydrophobic patches to regulate stability/solubility of the secreted *N*-glycoproteins (19) or to protect these more exposed proteins from proteolytic degradation in the extracellular environment. It could be speculated that the membrane-embedded nature of cell-surface glycoproteins would make them more stable by not facing solubility issues in their local environment and less vulnerable to proteolytic digestion, thereby having less requirement for steric protection provided by a bulky fucose residue proximal to the protein surface.

We and others have observed that  $\alpha$ -mannose is an unusual terminating structural determinant in the extracellular environment (61, 62). This may partly be explained by the intracellular functions of mannose (and glucose) terminating *N*-glycans (16, 17). The presence of several mannose recognizing lectins in the extracellular environment including mannan binding protein (MBP), DC-SIGN, and macrophage mannose receptors may be relevant in the context of apoptosis when mannose terminating *N*-glycoproteins are exposed to the extracellular environment. In particular, MBP is a key player and a first line of defense in innate immunity, enabling phagocytosis of apoptotic cells through its binding to exposed immature or under-processed glycans or to pathogens carrying mannosylated glycoproteins (63, 64). Hiding mannose inside cells under physiological conditions could thus be viewed as being

critical to avoiding the unnecessary onset of inflammation and auto-immunity. The presence of extracellular  $\alpha$ -mannosylation would, as such, be indicative of pathophysiological conditions. In support of this hypothesis, high-mannose containing glycoforms of intracellular adhesion molecule 1 and EGF receptor on cell-surfaces were shown to contribute to endothelial inflammation (61) and correlated with poor prognosis of various cancers, respectively (61, 62).

It has been noted that the structure and function of the protein N-glycome is different within and outside human cells and that these differences may be shaped by evolutionary forces (60). We are the first to systematically investigate and mechanistically explain some aspects of subcellular-specific N-glycosylation. We conclude that human cells have developed protein structure-specific mechanisms including differential N-glycosylation site accessibilities to generate subcellular glycoproteomes that display distinct N-glycosylation phenotypes using a shared biosynthetic machinery. Establishing this relationship is of general significance to glycobiologists and in particular to molecular immunologists due to the functional relevance of N-glycan determinants acting as ligands for the spectrum of endogenous lectins involved in facilitating an efficient immune response.

## ACKNOWLEDGMENTS

This work was supported by Macquarie University Research Excellence Scheme postgraduate scholarship and ARC Super Science (FS110200026) and Discovery (DP110104958) Grants. Morten Thaysen-Andersen was funded an Early Career Fellowship by Cancer Institute NSW. The authors declare no conflict of interest.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/Journal/10.3389/fimmu.2014.00404/abstract>

## REFERENCES

- Freeze HH. Genetic defects in the human glycome. *Nat Rev Genet* (2006) 7(7):537–51. doi:10.1038/nrg1894
- Freeze HH. Understanding human glycosylation disorders: biochemistry leads the charge. *J Biol Chem* (2013) 288(10):6936–45. doi:10.1074/jbc.R112.429274
- Cylwik B, Naklicki M, Chrostek L, Gruszewska E. Congenital disorders of glycosylation. Part I. Defects of protein N-glycosylation. *Acta Biochim Pol* (2013) 60(2):151–61.
- Venkatakrishnan V, Packer NH, Thaysen-Andersen M. Host mucin glycosylation plays a role in bacterial adhesion in lungs of individuals with cystic fibrosis. *Expert Rev Respir Med* (2013) 7(5):553–76. doi:10.1586/17476348.2013.837752
- Scott DW, Patel RP. Endothelial heterogeneity and adhesion molecules N-glycosylation: implications in leukocyte trafficking in inflammation. *Glycobiology* (2013) 23(6):622–33. doi:10.1093/glycob/cwt014
- Stuchlová Horynová M, Raška M, Clausen H, Novak J. Aberrant O-glycosylation and anti-glycan antibodies in an autoimmune disease IgA nephropathy and breast adenocarcinoma. *Cell Mol Life Sci* (2013) 70(5):829–39. doi:10.1007/s00018-012-1082-6
- Christiansen MN, Chik J, Lee L, Anugraham M, Abrahams JL, Packer NH. Cell surface protein glycosylation in cancer. *Proteomics* (2014) 14(4–5):525–46. doi:10.1002/pmic.201300387
- Boscher C, Dennis JW, Nabi IR. Glycosylation, galectins and cellular signaling. *Curr Opin Cell Biol* (2011) 23(4):383–92. doi:10.1016/j.ceb.2011.05.001
- Schwarz F, Aebi M. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol* (2011) 21(5):576–82. doi:10.1016/j.sbi.2011.08.005
- Helenius A, Aebi M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* (2004) 73:1019–49. doi:10.1146/annurev.biochem.73.011303.073752
- van Kooyk Y, Rabinovich GA. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat Immunol* (2008) 9(6):593–601. doi:10.1038/nri.f.203
- Figdor CG, van Kooyk Y, Adema GJ. C-type lectin receptors on dendritic cells and Langerhans cells. *Nat Rev Immunol* (2002) 2(2):77–84. doi:10.1038/nri723
- Kawasaki T, Li M, Kozutsumi Y, Yamashina I. Isolation and characterization of a receptor lectin specific for galactose/N-acetylgalactosamine from macrophages. *Carbohydr Res* (1986) 151:197–206. doi:10.1016/S0008-6215(00)90340-9
- van Vliet SJ, Saeland E, van Kooyk Y. Sweet preferences of MGL: carbohydrate specificity and function. *Trends Immunol* (2008) 29(2):83–90. doi:10.1016/j.it.2007.10.010
- Rabinovich GA, van Kooyk Y, Cobb BA. Glycobiology of immune responses. *Ann N Y Acad Sci* (2012) 1253:1–15. doi:10.1111/j.1749-6632.2012.06492.x
- Aebi M. N-linked protein glycosylation in the ER. *Biochim Biophys Acta* (2013) 1833(11):2430–7. doi:10.1016/j.bbamcr.2013.04.001
- Aebi M, Bernasconi R, Clerc S, Molinari M. N-glycan structures: recognition and processing in the ER. *Trends Biochem Sci* (2010) 35(2):74–82. doi:10.1016/j.tibs.2009.10.001
- Parodi AJ. Protein glucosylation and its role in protein folding. *Annu Rev Biochem* (2000) 69:69–93. doi:10.1146/annurev.biochem.69.1.69
- Thaysen-Andersen M, Packer NH. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology* (2012) 22(11):1440–52. doi:10.1093/glycob/cws110
- Sumer-Bayraktar Z, Nguyen-Khuong T, Jayo R, Chen DD, Ali S, Packer NH, et al. Micro- and macroheterogeneity of N-glycosylation yields size and charge isoforms of human sex hormone binding globulin circulating in serum. *Proteomics* (2012) 12(22):3315–27. doi:10.1002/pmic.201200354
- Parekh RB, Dwek RA, Thomas JR, Opdenakker G, Rademacher TW, Wittwer AJ, et al. Cell-type-specific and site-specific N-glycosylation of type I and type II human tissue plasminogen activator. *Biochemistry* (1989) 28(19):7644–62. doi:10.1021/bi00445a021
- Rudd PM, Dwek RA. Glycosylation: heterogeneity and the 3D structure of proteins. *Crit Rev Biochem Mol Biol* (1997) 32(1):1–100.
- Mun JY, Lee KJ, Seo H, Sung MS, Cho YS, Lee SG, et al. Efficient adhesion-based plasma membrane isolation for cell surface N-glycan analysis. *Anal Chem* (2013) 85(15):7462–70. doi:10.1021/ac401431u
- Bones J, Mittermayr S, O'Donoghue N, Guttman A, Rudd PM. Ultra performance liquid chromatographic profiling of serum N-glycans for fast and efficient identification of cancer associated alterations in glycosylation. *Anal Chem* (2010) 82(24):10208–15. doi:10.1021/ac102860w
- North SJ, Hitchen PG, Haslam SM, Dell A. Mass spectrometry in the analysis of N-linked and O-linked glycans. *Curr Opin Struct Biol* (2009) 19(5):498–506. doi:10.1016/j.sbi.2009.05.005
- Hua S, Lebrilla C, An HJ. Application of nano-LC-based glycomics towards biomarker discovery. *Bioanalysis* (2011) 3(22):2573–85. doi:10.4155/bio.11.263
- Jensen PH, Karlsson NG, Kolarich D, Packer NH. Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc* (2012) 7(7):1299–310. doi:10.1038/nprot.2012.063
- Parker BL, Thaysen-Andersen M, Solis N, Scott NE, Larsen MR, Graham ME, et al. Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity. *J Proteome Res* (2013) 12(12):5791–800. doi:10.1021/pr400783j
- Nwosu CC, Seipert RR, Strum JS, Hua SS, An HJ, Zivkovic AM, et al. Simultaneous and extensive site-specific N- and O-glycosylation analysis in protein mixtures. *J Proteome Res* (2011) 10(5):2612–24. doi:10.1021/pr2001429
- Zauner G, Koeleman CA, Deelder AM, Wührer M. Nano-HPLC-MS of glycopeptides obtained after nonspecific proteolysis. *Methods Mol Biol* (2013) 951:113–27. doi:10.1007/978-1-62703-146-2\_9
- Thaysen-Andersen M, Packer NH. Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteomes. *Biochim Biophys Acta* (2014) 1844(9):1437–52. doi:10.1016/j.bbapap.2014.05.002
- Kolarich D, Lepenies B, Seeberger PH. Glycomics, glycoproteomics and the immune system. *Curr Opin Chem Biol* (2012) 16(1–2):214–20. doi:10.1016/j.cbpa.2011.12.006

33. Balch WE, Dunphy WG, Braell WA, Rothman JE. Reconstitution of the transport of protein between successive compartments of the Golgi measured by the coupled incorporation of N-acetylglucosamine. *Cell* (1984) **39**(2 Pt 1):405–16. doi:10.1016/0092-8674(84)90019-9
34. Packer NH, Lawson MA, Jardine DR, Redmond JW. A general approach to desalting oligosaccharides released from glycoproteins. *Glycoconj J* (1998) **15**(8):737–47. doi:10.1023/A:1006983125913
35. Leymarie N, Griffin PJ, Jonscher K, Kolarich D, Orlando R, McComb M, et al. Interlaboratory study on differential analysis of protein glycosylation by mass spectrometry: the ABRF glycoprotein research multi-institutional study 2012. *Mol Cell Proteomics* (2013) **12**(10):2935–51. doi:10.1074/mcp.M113.030643
36. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* (2011) **8**(10):785–6. doi:10.1038/nmeth.1701
37. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* (2001) **305**(3):567–80. doi:10.1006/jmbi.2000.4315
38. Hiller K, Grote A, Scheer M, Münch R, Jahn D. PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res* (2004) **32**:W375–9. doi:10.1093/nar/gkh378
39. Kall L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* (2004) **338**(5):1027–36. doi:10.1016/j.jmb.2004.03.016
40. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* (2002):310–22.
41. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J* (1986) **5**(4):823–6.
42. Hubbard SJ, Thornton JM. NACCESS Computer Program. Department of Biochemistry and Molecular Biology, University College London (1993). Available from: <http://wolf.bms.umist.ac.uk/naccess/>
43. Engelen S, Trojan LA, Sacquin-Mora S, Lavery R, Carbone A. Joint evolutionary trees: a large-scale method to predict protein interfaces based on sequence sampling. *PLoS Comput Biol* (2009) **5**(1):e1000267. doi:10.1371/journal.pcbi.1000267
44. Tuncbag N, Gursoy A, Nussinov R, Keskin O. Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat Protoc* (2011) **6**(9):1341–54. doi:10.1038/nprot.2011.367
45. Joseph AP, Valadié H, Srinivasan N, de Brevern AG. Local structural differences in homologous proteins: specificities in different SCOP classes. *PLoS One* (2012) **7**(6):e38805. doi:10.1371/journal.pone.0038805
46. Chothia C. The nature of the accessible and buried surfaces in proteins. *J Mol Biol* (1976) **105**(1):1–12. doi:10.1016/0022-2836(76)90191-1
47. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* (1971) **55**(3):379–400. doi:10.1016/0022-2836(71)90324-X
48. An HJ, Gip P, Kim J, Wu S, Park KW, McVaugh CT, et al. Extensive determination of glycan heterogeneity reveals an unusual abundance of high mannose glycans in enriched plasma membranes of human embryonic stem cells. *Mol Cell Proteomics* (2012) **11**(4):M111010660. doi:10.1074/mcp.M111.010660
49. Hua S, Saunders M, Dimapasoc LM, Jeong SH, Kim BJ, Kim S, et al. Differentiation of cancer cell origin and molecular subtype by plasma membrane N-glycan profiling. *J Proteome Res* (2013) **13**(2):961–8. doi:10.1021/pr400987f
50. Liu X, Nie H, Zhang Y, Yao Y, Maitikabili A, Qu Y, et al. Cell surface-specific N-glycan profiling in breast cancer. *PLoS One* (2013) **8**(8):e72704. doi:10.1371/journal.pone.0072704
51. de Leoz ML, Young LJ, An HJ, Kronewitter SR, Kim J, Miyamoto S, et al. High-mannose glycans are elevated during breast cancer progression. *Mol Cell Proteomics* (2011) **10**(1):M110.002717. doi:10.1074/mcp.M110.002717
52. Sethi MK, Thaysen-Andersen M, Smith JT, Baker MS, Packer NH, Hancock WS, et al. Comparative N-glycan profiling of colorectal cancer cell lines reveals unique bisecting GlcNAc and alpha-2,3-linked sialic acid determinants are associated with membrane proteins of the more metastatic/aggressive cell lines. *J Proteome Res* (2014) **13**(1):277–88. doi:10.1021/pr400861m
53. Frank M, Schloissnig S. Bioinformatics and molecular modeling in glycobiology. *Cell Mol Life Sci* (2010) **67**(16):2749–72. doi:10.1007/s00018-010-0352-4
54. Lutteke T. Analysis and validation of carbohydrate three-dimensional structures. *Acta Crystallogr D Biol Crystallogr* (2009) **65**(Pt 2):156–68. doi:10.1107/S0907444909001905
55. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* (2004) **14**(2):103–14. doi:10.1093/glycob/cwh008
56. Petrescu AJ, Wormald MR, Dwek RA. Structural aspects of glycomes with a focus on N-glycosylation and glycoprotein folding. *Curr Opin Struct Biol* (2006) **16**(5):600–7. doi:10.1016/j.sbi.2006.08.007
57. Sharon N. Carbohydrate – lectin interactions in infectious disease. In: Kahane I, Ofek I, editors. *Toward Anti-Adhesion Therapy for Microbial Diseases*. Springer (1996). p. 1–8. doi:10.1007/978-1-4613-0415-9\_1
58. Peterson R, Cheah WY, Grinyer J, Packer N. Glycoconjugates in human milk: protecting infants from disease. *Glycobiology* (2013) **23**(12):1425–38. doi:10.1093/glycob/cwt072
59. Crocker PR, Paulson JC, Varki A. Siglecs and their roles in the immune system. *Nat Rev Immunol* (2007) **7**(4):255–66. doi:10.1038/nri2056
60. Varki A. Nothing in glycobiology makes sense, except in the light of evolution. *Cell* (2006) **126**(5):841–5. doi:10.1016/j.cell.2006.08.022
61. Johns TG, Mellman I, Cartwright GA, Ritter G, Old LJ, Burgess AW, et al. The antitumor monoclonal antibody 806 recognizes a high-mannose form of the EGF receptor that reaches the cell surface when cells over-express the receptor. *FASEB J* (2005) **19**(7):780–2. doi:10.1096/fj.04-1766fje
62. Scott DW, Dunn TS, Ballesta ME, Litovsky SH, Patel RP. Identification of a high-mannose ICAM-1 glycoform: effects of ICAM-1 hypoglycosylation on monocyte adhesion and outside in signaling. *Am J Physiol Cell Physiol* (2013) **305**(2):C228–37. doi:10.1152/ajpcell.00116.2013
63. Ogden CA, deCathelineau A, Hoffmann PR, Bratton D, Ghebrehiwet B, Fadok VA, et al. C1q and mannose binding lectin engagement of cell surface calreticulin and CD91 initiates macropinocytosis and uptake of apoptotic cells. *J Exp Med* (2001) **194**(6):781–95. doi:10.1084/jem.194.6.781
64. Takahashi K, Ezekowitz RA. The role of the mannose-binding lectin in innate immunity. *Clin Infect Dis* (2005) **41**(Suppl 7):S440–4. doi:10.1086/431987

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 July 2014; accepted: 07 August 2014; published online: 25 August 2014.

Citation: Lee LY, Lin C-H, Fanayan S, Packer NH and Thaysen-Andersen M (2014) Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants. *Front. Immunol.* 5:404. doi: 10.3389/fimmu.2014.00404  
This article was submitted to *Immunotherapies and Vaccines*, a section of the journal *Frontiers in Immunology*.

Copyright © 2014 Lee, Lin, Fanayan, Packer and Thaysen-Andersen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Crossroads between bacterial and mammalian glycosyltransferases

Inka Brockhausen<sup>1,2</sup> \*

<sup>1</sup> Department of Medicine, Queen's University, Kingston, ON, Canada

<sup>2</sup> Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON, Canada

## Edited by:

Elizabeth Yuriev, Monash University, Australia

## Reviewed by:

Thomas Dandekar, University of Würzburg, Germany

Anne Harduin-Lepers, Centre National de la Recherche Scientifique, France

## \*Correspondence:

Inka Brockhausen, Department of Biomedical and Molecular Sciences, Queen's University, 18 Stuart Street, Kingston, ON K7L3N6, Canada  
e-mail: brockhau@queensu.ca

Bacterial glycosyltransferases (GT) often synthesize the same glycan linkages as mammalian GT; yet, they usually have very little sequence identity. Nevertheless, enzymatic properties, folding, substrate specificities, and catalytic mechanisms of these enzyme proteins may have significant similarity. Thus, bacterial GT can be utilized for the enzymatic synthesis of both bacterial and mammalian types of complex glycan structures. A comparison is made here between mammalian and bacterial enzymes that synthesize epitopes found in mammalian glycoproteins, and those found in the O antigens of Gram-negative bacteria. These epitopes include Thomsen–Friedenreich (TF or T) antigen, blood group O, A, and B, type 1 and 2 chains, Lewis antigens, sialylated and fucosylated structures, and polysialic acids. Many different approaches can be taken to investigate the substrate binding and catalytic mechanisms of GT, including crystal structure analyses, mutations, comparison of amino acid sequences, NMR, and mass spectrometry. Knowledge of the protein structures and functions helps to design GT for specific glycan synthesis and to develop inhibitors. The goals are to develop new strategies to reduce bacterial virulence and to synthesize vaccines and other biologically active glycan structures.

**Keywords: glycosyltransferases, protein structure, specificities, glycoprotein epitopes, glycan mimics**

## INTRODUCTION

Glycans play important roles in most biological processes in health and disease. Bacteria and human beings have a close relationship in the intestine, which can be symbiotic or pathogenic. Bacteria often produce human-like glycan structures with bacteria-specific glycosyltransferases (GT)s that have given them a selective advantage for adhesion, colonization, and survival. Knowledge of these enzymes can help us understand human counterparts of GTs, and provide a convenient technology to synthesize both bacterial and human glycans. Bacterial GTs can be easily expressed and stored; they are more soluble and often remarkably active and stable.

Currently, GTs from many different organisms have been classified into 96 GT families in the Carbohydrate-active Enzymes (CAZy) classification system (<http://www.cazy.org>), based on their sequence similarity derived from GenBank (<ftp://ftp.ncbi.nih.gov/genbank/> or EMBL or DDBJ) (1, 2). Very few of the bacterial GTs have been biochemically and functionally characterized, thus proposed enzymes are assigned based on similarity searches. The CAZy database also contains genetic, structural, mechanistic, and functional information of known GTs. The former *Escherichia coli* (EC) nomenclature for GTs as well as the currently accepted nomenclature and alternative names for GTs are included. A number of databases provide sequence analyses of GTs (e.g., NCBI BLAST, PFAM, INTERPRO, DBCAN, Swiss-Prot – ExPASy).

For searches of glycan structures, a number of databases are useful (3). For example, GLYCOsuiteDB contains information on N- and O-linked glycans and glycoproteins and Glycobase on N- and O-Glycan structures. For glycomics analyses by

mass spectrometry (MS), GlycoMaster DB at <http://www-novo.cs.uwaterloo.ca:8080/GlycoMasterDB> is helpful (4). The current *E. coli* O-antigen database (ECODAB) contains known O antigen structures of *E. coli*, the analytic data available, and has links to genes involved in O antigen synthesis from the O antigen gene cluster (5). Many of the *E. coli* antigens can be found in other bacterial strains. Finally, the Consortium for Functional Glycomics (<http://www.functionalglycomics.org/>) provides a large database for glycan functions.

Because of wide-spread development of antibiotic resistance, we need new anti-bacterial strategies, and bacterial GTs are virulence factors that could be targeted. The understanding of GTs can help in the production of vaccines to protect against bacterial infections, cancer, and for application in inflammation and autoimmune disease. In this review, we will compare mammalian and bacterial GTs that show remarkable similarity of action, protein folding, or mechanisms, in spite of surprisingly large differences in amino acid sequences.

## MAMMALIAN GLYCOPROTEINS AND BACTERIAL GLYCANS

Mammalian glycoproteins are involved in virtually all cellular activities; they serve as ligands for antibodies or lectins, or as receptors involved in signaling, cellular interactions, cell growth, differentiation, and cell death (6–11). Glycans are important in the inflammatory response, the innate and adaptive immune system, and cancer metastasis, as well as microbial colonization and infections. Glycoproteins have many functional epitopes attached to either N-glycans or O-glycans, and the amounts of many of these epitopes can be altered in disease, for example, in cancer. Although

there is remarkable diversity in glycan structures in mammals, and hundreds of different chains can be found in glycoproteins, only six sugar residues (Man, GlcNAc, GalNAc, Gal, sialic acid, Fuc) are forming the extended and branched varieties of glycans with few modifications such as O-acetylation and sulfation. N- and O-glycans can affect the chemical and physical properties and the conformations of proteins and the accessibility of peptide epitopes.

Bacteria display an astounding variety of unusual sugars and sugar linkages as well as modifications of sugars that are foreign to human beings and, therefore, can trigger immune responses. However, a number of specific bacterial glycans are mimics of mammalian glycoprotein epitopes (Table 1). Partial structures of O-antigenic polysaccharides of Gram-negative bacteria (ECODAB) often mimic human glycans and may help bacteria to evade the immune system and promote colonization. The mimicry may prevent the production of effective vaccines to protect against bacterial infections, which requires new considerations of anti-bacterial strategies. About half of the *EC* strains have some form of mammalian epitope within their O antigens. This includes Gal $\beta$ 1-3GlcNAc $\beta$ -, and Gal $\beta$ 1-4GlcNAc $\beta$ -linkages, which are part of the glycan backbone structures (type 1 and type 2, respectively) in mammalian glycoproteins. In bacteria, those are internal structures within the O antigen repeating unit. The cancer-associated Thomsen–Friedenreich (TF or T antigen, O-glycan core 1) is common in glycoproteins and also found in several O antigens of *E. coli*. Blood group O, A, and B, sialylated glycans, and polysialic acid are mimics found in a number of bacterial strains. The fact that bacteria are able to synthesize these human-like structures suggests that they have the appropriate biosynthetic enzymes (Table 2), although this would be difficult to anticipate from the inspection of the amino acid sequences of their GTs. Biochemical characterization of bacterial enzymes and structure/function studies are important prerequisites to utilize these enzymes in chemoenzymatic synthesis of mammalian glycoprotein epitope structures.

### ROLE OF O ANTIGENS

The LPS of Gram-negative bacteria are essential structures of the outer membrane. LPS binds to the LPS-binding protein, requiring the CD14/TLR4/MD2 receptor complex, which elicits a strong response during infections, through TLR4 signaling. LPS consist of a lipid A base (endotoxin), which carries a relatively invariable inner oligosaccharide core, strain-specific outer core oligosaccharides, and the serotype-specific outer O antigen polysaccharide. O antigens are polysaccharides composed of up to 50 repeating units of oligosaccharides with one (homopolymeric) to 10 sugars (heteropolymeric) that play a role in bacterial adhesion and colonization, affect pathogenicity and survival, and can be bacteriophage receptors. The enormous structural variability of O antigens is mediated by many specific GTs and other enzymes that modify O antigens, thus increasing structural diversity, e.g., by adding phosphate, acetyl groups, or branching sugar residues. The LPS molecules are necessary for stabilization of the outer membrane and form a barrier against penetration of toxins. In particular, the O antigens serve to evade complement; they protect against phagocytosis and give the bacteria a strain-specific and diversity-selective advantage. The molecular mimicry found

**Table 1 | Glycan mimics: examples of mammalian glycoprotein epitopes also found in the lipopolysaccharides (LPS) or lipooligosaccharides of Gram-negative bacteria.**

Glycoprotein epitope	Structure	Bacteria
T antigen	Gal $\beta$ 1-3GalNAc $\alpha$ -	<i>EC</i> O104, <i>EC</i> O5, <i>EC</i> O127
Sialyl-T antigen	Sialyl $\alpha$ 2-3Gal $\beta$ 1-3GalNAc $\alpha$ -	<i>EC</i> O104
Type 1 chain	Gal $\beta$ 1-3GlcNAc-	<i>EC</i> O7, <i>EC</i> O55
Type 2 chain	Gal $\beta$ 1-4GlcNAc-	<i>EC</i> O91
Lewis a	Fuc $\alpha$ 1-4[Gal $\beta$ 1-3]GlcNAc $\beta$ -	<i>Hp</i>
Lewis b	Fuc $\alpha$ 1-4[Fuc $\alpha$ 1-2Gal $\beta$ 1-3]GlcNAc $\beta$ -	<i>Hp</i>
Lewis x	Gal $\beta$ 1-4[Fuc $\alpha$ 1-3]GlcNAc $\beta$ -	<i>Hp</i>
Lewis y	Fuc $\alpha$ 1-2Gal $\beta$ 1-4[Fuc $\alpha$ 1-3]GlcNAc $\beta$ -	<i>Hp</i>
H antigen	Fuc $\alpha$ 1-2Gal $\beta$ -	<i>EC</i> O86
Blood group A	GalNAc $\alpha$ 1-3[Fuc $\alpha$ 1-2]Gal $\beta$ -	<i>Bo</i>
Blood group B	Gal $\alpha$ 1-3[Fuc $\alpha$ 1-2]Gal $\beta$ -	<i>EC</i> O86
Linear B	Gal $\alpha$ 1-3Gal $\beta$ -	<i>Nm</i>
P blood group	Gal $\alpha$ 1-4Gal $\beta$ -	<i>Cj</i> , <i>Nm</i>
Polysialic acids, PSA	[Sialyl $\alpha$ 2-8] <sub>n</sub>	<i>EC</i> , <i>Nm</i>
Sialyl $\alpha$ 2-3	Sialyl $\alpha$ 2-3Gal $\beta$ -	<i>Cj</i> , <i>EC</i> O104
Sialyl $\alpha$ 2-6	Sialyl $\alpha$ 2-6Gal $\beta$ -	<i>Psp</i>
Fuc $\alpha$ 1-6	Fuc $\alpha$ 1-6GlcNAc $\beta$ -	<i>Rsp</i>

*Bo*, *Bacteroides ovatus*; *Cj*, *Campylobacter jejuni*; *EC*, *Escherichia coli*; *Hp*, *Helicobacter pylori*; *Nm*, *Neisseria meningitidis*; *Psp*, *Photobacterium sp.*; *Rsp*, *Rhizobium sp.*

in a large proportion of bacteria adds to their ability to prevent recognition by the host immune system and thus promotes virulence.

A number of bacteria do not have an extended O-antigenic polysaccharide but instead have a short lipooligosaccharide that may have structural identity with human glycoproteins or glycolipids and may lead to pathological conditions. The close relationship between bacteria and human beings is also apparent in the abundance of bacterial lectins that bind to mammalian glycoproteins and thus promote adhesion to mammalian tissues.

### N-GLYCOSYLATION OF MAMMALIAN AND BACTERIAL GLYCOPROTEINS

In eukaryotic cells, N-glycans are assembled first on a dolichol-phosphate (P-Dol) intermediate on the cytoplasmic side of the endoplasmic reticulum (ER) membrane (7, 12) (Figure 1). GlcNAc-phosphate is transferred by GlcNAc-phosphotransferase in a reversible reaction, inhibited by tunicamycin, to P-Dol, followed by the transfer of another GlcNAc residue to form chitobiose. This is followed by five Man residues, all transferred from nucleotide sugar donor substrates to form the common N-glycan core structure Man<sub>5</sub>-chitobiose linked to PP-Dol. This heptasaccharide is flipped across the membrane and further addition of sugars occurs on the inside of the ER lumen through transfer from Man-P-Dol and Glc-P-Dol donor substrates. After completion of the lipid-linked N-glycan, it is transferred *en bloc* to the Asn residue(s) of Asn-X-Ser/Thr sequons in a glycoprotein by the oligosaccharyltransferase complex (OST), and Glc and Man

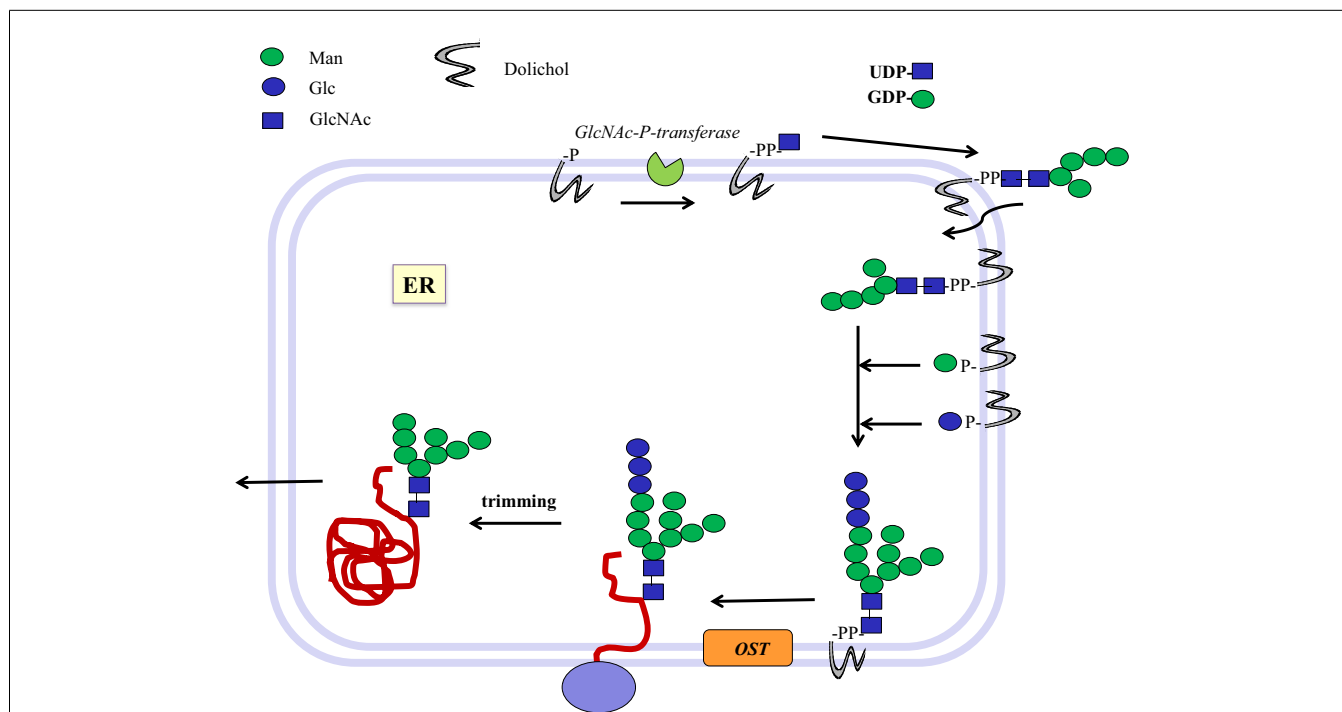
**Table 2 | Comparison of characterized mammalian and bacterial enzymes that catalyze a similar reaction.**

Human enzyme (Accession No.)	Acceptor	Bacterial enzyme (Species) (Accession No.)	Acceptor	Identity
Core 1 $\beta$ 3-Gal-transferase, C1GALT1, T synthase <b>(NP_064541, Q9NS00)</b>	GalNAc $\alpha$ -	Wbwc (ECO104) <b>(AAK64375, Q93NP5)</b>	GalNAc $\alpha$ -PP-lipid	<b>10.5</b>
		WbiP (ECO127) <b>(YP_002329684, B7UT68)</b>	GalNAc $\alpha$ -R	<b>10.6</b>
		CgtB ( <i>Cj</i> ) <b>(ABL75151, A2I3T3)</b>	GalNAc $\beta$ -R	<b>8.4</b>
$\beta$ 4-Gal-transferase 1, B4GALT1 <b>(P15291)</b>	GlcNAc $\beta$ -	LgtB ( <i>Nm</i> ) <b>(EJU76624, J8YER7)</b>	GlcNAc $\beta$ -R	<b>9.3</b>
		WfeD ( <i>Sb</i> ) <b>(ACD37055, B5L3X1)</b>	GlcNAc $\alpha$ -PP-R	<b>8.6</b>
$\beta$ 3-Gal-transferase 5, B3GALT5 <b>(NP_149361, Q9Y2C3)</b>	GlcNAc $\beta$ -	WbgO (ECO55) <b>(ADD57106, D3QY14)</b>	GlcNAc $\beta$ -R	<b>11.4</b>
		WbbD (ECO7) <b>(AAC27537, Q03084)</b>	GlcNAc $\alpha$ -PP-lipid	<b>13</b>
$\beta$ 3-GlcNAc-transferase 2, B3GNT2 <b>(AAD09764, Q9NY97)</b>	Gal $\beta$ -	LgtA ( <i>Nm</i> ) <b>(AAL12840, Q93EK6)</b>	lactose	<b>11.6</b>
$\alpha$ 3-Gal-transferase, ABO transferase GTB <b>(AAD26574, Q9UQ64)</b>	Fuc $\alpha$ 2Gal $\beta$ -	WbnI (ECO86) <b>(AAV80756, Q5JBG6)</b>	Fuc $\alpha$ 2Gal $\beta$ -R	<b>20.6</b>
$\alpha$ 4-Gal-transferase, A4GALT <b>(AAO39150, Q540I6)</b>	Gal $\beta$ -	CgtD ( <i>Cj</i> ) <b>(AAM90647, Q8KWQ9)</b>	Gal $\beta$ -R	<b>11.6</b>
		LgtC ( <i>Nm</i> ) <b>(AAL12839, Q93EK7)</b>	Gal $\beta$ -R	<b>10.7</b>
$\alpha$ 3-GalNAc-transferase, ABO transferase GTA <b>(V5ZDN8, P16442)</b>	Fuc $\alpha$ 2Gal $\beta$ -	BoGT56a ( <i>Bo</i> ) <b>(4AYL_A, A7LVT2)</b>	Fuc $\alpha$ 2Gal $\beta$ -R	<b>21.1</b>
		GTA-like ( <i>Hm</i> ) <b>(CBG40459, D3UIY4)</b>	Fuc $\alpha$ 2Gal $\beta$ -R	<b>19.7</b>
$\alpha$ 3-Sialyl-transferase ST3GAL1 <b>(NP_775479, Q11201)</b>	Gal $\beta$ 3GalNAc-	NST ( <i>Nm</i> ) <b>(C6S6J1)</b>	Lactose $\beta$ -benzyl	<b>13.2</b>
		WbwA (ECO104) <b>(AAK64371, Q93NP9)</b>	Gal $\beta$ 3GalNAc $\alpha$ -PP-lipid	<b>13.6</b>
$\alpha$ 3-Sialyl-transferase, ST3GAL4 <b>(Q11206)</b>	Gal $\beta$ 3-	CstI ( <i>Cj</i> ) <b>(ADA70348, D2KQ02)</b>	Gal $\beta$ 3-R	<b>10.2</b>
$\alpha$ 6-Sialyl-transferase, ST6GAL1 <b>(P15907)</b>	Gal $\beta$ 3-	Pm0188 ( <i>Pm</i> ) <b>(Q9CP67)</b>	Gal $\beta$ -R	<b>14.6</b>
		$\alpha$ 6Sialyltransferase ( <i>Psp</i> ) <b>(A8QYL1)</b>	Gal $\beta$ -R	<b>14.6</b>
Poly $\alpha$ 8-sialyl-transferase 2, ST8SIA2 <b>(AAH10645, Q92186)</b>	sialyl-Gal-	PST ( <i>Nm</i> ) <b>(AAS90326, Q6PUE5)</b>	GD1 ganglioside	<b>9.8</b>
		PST ( <i>EC</i> ) <b>(B1LEB7)</b>	Sialyl-R	<b>5.4</b>
$\alpha$ 2-Fuc-transferase, FUT2 <b>(CAQ81982, A8K2L2)</b>	Gal $\beta$ -	WbwK (ECO86) <b>(AAO37719, Q5JBG3)</b>	Gal $\beta$ 3GalNAc-	<b>11.9</b>
		WbsJ (ECO128) <b>(AAO37698, Q6XQ53)</b>	Gal $\beta$ 4-R	<b>15.8</b>
		FutC ( <i>Hp</i> ) <b>(ABO61750, A4L7J1)</b>	Gal $\beta$ -	<b>17.5</b>
$\alpha$ 3-Fuc-transferase, FUT4 <b>(NP_002024, P22083)</b>	Gal $\beta$ 4GlcNAc $\beta$ -	FutA ( <i>Hp</i> ) <b>(YP_003057735, C7BXF2)</b>	Gal $\beta$ 4GlcNAc $\beta$ -R	<b>10</b>
		FutB ( <i>Hp</i> ) <b>(YP_003057467, C7BZU7)</b>	Gal $\beta$ 4GlcNAc $\beta$ -R	<b>10.2</b>
		$\alpha$ 3FUT ( <i>Hh</i> ) <b>(Q7VFA1)</b>	(Sialyl-)Gal $\beta$ 4GlcNAc $\beta$ -R	<b>10.4</b>
$\alpha$ 3/4-Fuc-transferase, FUT3 <b>(P21217)</b>	Gal $\beta$ 4/3GlcNAc $\beta$ -	FucTa ( <i>Hp</i> ) <b>(AAF35291, Q9L8S4)</b>	Gal $\beta$ 4/3GlcNAc $\beta$ -R	<b>14.5</b>
$\alpha$ 6-Fuc-transferase, FUT8 <b>(AAI42959, A5PLL2)</b>	N-glycan core	NodZ ( <i>Rsp</i> ) <b>(G9ACH1)</b>	Chitobiose	<b>7.6</b>

The donor substrates of mammalian and bacterial enzymes that synthesize the same linkage are the same but the preferred acceptor substrate may differ. The bacterial enzyme can mimic the mammalian enzyme and can be used to synthesize mammalian glycan structures but the amino acid sequences have little sequence identities. The percentage of amino acid identity is much higher among eukaryotic GT families and between GTs with similar action and using similar acceptors in bacteria. The accession numbers in bold (Accession No.) are from Pubmed and/or UniProt database. *Bo*, *Bacteroides ovatus*; *Cj*, *Campylobacter jejuni*; *EC*, *E. coli*; *Hh*, *Helicobacter hepaticus*; *Hm*, *Helicobacter mustelae*; *Hp*, *Helicobacter pylori*; *Nm*, *Neisseria meningitidis*; *Pm*, *Pasteurella multocida*; *Psp*, *Photobacterium sp.*; *Rsp*, *Rhizobium sp.*; *Sb*, *Shigella boydii*; *PP*, diphosphate.

residues are selectively cleaved by glycosidases. After transfer to the Golgi, further removal of Man residues occurs, and GlcNAc-transferase I (GnT I, MGAT1) adds the first of the N-glycan antennae in  $\beta$ 1–2 linkage to the Man $\alpha$ 1–3 arm of the core. This can then be followed by several steps that depend on the presence of this first GlcNAc antenna and the expression of processing enzymes, which remove two Man residues from the Man $\alpha$ 1–6 arm, add

Fuc $\alpha$ 1–6 to the inner chitobiose GlcNAc, and add further antennae to form complex type N-glycans. N-glycans can be extended by repeating Gal-GlcNAc residues to form type 1 or type 2 chains (Table 1); they may be branched by GlcNAc $\beta$ 1–6 linkages and may be decorated with specific functional epitopes and blood group determinants (7). This results in hundreds of different N-glycan structures, depending on the glycosylation potential of the cell.



**FIGURE 1 | Biosynthesis of N-glycosylated glycoproteins in eukaryotes.**

N-glycosylation is initiated at the endoplasmic reticulum (ER) membrane using nucleotide sugar donor substrates and a membrane-bound acceptor phospholipid with multiple isoprenyl units (dolichol-phosphate, P-Dol). The first sugar (GlcNAc) is transferred as GlcNAc-phosphate from UDP-GlcNAc by GlcNAc-P-transferase, resulting in GlcNAc-diphosphate-dolichol (GlcNAc-PP-Dol). This step can be inhibited by the UDP-GlcNAc analog tunicamycin. On the outside face of the ER membrane, another GlcNAc is added to form chitobiose, followed by five Man residues to form a heptasaccharide (Man<sub>5</sub>GlcNAc<sub>2</sub>)-PP-Dol. This heptasaccharide is flipped to the inside of the ER where the chain grows by transfer of sugars from membrane-bound Man-P-Dol and Glc-P-Dol. The completed saccharide Glc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub> is then transferred by an oligosaccharyltransferase complex (OST) to the Asn residue in an Asn-x-Ser/Thr sequon of nascent proteins. After trimming of sugar residues in the ER by removal of Glc and Man residues to

the Man<sub>5</sub>GlcNAc<sub>2</sub> structure, glycoproteins are exported to the Golgi where further trimming occurs by mannosidases. Many N-glycan chains are processed to the complex type by the addition of GlcNAc residues by GlcNAc-transferases I to V (MGAT1 to 5). Chains grow further by the addition of Gal-GlcNAc sequences and termination by sialyl-, Fuc-, Gal-, GlcNAc-, and GalNAc-transferases, which are all highly specific for both the donor and the acceptor substrates and with few exceptions form only one type of linkage between sugars. This creates a multitude of hundreds of different structures and epitopes with many possible functions, depending on the final destination of the glycoprotein, e.g., in the cell membrane or in secretions. Glycoprotein biosynthesis is regulated at many different levels, e.g., by the synthesis and delivery of nucleotide sugar substrates, the expression, activities and localization of glycosyltransferases and trimming hydrolases, the competition of enzymes for common substrates, levels of metal ion activating factors, localization of enzymes involved, and rate of transport of glycoproteins.

Not all N-glycosylation sites carry N-glycans, and there are differences in chain processing between different glycosylation sites of the same protein. The peptide has been shown to interact with the glycan chains, and this controls the conformations of the glycan and the peptide and leads to site-specific glycosylation. Many sequentially acting and competing GTs assemble glycoproteins in a cell type-specific pattern. Most of the GTs involved exist as families of enzymes (Table 3). Several of these have been shown to be localized in specific Golgi compartments according to their action within the complex pathways.

In the mammalian biosynthetic pathways, the sequence of sugar additions is controlled by the gene expression, the relative activities of competing enzymes, the enzyme localizations, levels of substrates and cofactors, and the distinct substrate specificities of GTs. These types of controls still need to be investigated for glycosylation reactions in bacteria.

Bacteria such as *Campylobacter jejuni* (Cj) also have N-glycosylated proteins (13). An oligosaccharide is first assembled

on undecaprenol-phosphate (P-Und), an analog of P-Dol, in the cytoplasmic compartment. The sugar-PP-Und is then flipped to the periplasmic space where the glycan chain is transferred *en bloc* to protein by oligosaccharyltransferases. These GTs have a broad specificity toward their donor substrates but also require a sequon, Asp/Glu-x-Asn-y-Ser/Thr, where x and y cannot be Pro, in the protein acceptor, that bears close resemblance to the mammalian N-glycosylation sequon (Figure 1).

#### PROTEIN O-GLYCOSYLATION

O-glycans of glycoproteins and mucins are assembled in mammals without a lipid intermediate and without removal of sugar residues by glycosidases (14, 15). The first sugar is always GalNAc $\alpha$ -linked to Ser or Thr (the cancer-associated Tn antigen). All sugars are transferred from nucleotide sugars in the Golgi, resulting in extended and branched O-glycans with hundreds of different structures. The most common structure is Gal $\beta$ 1-3GalNAc, core 1, the T antigen, which is normally masked by the addition of other

**Table 3 | Examples of glycosyltransferase families (CAZy).**

GT family	Mechanism	Fold	Glycosyltransferases	Known structures
2	I	GT-A	WbbD, WbwC, WbiP, CgtB, WbgO, WfaP, WfgD, LgtA	7
6	R	GT-A	GTA, GTB, A3GALT, Wbni, BoGT56a, GTA-like ( <i>Hm</i> )	4
7	I	GT-A	B4GALT	4
8	R	GT-A	LgtC	5
10	I	GT-B	$\alpha$ 3/4FUT, FutA, FutB, FucTa, FUT3, FUT4-7, 9-11	1
11	I	GT-B	$\alpha$ 2FUT, $\alpha$ 3FUT ( <i>Hh</i> ), WbwK, Wbsj, FutC	–
13	I	GT-A	GnT I, MGAT1	1
14	I	GT-A	C2GnT1, IGnT, GCNT2	1
16	I	?	MGAT2	–
17	I	?	MGAT3	–
18	I	?	MGAT5	–
23	I	GT-B	$\alpha$ 6FUT8, NodZ	2
25	I	?	LgtB	–
26	I	?	WfeD	–
27	R	GT-A	GALNT	3
29	I	GT-A	Animal sialyl-T, ST3GAL, ST6GAL, ST6GALNAC, ST8SIA	3
31	I	GT-A	C1GALT, B3GALT5	1
32	R	?	CgtD, Lgt5, A4GALT	–
38	I	?	Bacterial PST	–
42	I	GT-A	Bacterial Sialyl-T, CstI, CstII	2
52	I	GT-B	Bacterial Sialyl-T, NST, WbwA	1
54	I	?	MGAT4	–
80	I	GT-B	Bacterial Sialyl-T, Pm0188, $\alpha$ 6Sialyl-T ( <i>Psp</i> ), PmST1, multifunctional	4

Fold, expected overall fold; Hh, *Helicobacter hepaticus*; Hm, *Helicobacter mustelae*; I, Inverting mechanism; Psp, *Photobacterium sp.*; R, Retaining mechanism; -T, -transferase.

residues but exposed in many cancer cells. In a number of cells, core 1 is branched by core 2  $\beta$ 6-GlcNAc-transferase (C2GnT) or extended in a fashion that is similar to the synthesis of complex N-glycans.

GalNAc is transferred from UDP-GalNAc by up to 20 polypeptide GalNAc-transferases (GALNTs) (14–16). All GALNTs are

classified in the GT27 family with a GT-A fold (Table 3). They have a catalytic domain linked to a lectin (ricin-like) domain at the C terminus. This lectin domain has three subdomains and may play an important role in binding products or substrates containing GalNAc residues. A crystal structure of mouse GALNT1 with Mn<sup>2+</sup> supported the importance of a DxH motif and the role of Asp209, His211, and His344 (17) (Table 4). The conformations of human GALNT2 (18) crystallized with UDP and with or without an acceptor peptide showed a loop formed over UDP. It appeared that the acceptor peptide connected the otherwise separate catalytic and lectin domains. Kinetic analyses showed that the presence of GalNAc in the acceptor was beneficial for activity. Human GALNT10 was crystallized complexed with UDP, GalNAc, and Mn<sup>2+</sup> (19). GalNAc-peptides appear to bind to the second beta-subdomain of the lectin domain. Binding of the donor induces a conformation change that opens the acceptor-binding site. These three crystallized ppGalNAcTs are similar in overall structure and mechanism.

An equivalent GALNT that transfers GalNAc to protein has not been identified in bacteria, although bacteria are known to O-glycosylate Ser/Thr residues of proteins with various sugar residues. In contrast to mammalian O-glycosylation, bacteria transfer a pre-assembled oligosaccharide to Ser/Thr. Bacterial protein OGTs have no sequence homology to GALNT and their action is reminiscent to that of OST in the N-glycosylation pathway. In several bacteria, for example in *Campylobacter* and *Neisseria*, an oligosaccharide or monosaccharide is first pre-assembled on PP-lipid in the cytoplasmic compartment, flipped to the periplasm and then transferred *en bloc* to Ser/Thr residues of proteins. These enzymes have a relaxed oligosaccharide donor specificity (46). Oligosaccharyltransferase PglL (which has not yet been assigned to a GT family) from *Neisseria meningitidis* (*Nm*) can transfer many different glycans from sugar-PP-Und or sugar-PP-lipid (including sugar-PP-Dol) donor substrates to protein in the periplasmic space. UDP-*N*-diacetyl-bacillosamine was also a donor substrate *in vitro*, showing that even nucleotide sugars can be donors and a single sugar could be transferred to protein. Mutagenesis experiments showed that PglL from *Nm* requires His349 for activity and for interaction with the lipid-linked oligosaccharide (47).

**BIOSYNTHESIS OF BACTERIAL O ANTIGENS**

There are many similarities in the pathways and mechanisms by which bacterial O antigens and mammalian glycoproteins are synthesized. In Gram-negative bacteria, O antigens are synthesized by specific GTs at the cytosolic face of the inner membrane where the nucleotide sugar donor substrates are present, as well as the membrane-bound P-Und, an analog of the mammalian P-Dol, as the acceptor substrate for the transfer of the first sugar (48) (Figure 2). The first GT to act is always a sugar-phosphate transferase that produces the sugar-PP-Und substrate for subsequent transfer of monosaccharides by GTs. Most *E. coli* have GlcNAc or GalNAc at the reducing end of the repeating unit, thus sugar-phosphate transferase WecA and its orthologs are responsible for the first reaction, maintaining the  $\alpha$ -anomeric configuration of GlcNAc. 4-Epimerases may also be involved in interconverting GlcNAc and GalNAc in the activated form (UDP-GlcNAc/UDP-GalNAc) or after the sugar transfer (49).



**Table 4 | Mammalian and bacterial glycosyltransferases with known crystal structure relevant to glycoprotein determinants or their mimics.**

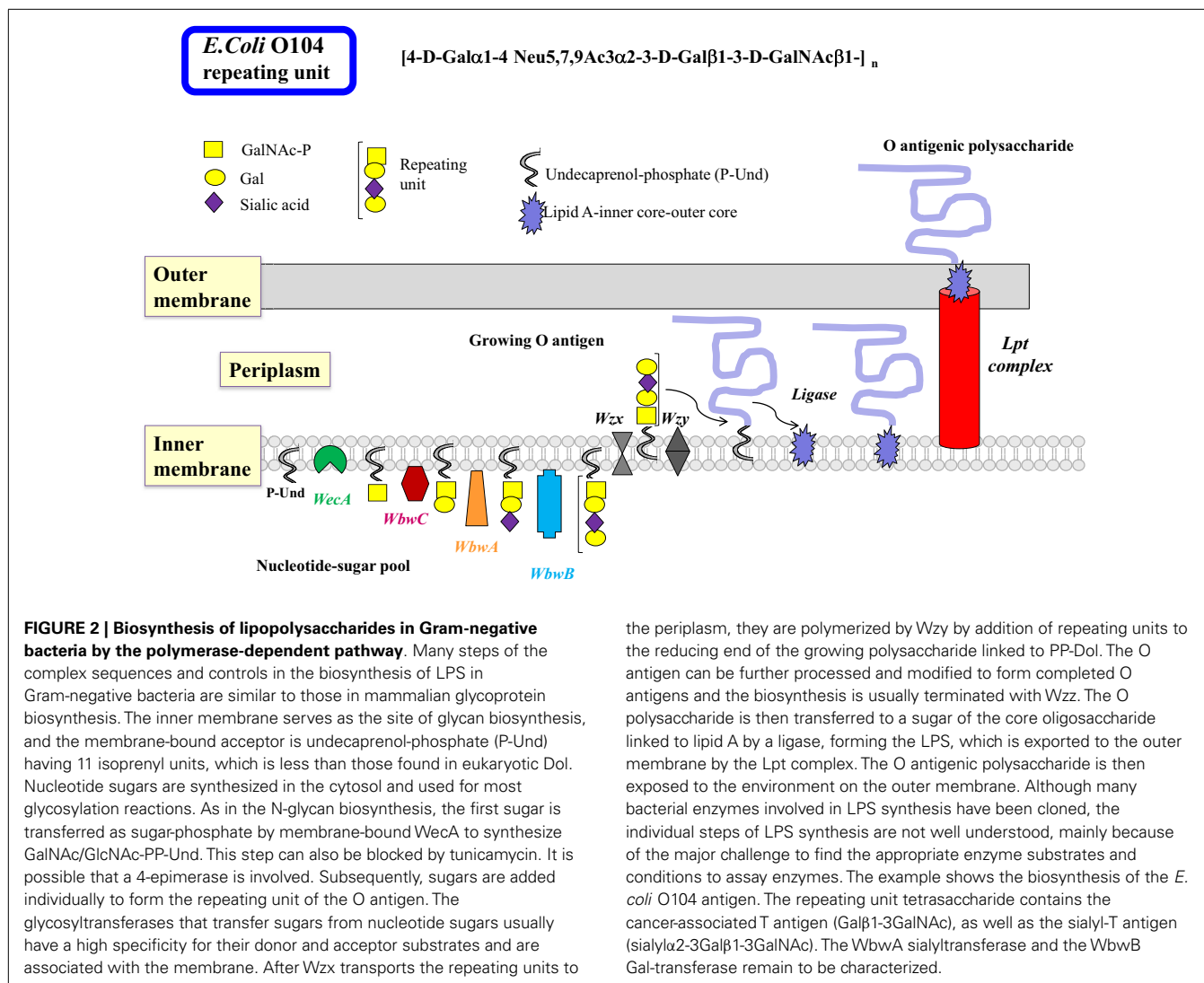
Enzyme name		Species	Donor	Acceptor	GT	Fold	DxD	Reference
<b>INVERTING</b>								
GnT I, MGAT1		Rabbit	UDP-GlcNAc	Man $\alpha$ 3-R	13	GT-A	DxD	(19)
C2GnT, GCNT1		Mouse	UDP-GlcNAc	Gal $\beta$ 3GalNAc-R	14	GT-A	E320	(20, 21)
$\beta$ 4-Gal-T1, B4GALT1		Human	UDP-Gal	GlcNAc $\beta$ -R	7	GT-A	DxD	(22)
$\beta$ 4-Gal-T1, B4GALT1		Bovine	UDP-Gal	GlcNAc $\beta$ -R	7	GT-A	DxD	(23)
$\beta$ 4-Gal-T7, B4GALT7		Human	UDP-Gal	GlcNAc $\beta$ -R	7	GT-A	DxD	(24)
$\alpha$ 3-FUT		<i>Hp</i>	GDP-Fuc	Gal $\beta$ 4GlcNAc	10	GT-B	E95, E249	(25)
$\alpha$ 6-FUT8		Human	GDP-Fuc	N-glycan core Gn	23	GT-B	D453	(26)
NodZ		<i>Rsp</i>	GDP-Fuc	chitobiose	23	GT-B		(27)
ST3GAL1		Pig	CMP-SA	Gal $\beta$ 3GalNAc-R	29	GT-A	H319	(28)
$\alpha$ 3-Sialyl-T		<i>Pp</i>	CMP-SA	Gal-R	80	GT-B	H317	(29)
$\alpha$ 3-Sialyl-T	CstI	<i>Cj</i>	CMP-SA	Gal $\beta$ -R	42	GT-A	H202	(30)
$\alpha$ 3-Sialyl-T	NST	<i>Nm</i>	CMP-SA	Lactose-Bn	52	GT-A	H280	(31)
ST6GAL1		Rat	CMP-SA	Gal $\beta$ 4GlcNAc	29	GT-A	H367	(32)
ST6GAL1		Human	CMP-SA	Gal $\beta$ 4GlcNAc	29	GT-A	H370	(33)
$\alpha$ 6-Sialyl-T	PM0188	<i>Pm</i>	CMP-SA	Lactose, Gal, GalNAc, SA	80	GT-B	D141	(34)
$\alpha$ 6-Sialyl-T	pst6-224	<i>Psp</i>	CMP-SA	Lactose	80	GT-B		(35)
$\alpha$ 3/8-Sialyl-T	CstII	<i>Cj</i>	CMP-SA	Gal $\beta$ -R	42	GT-A	H188	(36)
$\alpha$ 3-Sialyl-T multi	PmST1	<i>Pm</i>	CMP-SA	Lactose, Gal, GalNAc, sialic acid	80	GT-B	D141	(37)
<b>RETAINING</b>								
GALNT1		Mouse	UDP-GalNAc	Thr/Ser	27	GT-A-ricin		(17)
GALNT2		Human	UDP-GalNAc	Thr/Ser	27	GT-A-ricin		(18)
GALNT10		Human	UDP-GalNAc	Thr/Ser	27	GT-A-ricin		(38)
ABO transferase	GTB	Human	UDP-Gal	H antigen	6	GT-A	E303 H347	(39) (40)
$\alpha$ 3-Gal-T	Linear B Gal-T	Bovine	UDP-Gal	Gal $\beta$ 4GlcNAc	6	GT-A	E317, H271	(41, 42)
ABO transferase	GTA	Human	UDP-GalNAc	H antigen	6	GT-A	E303 H347	(39, 43)
$\alpha$ 3-GalNAc-T	BoGT6a	<i>Bo</i>	UDP-GalNAc	Fuc-lactose	6	GT-A	NxN	(44)
$\alpha$ 4-Gal-T	LgtC	<i>Nm</i>	UDP-Gal	Lactose, oligosac	8	GT-B	4DxD, H188	(45)

Enzymes are listed that have been characterized and crystallized and shown to be involved in the synthesis of glycoproteins or their glycan mimics in bacteria. *Bo*, *Bacteroides ovatus*; *Bs*, *Bacillus subtilis*; *Cj*, *Campylobacter jejuni*; DxD, presence of a DxD motif or its analog or proposed catalytic residue; GT, Glycosyltransferase family (CAZy); *Hp*, *Helicobacter pylori*; multi, multifunctional enzyme; Inverting: enzymes invert the anomeric configuration of the sugar of the donor substrate to form enzyme product; N-glycan core Gn, reducing GlcNAc residue of the N-glycan core structure; *Nm*, *Neisseria meningitidis*; *Pm*, *Pasteurella multocida*; *Pp*, *Photobacterium phosphoreum*; *Psp*, *Photobacterium sp.*; *Rsp*, *Rhizobium sp.*; Retaining: enzymes retain the anomeric configuration of the sugar of the donor substrate in the enzyme product; SA, sialic acid; -T, -transferase.

The common heteropolymeric O antigens are synthesized by sequential transfer of sugar units by donor- and acceptor-specific, membrane-associated GTs. The specificities of these bacterial GTs are distinct and comparable to eukaryotic GTs. A completed repeating unit is then translocated across the inner membrane to the periplasmic side by the multiple membrane-spanning flippase Wzx, a process resembling the transfer of Man<sub>5</sub>GlcNAc<sub>2</sub>-PP-Dol intermediate across the ER membrane. Polymerization involves the addition of repeating units to the reducing end of the growing chain by Wzy polymerase. This enzyme has 12 predicted trans-membrane domains with the catalytic domain in the periplasm that has some specificity for the structure of the repeating unit. Wzy may invert the anomeric linkage of the first sugar in the polysaccharide since many repeating units have the GlcNAc $\beta$ -linkage in the O antigen. Many genes specifically involved in the synthesis

of the O antigen are found in the O antigen gene cluster. The presence of the *wzy* gene suggests that the O antigen is synthesized by the Wzy-dependent pathway (Figure 2). A much less specific chain terminator Wzz then helps to restrict the number of repeating units assembled in the O antigen. This is followed by a ligase (polysaccharide transferase)-catalyzed transfer of the O antigen to a specific sugar of the outer core structure, synthesizing the complete LPS. This releases PP-Und, which is recycled to P-Und. LPS is then extruded to the outer membrane by the Lpt complex (50).

The less common homopolymeric O antigens, such as the D-Rha polymers of *Pseudomonas aeruginosa* (*Pa*) and the D-Man polymers of *E. coli* O9, are synthesized by the transfer of mono-saccharides from nucleotide sugars to R-GlcNAc-PP-Und in a processive fashion in the ABC transporter-dependent pathway (Figure 3) (51). Some of the processive GTs can have multiple



catalytic domains, e.g., Man-transferase WbdA. The entire O antigen is assembled on the cytosolic side, and terminated by termination reactions, e.g., methylation. This is followed by translocation of the large O-antigen-PP-Und by the Wzm exporter and the ATP-binding Wzt to the periplasm where it is further processed. The presence of *wzm* and *wzt* genes in the O antigen gene cluster would suggest that this pathway is operative.

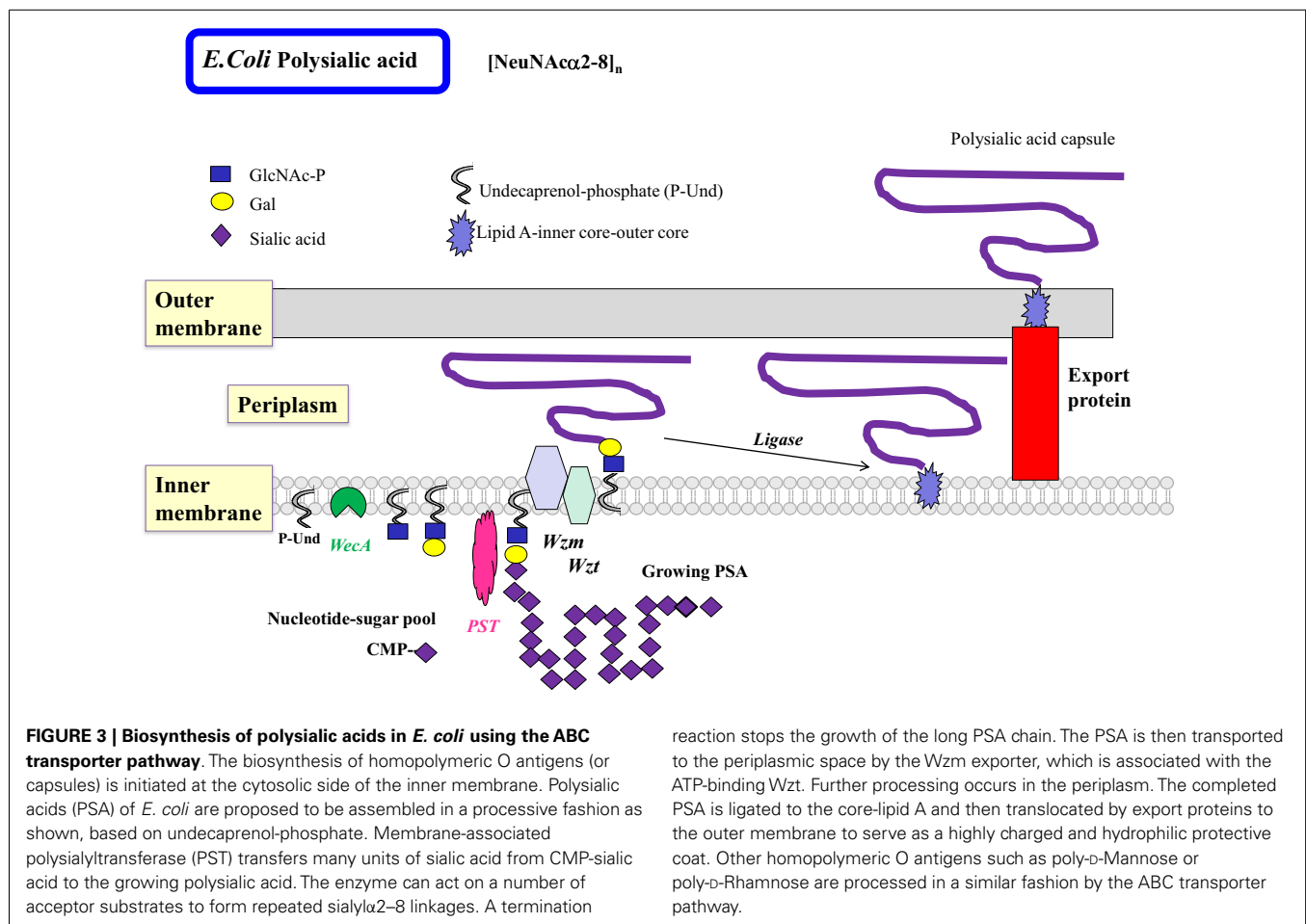
The events utilizing membrane-bound acceptor substrates in bacteria are similar to those of the early N-glycan synthesis in eukaryotes at the ER inner membrane (Figure 1). In both mammals and bacteria, isoenzymes are known that can synthesize the same linkage, often with slightly different substrate specificity. These isoenzymes are interesting models to study the catalytic sites and requirement for specific amino acids critical for catalysis and specificity.

### CHARACTERIZATION OF GLYCOSYLTRANSFERASES

Chemical synthesis has been used to produce natural-like or unnatural glycans but the stereochemistry and regio-selectivity is difficult to achieve. Nature has developed GTs, excellent tools

to synthesize an amazing diversity of glycan structures with defined anomeric configurations and linkages. GT reactions do not require harsh conditions or protection of reactive groups. GTs have distinct specificities for their donor and acceptor substrates. More than 100,000 genes from various species are thought to encode GTs, and organisms have 1–2% of their genes dedicated to GTs (52).

In order to assess the requirements and characteristics of GT activities, specific and accurate enzyme assays have to be developed. Nucleotide sugar donor substrates for mammalian glycoprotein biosynthesis are usually commercially available but for bacterial enzymes may have to be chemically or enzymatically synthesized. It would be difficult to extract the natural donor and acceptor substrates from bacteria in the pure form. Therefore, syntheses for bacteria-specific donor substrate analogs have been developed, e.g., for UDP-QuiNAc (UDP-6-deoxy-GlcNAc) found in *E. coli* and *Pseudomonas aeruginosa* (PA) (53) or for GDP-D-Rha found in PA (54). Oligosaccharides linked to a synthetic aglycone group may be suitable acceptors for both, mammalian GTs and bacterial GTs. However, bacterial GTs that act early in the



O antigen synthesis pathway seem to require sugar-diphosphate-lipids as acceptors, which are difficult to synthesize. We developed the natural acceptor analog GlcNAcα-diphosphate-lipid to mimic the product of the first sugar-phosphate addition (55), which was very active as an acceptor. In order to isolate the enzyme product from the assay mixture for quantification, a number of different chromatographic methods have been employed, including hydrophobic or anion exchange methods, HPLC, TLC, and capillary electrophoresis. Enzyme-coupled assays or lectin and antibody binding have also been used to determine activities. Methods to assay specific GTs are essential prerequisites to study their properties and optimal conditions, substrate specificities, and to develop inhibitors.

GTs can be classified based on similarities of their amino acid sequences, according to the sugar they transfer, and the stereochemistry of the reaction in the CAZy database. If at least 100 amino acids in two different stretches of the protein have significant similarity to other members of the same family but not to other families, GTs are assigned to a specific family with the same predicted fold, and being either inverting or retaining GT (Tables 3 and 4). However, not all known sequences fit into a GT family or are reclassified when the specific function of the GT has been established, and the number of families are growing. Sequence similarity of unknown proteins can be used to predict function

and protein folding. However, the final proof of function has to be obtained by biochemical analysis of enzymes. Most GTs in bacteria have not been functionally characterized, and this area is both challenging and tedious, often because the appropriate donor and acceptor substrates have to be especially prepared.

Crystal structures for GTs from eukaryotic and prokaryotic sources have been helpful in delineating the catalytic actions of GTs. It is interesting that this large and important class of thousands of enzymes that bind to many different nucleotide sugars as well as to a very large variety of monosaccharides, oligosaccharides, glycopeptides, and glycolipids occurs in only two major fold types, GT-A and GT-B. GTs, thus, have a relatively conserved three-dimensional architecture within their catalytic sites and share mechanisms, resulting in an extremely large number of product structures with linear or branched glycans of mostly unknown functions.

The binding of substrates to GTs and the transfer reactions have been shown to involve conformational changes in the enzyme proteins. GT-A folded enzymes have two tightly associated α/βα Rossman nucleotide-binding-like domains with two α-helices surrounding an open twisted, central β-sheet. The donor and acceptor substrates bind in different domains. The GT-B folded enzymes have two β/α/β Rossman-like domains, which are less tightly associated with each other and have the active site in the cleft in

between domains (56). Usually, the sugar donor substrate binds first. This induces a conformational change in the enzyme forming a lid over the nucleotide sugar, facilitating the binding of the acceptor substrate and catalysis in an ordered sequential, regio- and stereo-specific mechanism (57, 58). Internal disordered loops seem to be a common feature in mammalian and bacterial enzymes (40). Upon substrate binding, a disordered, short protein loop becomes ordered when donor substrate is bound. A change in orientation and conformation of the resulting ordered loop appears to facilitate binding of the second substrate and catalysis. Thus, the function of an ordered loop could be to allow catalysis, possibly by excluding water that would hydrolyze the donor substrate, or to form a lid over the nucleotide binding site allowing acceptor to bind, or to allow movement, and facilitating the reaction.

Generally, GTs have a distinct acceptor substrate specificity and with few exceptions, utilize only one type of nucleotide sugar donor substrate. Although few of the bacterial GTs have been biochemically characterized, it appears that both bacterial and mammalian GTs generally have similar properties with respect to their optimal pH, metal ion requirement, and donor specificity, although some bacterial GTs have a more promiscuous acceptor specificity (59). They can, thus, synthesize unnatural linkages that may find application as inhibitors or for biological studies. For example,  $\beta$ 4-Gal-transferase LgtB from *Helicobacter pylori* (Hp) has been used to synthesize thio-glycosides.

A comparison of mammalian and corresponding bacterial GTs (Table 2) shows that there is a low percentage of amino acid identity (often <12%), although the activities are comparable and the sugar transfer reactions follow similar mechanisms. Exceptions are ABO transferases, GTA and GTB, that synthesize blood group A and B, respectively, similarly in human beings and in certain bacteria, and show about 20% identity. Some of the  $\alpha$ 2 and  $\alpha$ 3/ $\alpha$ 4-Fuc-transferases also have similar activities when comparing human and bacterial GTs and show 14.5–17.5% identity. This suggests an exchange of genes between mammals and bacteria or a common evolutionary origin. The similarity and identity between GTs with similar function in bacteria or within the eukaryotic GT families can be much higher. The arrangements of amino acids in the catalytic site may therefore be similar, leading to the binding of the same nucleotide sugars and acceptors with the transfer of the sugar in a specific linkage. The requirement of a metal ion to stabilize the negative charge of the nucleotide sugar may also be the same. An evolutionally conserved feature of GTs is that the catalytic mechanism usually involves a catalytic base.

Inverting GT (52, 57, 58) inverts the anomeric configuration of the sugar in the donor substrate. This inversion is expected to follow a single displacement where the catalytic base deprotonates the hydroxyl group of the acceptor to be glycosylated, which then becomes an active nucleophile attacking carbon-1 of the sugar of the donor substrate. This mechanism involves an  $S_N2$  reaction and an oxocarbenium ion transition state. Crystal structures show that the catalytic base (Asp, Glu, or His) is properly positioned near the hydroxyl to be glycosylated. In many cases, this catalytic residue is within a conserved DxD motif (Table 5). Both the DxD motif and the negatively charged phosphate group of the nucleotide leaving group may be stabilized by a divalent metal

**Table 5 | Conserved peptide motifs in glycosyltransferases.**

	Peptide motif	Reference
Many GTs	DxD	(60)
Beta4-Gal-transferases	DxD, GWGxED	(60, 61)
Beta3-Gal-transferases	DxD, 2 motifs	(61)
Fuc-transferases	2-Fuc motif	(60, 62, 63)
	Two 3-Fuc motifs	(60, 62, 63)
	6-Fuc motif	(60, 62, 63)
	Three 2/6-Fuc motifs	(60, 63, 64)
	HxRRxD	
Beta3-GlcNAc-transferases	DxD, 2 motifs	(61)
C2GnT	SPDE	(20, 21)
Sialyltransferases, mammalian	Sialyl-motif L	(65)
	Sialyl-motif S	(65)
	Sialyl-motif VS	(65)
	Sialyl-motif III	(65)
Sialyltransferases, bacterial	D(E)-D(E)-G	(66)
	HP	
Polysialyltransferases, mammalian	PSTD motif	(67)
Polypeptide GalNAc-transferases	Ricin-like lectin domain	
ABO transferases	DxD	(60)

ion, but positively charged amino acids could also serve this function (20). Inverting GTs have been shown to act with a sequential ordered mechanism.

GTs that retain the anomeric linkage of the nucleotide sugar may function in a double displacement mechanism (58). Thus, for retaining GTs, a short-lived glycosyl-enzyme intermediate may form. This is followed by a shift in protein conformation that allows a nucleophilic attack on the anomeric center of the sugar by the deprotonated hydroxyl of the acceptor substrate to be glycosylated, maintaining the original anomeric linkage. A double displacement mechanism has been proposed for GalNAc-transferase GTA and Gal-transferase GTB, and a covalent glycosyl-enzyme intermediate through Cys303 was found (68). Other mechanisms may be possible and need to be investigated for retaining enzymes (58). GTs can also transfer sugar to water and thus have a nucleotide sugar hydrolase activity.

Mammalian GTs are single or multiple membrane-spanning proteins in the ER or single transmembrane-spanning type II membrane proteins in the Golgi, with a short cytosolic domain, a transmembrane anchor domain, and a stem region that helps the globular catalytic domain to protrude into the Golgi lumen. In the bacterial inner membrane, the first enzyme that adds the sugar-phosphate to P-Und such as WecA, as well as related sugar-phosphate transferases has multiple membrane-spanning domains. The remaining bacterial GTs that assemble O antigen repeating units do not have a transmembrane domain but have short hydrophobic stretches that may contribute to an association with membrane components. It is possible that both, mammalian and bacterial GTs, exist in protein/membrane complexes that activate enzymes and make the assembly of glycan chains highly efficient.

## A LARGE FAMILY OF Gal-TRANSFERASES

Families of at least 5  $\beta$ 3-Gal-transferases (B3GALTs) and at least 7  $\beta$ 4-Gal-transferases (B4GALTs) participate in forming the extensions of glycoproteins (69) that are the basis for the attachments of epitopes including the Lewis<sup>x</sup> antigen, the selectin ligand involved in the inflammatory response (8). These inverting metal ion-dependent GTs have a DxD motif, bind UDP-Gal and a number of GlcNAc-terminating acceptor substrates.

### THE B4GALT FAMILY

The crystal structures of both, human and bovine  $\beta$ 4-Gal-transferases 1 (B4GALT1) in complexes with donor and acceptor substrates and several mutants, have been thoroughly studied (22). UDP-Gal binds in a deep catalytic pocket of the bovine B4GALT1 together with  $Mn^{2+}$ , in the vicinity of Asp252, Asp318, and Glu317 residues. The conformational change induced by binding UDP-Gal creates the binding site for GlcNAc-terminating oligosaccharides. The GlcNAc moiety, which needs to be in the  $\beta$ -anomeric configuration is bound by Phe280, Phe360, Tyr286, Arg259, and Ile363. The enzyme has three DxD sequences.

In the bovine B4GALT1 enzyme, the first Asp254 residue in the DVD motif has contact with UDP and  $Mn^{2+}$  but mutations of Asp318 or Asp320 within the DDD sequence show that these residues are essential for activity. His344 normally interacts with  $Mn^{2+}$ . A His344Met mutant is active in the presence of  $Mg^{2+}$ , instead of  $Mn^{2+}$  and maintains a closed conformation bound to  $Mg^{2+}$  and UDP-hexanolamine, allowing an acceptor to bind. The mutant is, thus, useful to study the role of conformational changes and the binding of various acceptors (70, 71).

The catalytic domain of B4GALT1 has a short and a larger flexible loop containing the metal ion binding site. The binding of the donor and metal ion induces conformational changes in the long flexible loop, which changes from the open to the closed conformation, creating a lid over the bound nucleotide sugar. This opens an acceptor-binding site at the C terminus of the flexible loop. After the transfer reaction, the loop changes back to the open conformation, releasing the nucleotide (72).  $\beta$ 4-Gal-transferase 7 (B4GALT7) is another member of the same family, involved in priming glycosaminoglycan synthesis by adding Gal to Xylose (24). B4GALT7 also works in an  $S_N2$  type mechanism and changes conformation from closed to open conformation upon binding UDP and  $Mn^{2+}$ . The mammalian  $\beta$ 4-Gal-transferases have a common B4GALT motif GWGxED, which is not found in  $\beta$ 3-Gal-transferases or in the bacterial counterparts of B4GALT (61) (Table 5).

$\beta$ 4-Gal-transferases that synthesize Gal $\beta$ 1-4GlcNAc sequences are also found in bacteria. For example,  $\beta$ 4-Gal-transferase LgtB from *Helicobacter pylori* (Hp) can synthesize Gal $\beta$ 4-S-GlcNAc and Gal $\beta$ 4-Man linkages (59). The repeating unit of *Shigella boydii* (Sb) also contains the Gal $\beta$ 1-4GlcNAc sequence, which is synthesized by  $\beta$ 4-Gal-transferase WfeD (73). The sequences of human  $\beta$ 4-Gal-transferase and WfeD have about 9% identity; yet, the reaction catalyzed is similar. Both enzymes are inverting GTs, bind UDP-Gal and GlcNAc-R acceptor substrates, are activated by  $Mn^{2+}$ , and have a DxD motif. Interestingly, we found that both enzymes are also activated by  $Pb^{2+}$ , although the activation of the bacterial enzyme is much higher and is similar to  $Mn^{2+}$  activation. While

human  $\beta$ 4-Gal-transferase is in the GT7 family with a GT-A fold, the structure and predicted fold of the WfeD in GT family 26 is uncertain (Table 3). The human enzyme does not accept the negatively charged bacterial acceptor substrate, GlcNAc-PP-lipid, and vice versa, the bacterial enzyme cannot act on GlcNAc $\beta$ -Bn, which is the standard acceptor for assays of the human enzyme. Mutagenesis of WfeD showed that the central Glu101 residue of the DxE sequence is essential for activity. Lys211 was also found to be important, possibly by binding one or two phosphate group(s) of the acceptor substrate (73). Lys residues are apparently not involved in catalysis of the human enzyme. WfeD is not inhibited by GlcNAc $\beta$ -naphthyl, which is a potent inhibitor of the mammalian  $\beta$ 4-Gal-transferase (74).

### THE FAMILY OF $\beta$ 3Gal-TRANSFERASES (B3GALT)

Human glycoproteins can be extended with Gal $\beta$ 1-3GlcNAc (type 1) sequences that are also found in O antigens, e.g., in the repeating unit structure of the *E. coli* O7 antigen. There are five enzymes that synthesize the Gal $\beta$ 1-3GlcNAc linkage on a variety of acceptors in mammals. They are inverting GTs having a DxD motif and a requirement for divalent metal ions such as  $Mn^{2+}$  (15, 69). B3GALT5 has a distinct specificity for O-glycan core 3 (GlcNAc $\beta$ 1-3GalNAc-) acceptors. However, crystal structures are not available for  $\beta$ 3-Gal-transferases. Members of the  $\beta$ 3-Gal-transferase family have two common peptide motifs, in addition to the DxD motif (Table 5).

A  $\beta$ 3-Gal-transferase WbbD from *E. coli* O7 was detected that can act on GlcNAc $\alpha$ -PP-lipids where apparently the lipid structure is of minor contribution to the activity (75). The enzyme belongs to the GT2 family with a predicted GT-A fold and synthesizes the disaccharide Gal $\beta$ 3GlcNAc  $\alpha$ -linked to PP-lipid as the second step in repeating unit synthesis. Deletion of the enzyme eliminates the synthesis of O antigen on LPS. This supports the idea that an inhibition of this second step is successful in creating bacteria that are more susceptible to the mammalian immune system.

### BIOSYNTHESIS OF THE THOMSEN-FRIEDENREICH (TF) ANTIGEN

The cancer-associated T antigen, Gal $\beta$ 1-3GalNAc-, core 1, is the precursor for most O-glycans. In cancer, core 1 is often found in the unsubstituted form, while in normal glycoproteins, it is substituted by other sugars and is thus not recognized by anti-T antibodies. Sialylation of core 1 is also common in glycoproteins and often overexpressed in cancer and is recognized as the sialyl-T antigen (15). Several bacteria carry the T antigen as an internal structure within their O antigen repeating unit. The Shiga toxin producing O104 serogroup of *E. coli* is unusual in that it contains the T antigen in its O antigen repeating unit, as well as the sialyl-T antigen, sialyl $\alpha$ 2-3Gal $\beta$ 1-3GalNAc- (ECODAB).

The core 1 structure in human beings is synthesized by core 1  $\beta$ 3-Gal-transferase (T synthase, C1GALT1) and deficiencies of the enzyme are associated with pathological conditions including cancer. T synthase is the only known GT that requires the co-expression of a chaperone protein, Cosmc, C1GALT1C1 (76). C1GALT is a GT31 family member with a predicted GT-A fold, requires  $Mn^{2+}$  for activity and prefers GalNAc $\alpha$ -glycopeptides as substrates but can also transfer Gal from UDP-Gal to GalNAc $\alpha$ -benzyl and related substrates (77).



The GTs responsible for the synthesis of the T antigen in bacteria have a similar function (Table 2). The T synthase WbwC in the *E. coli* O104 strain is within the GT2 family (Table 3), and has only 10.5% identity compared to human C1GALT. No chaperone is necessary for the expression and activity of the bacterial enzyme (78). Both, human C1GALT and WbwC have a GT-A fold and DxD motifs, utilize UDP-Gal as a donor and require  $Mn^{2+}$  as a cofactor. However, in contrast to C1GALT, WbwC has a specificity for GalNAc $\alpha$ -diphosphate-lipid acceptor, while GalNAc $\alpha$ -peptides are not substrates. At this time, no crystal structure is available for T synthases but it is conceivable that the three-dimensional amino acid arrangements in the catalytic sites are similar. WbwC and human C1GALT could be distinguished using bis-imidazolium salt inhibitors, which showed that only WbwC, but not human C1GALT, was strongly inhibited with IC<sub>50</sub> values of 8  $\mu$ M (78). These inhibitors could selectively attack GTs in pathogenic bacteria. However, a potent inhibitor for T synthase has yet to be discovered (77).

### P BLOOD GROUP SYNTHESIS

Human blood group P (Table 1) and related, complex structures containing the Gal $\alpha$ 1–4 linkage are synthesized by  $\alpha$ 4-Gal-transferases (A4GALT), mainly using glycolipids with Gal residues as acceptors, e.g., lactosylceramide (79). However, a different  $\alpha$ 4-Gal-transferase from pigeon, related to  $\beta$ 4-Gal-transferase from the same species, but not to  $\beta$ 4-Gal-transferases from human beings, has been described that preferably acts on the N-glycans of glycoproteins (80).

A number of bacteria, including *Cj* (81), also express an  $\alpha$ 4-Gal-transferase with about 11% identity to the human enzyme (Table 2). The LgtC  $\alpha$ 4-Gal-transferase from *Nm* synthesizes the bacterial mimic of the human P blood group (45). The enzyme is a member of the GT8 family with a GT-A fold and follows a bi-bi kinetic mechanism where UDP-Gal binds first. The crystal structure of LgtC with analogs of UDP-Gal and lactose substrates suggests that Asp103 and Asp105 of one of the four DxD motifs, as well as His244, are in the vicinity of the donor substrate, while a  $Mn^{2+}$  ion coordinates the phosphates of UDP. The mainly helical C terminus is expected to form hydrophobic and electrostatic interactions with the bacterial membrane. Multiple conformational states of LgtC with and without bound substrate analogs were found by methyl-TROSY NMR (82), which is additional information that cannot be obtained by static crystal structure analysis.

### A NEW DxDD MOTIF IN GT2 TRANSFERASES

A new DxDD motif (Table 5), essential for activity, was discovered in WbwC (78). This motif is also present in WbdN, WfaP, WfgD, WbgO, WbiP, and CgtB (83–87). All of these GTs in the GT2 family having a DxDD motif are specific for the transfer of either Gal or Glc in  $\beta$ 1–3 linkage to GalNAc or GlcNAc. Mutagenesis showed that in WbiP from *E. coli* O127 (83), the first Asp of the DxDD sequence was critical for activity while the second contributed but was not essential. In WbwC from *E. coli* O104 and O5, all three Asp residues were mutated and found to be important for activity. The first Asp (D91) is probably the catalytic base. The other Asp residues may support the nucleophilic property of the catalytic base (78).

While WbwC synthesizes the Gal $\beta$ 1–3 linkage attached to the first GalNAc residue at the reducing end of the O antigen repeating unit, several other GTs having a DxDD motif in the GT2 family were shown to synthesize the T antigen at a more internal position of the repeating unit. These GTs have a different specificity from that of WbwC and do not require the diphosphate in the acceptor. The T synthase activities of variants of CgtB from *Cj* mainly act on  $\beta$ -linked GalNAc acceptors. Variants of CgtB have distinct acceptor specificities (86) and synthesize lipooligosaccharides, which mimic mammalian glycolipids and glycoproteins.

### GlcNAc-TRANSFERASES FORM BACKBONE STRUCTURES

Gal-transferases cooperate with five or more  $\beta$ 3-GlcNAc-transferases (B3GNT) within the GT31 family to form the type 1 and 2 backbone structures of mammalian glycan chains (15, 88, 89) (Table 3). B3GNTs have significant sequence similarity with Gal-transferases. It is not known if these enzymes are physically associated, although their combined action would suggest this. A family of three  $\beta$ 6GlcNAc-transferases (IGnT, GCNT2) then can add 1–6 branches to the linear chains. The  $\beta$ 3-GlcNAc-transferases, but not the  $\beta$ 6-GlcNAc-transferases, require divalent metal ions for activity. No crystal structures are yet available for B3GNTs.

In the N-glycosylation pathways, GnT I to V (MGAT1 to 5) (12) are responsible for forming GlcNAc-based antennae that can be further extended through repeating linear or branched GlcNAc $\beta$ 1–3Gal-disaccharides. MGAT1 is an inverting GT with a GT-A fold within the GT13 family. The crystal structure of rabbit GnT I with UDP-GlcNAc and  $Mn^{2+}$  supports an ordered sequential mechanism. The DxD motif is present as EDD, with Glu211 being the likely catalytic base (19). MGAT2, 3, 4, and 5 are all inverting GTs and have been classified in the GT16, GT17, GT54, and GT18 families respectively. Although the GT17 family also contains uncharacterized bacterial proteins, no bacterial equivalents of MGAT have been found in bacteria.

In the O-glycosylation pathways, the basis of most extended chains is core 2. Core 2  $\beta$ 6-GlcNAc-transferase C2GnT1 (GCNT1) adds a branch to O-glycan core 1 to form the core 2 structure GlcNAc $\beta$ 1–6(Gal $\beta$ 1–3)GalNAc-R (15). The enzyme has a GT-A fold and is classified in the GT14 family. The crystal structure of the catalytic domain of mouse C2GnT1 shows that the protein has four conserved intramolecular disulfide bonds (20, 21). Cys217, however, has to be reduced to support the activity, although it is not an essential residue (90). The human enzyme expressed in insect cells has two flexible N-glycans that protect the protein from degradation (91). C2GnT1 is an inverting GT that is active in the presence of EDTA and does not require  $Mn^{2+}$ . The crystal structure suggests that the conserved, basic amino acids Arg378 and Lys401 stabilize the diphosphate group of UDP-GlcNAc and thus serve the function of  $Mn^{2+}$ . The structure supports specificity studies of C2GnT1, showing an absolute requirement for the 4- and 6-hydroxyl groups of the Gal and GalNAc residues and the 2-acetamido group of GalNAc (77). Glu320 of the conserved SPDE sequence may be the catalytic base; it binds to the 4 and 6-oxygen of GalNAc and could thus deprotonate and activate the 6-hydroxyl to induce a nucleophilic attack on the C-1 of the GlcNAc moiety of UDP-GlcNAc (20, 21).

Bacteria do not appear to have C2GnT or GnT I equivalents, but they express type 1 of type 2 chains and  $\beta$ 3-GlcNAc-transferases comparable to the mammalian enzymes in their activities. For example, a  $\beta$ 3-GlcNAc-transferase from *Hp* is involved in the synthesis of lipooligosaccharides and GlcNAc $\beta$ 1–3Gal- extensions that resemble mammalian epitopes (92). The  $\beta$ 3-GlcNAc-transferase LgtA from *Nm* acts on lactose and has a relaxed donor specificity. It is most active with UDP-GlcNAc but can also utilize UDP-GalNAc (93). Both, the mammalian and bacterial  $\beta$ 3-GlcNAc-transferases accept a wide variety of acceptor substrates but have low sequence identity (**Table 2**) (15).

## FUCOSYLTRANSFERASES

Three different types of Fuc-transferases (FUT) are involved in glycoprotein biosynthesis. Peptide motifs have been identified that are specific for  $\alpha$ 2-Fuc-transferases,  $\alpha$ 3-Fuc-transferases, or  $\alpha$ 6-Fuc-transferases or are shared by  $\alpha$ 2- and  $\alpha$ 6-Fuc-transferases (62). The mammalian  $\alpha$ 3-Fuc-transferases (FUT3-7,9-11) are inverting enzymes (**Table 3**) and have two shared motifs with similar spacing, shared by eukaryotic and bacterial  $\alpha$ 3FUT. The  $\alpha$ 2-Fuc-transferases FUT1 and 2 have less than 30% sequence identity with their bacterial FUT counterparts, but have well-preserved  $\alpha$ 2-Fuc-transferase motifs. There are also similarities between  $\alpha$ 2- and  $\alpha$ 6-FUT (FUT8) (60) with three common motifs (I to III), shared among eukaryotic and bacterial enzymes (62). The different types of FUT may have evolved from a common ancestor by divergent evolution (63).

## FUCOSYLTRANSFERASES THAT SYNTHESIZE THE H ANTIGEN

The blood group O (H antigen, Fuc $\alpha$ 1-2Gal-R) is found in virtually all human beings and in certain bacteria and is the precursor substrate structure to form blood groups A and B. The enzymes that synthesize the H antigen in human beings are inverting  $\alpha$ 2-Fuc-transferases 1 and 2 (FUT1 and FUT2) that are closely related in sequence to the GT6 family ABO transferases GTA and GTB, although FUT1 and 2 have been classified into a different (GT11) family. FUT1 has a broad acceptor specificity for Gal $\beta$ -R while FUT2 prefers O-glycan core 1 (T antigen) as a substrate (15).

Similar enzymes (**Table 2**) have been identified in *Hp* as FutC (94), in *E. coli* O86 as WbwK (95), as WbsJ in *E. coli* O128 (64), and WbiQ in *E. coli* O127 (96). WbwK and WbiQ have a distinct specificity for the T antigen (95, 96) and do not act on Gal $\beta$ 1–4 glycans. These FUT, therefore, have an activity resembling that of human FUT2 and have 12–17.5% sequence identity. HpFutC2 (FutC) adds Fuc preferably to Lewis x acceptors but also uses Lewis a and type 1 chains (94). In contrast, WbsJ prefers acceptors with terminal Gal $\beta$ 1-4Glc structures (64). WbsJ functions in the absence of divalent metal ion and does not have a DxR motif. Especially the first Arg residue of the HxRRxD motif, conserved in  $\alpha$ 2- and  $\alpha$ 6-Fuc-transferases, is critical for activity due to its positive charge. Domain swapping between WbwK and WbsJ showed that the C-terminal motifs function in determining acceptor specificity (95). All of the identified  $\alpha$ 2-Fuc-transferases have significant homology in GT family 11 (**Table 3**) with a predicted GT-B fold but none have been crystallized.

## FUCOSYLTRANSFERASES INVOLVED IN THE SYNTHESIS OF LEWIS ANTIGENS

Lewis type antigens play essential roles in cell adhesion in the immune system and during inflammation, and aberrant amounts are often found in cancer. A family of mammalian, inverting  $\alpha$ 3-Fuc-transferases (FUT3–7, 9–11) is involved in Lewis antigen synthesis by linking Fuc to GlcNAc (9, 15, 97, 98). The enzymes vary in their acceptor substrate specificities and cell type expression and are in the GT10 family with a GT-B fold. FUT3 is an exceptional enzyme that has a dual specificity and adds Fuc $\alpha$ 1–3 on type 2 chains to synthesize Lewis x and y, as well as Fuc $\alpha$ 1–4 to type 1 chains to synthesize Lewis a and b (**Table 1**). FUT5 also shows some  $\alpha$ 4-Fuc-transferase activity. Human FUT3 and 5 have Trp111, responsible for type I acceptor recognition and 1–4 linkage synthesis. FUT that do not have this Trp synthesize the 1–3 linkage (99).

The bacterial  $\alpha$ 3FUTs show weak homology to mammalian FUT in two small segments of the catalytic domains ( $\alpha$ 3FUT motifs). They have about 10% sequence identity and a common GT-B fold but no transmembrane domain (25, 62). Two amphipathic  $\alpha$ -helices serve to anchor the enzymes in the membrane. The gastric pathogen *Hp* is a prime example of expressing human-like type 1 and type 2 chains that are fucosylated and include Lewis antigens, which may play a role in adhesion to gastric epithelial cells or in internalization. *Hp* have short O antigens (lipooligosaccharides) and the human glycan mimics help to mask the immunogenic determinants of *Hp*, thus evading immune surveillance and supporting persistent *Hp* infections. The different pH environments in the various regions of the stomach influence the expression of Lewis antigens, and likely the activities of GTs, leading to phase variations.

A number of bacteria have Fuc $\alpha$ 1–4 linkages but *Hp* is especially rich in Lewis a, b, x, and y structures and in  $\alpha$ 3/4-Fuc-transferase activities (100). All of the eukaryotic and most bacterial  $\alpha$ 3-Fuc-transferases are in the GT10 family. *Hp* has futA and futB genes encoding  $\alpha$ 3FUT, in addition to 1-3/4 FUT (FucTa). FucTa has the CNDAAHYSALH sequence near the C terminus that controls type I chain recognition. It seems that in this  $\alpha$ 3/4 FUT, it is Tyr instead of Trp that determines the acceptor preference. Thus, the Y350A mutant synthesizes Lewis x since it had dramatically reduced  $\alpha$ 4 FUT activity (100).

The crystal structure of  $\alpha$ 3-Fuc-transferase from *Hp* shows that a Glu95 residue is positioned closely to the anomeric carbon of Fuc of the donor GDP- $\beta$ Fuc and could be a catalytic base (25) while Glu249 could stabilize the intermediate oxonium ion. Mutants in these Glu residues are virtually inactive. Interestingly, tandem repeats of 7 amino acids (DDLRLNY) are found in this  $\alpha$ 3FUT. The 2–10 heptad repeats appear to connect the N terminus to 2 amphipathic helices at the C terminus and are thought to be involved in maintaining secondary structure and activity (101). The C terminal sequence appears to determine the stability and overall structure of the protein.

A different  $\alpha$ 3-Fuc-transferase HhFT2 from *Helicobacter hepaticus* (*Hh*) synthesizes the Lewis x as well as the sialyl-Lewis x antigen (102). This enzyme is a member of the GT11 family and has more homology to  $\alpha$ 2-Fuc-transferases such as WbsJ of GT11, but less to  $\alpha$ 3/4 FUT in GT family 10. It has 10.4% sequence

identity with the human enzyme FUT4. HhFT2 has three conserved motifs, one at the N terminus, one central, and one near the C terminus (Table 5).

#### SYNTHESIS OF THE Fuc $\alpha$ 1-6GlcNAc LINKAGE

The  $\alpha$ 6-Fuc-transferases add Fuc in  $\alpha$ 1–6 linkage to the reducing end GlcNAc of the N-glycan core. The human enzyme (FUT8) requires the prior action of GnT I and cannot act when the chitobiose of the N-glycan core carries an  $\alpha$ 3Fuc residue, or if the internal Man $\beta$  residue carries the bisecting GlcNAc. FUT8 is classified in family GT23 with a GT-B fold. The crystal structure of human FUT8 shows three domains: an N-terminal coiled-coil domain, a catalytic domain that resembles GT-B folded GTs, and a C-terminal SH3 domain, although its significance is unknown. The C-terminal part of the catalytic domain contains a Rossmann-like fold with three regions, conserved in  $\alpha$ 2-,  $\alpha$ 6-, and other Fuc-transferases. Both Arg365 and Arg366 are critical for binding to GDP-Fuc while Asp453 may be a critical catalytic base (26).

A bacterial  $\alpha$ 6-Fuc-transferase with similar activity in the GT23 family with a GT-B fold and only 8% sequence identity is NodZ from *Rhizobium* sp. (*Rsp*) (Tables 2 and 4) (103). The crystal structure of NodZ shows two domains of nearly equal size but with different shape, separated by a central cleft (27). There are three conserved sequence motifs near the C terminus that play a role in GDP-Fuc binding or catalysis.

#### GLYCOSYLTRANSFERASES THAT SYNTHESIZE BLOOD GROUPS A AND B

The two human ABO transferases that synthesize the antigenic blood group A and B determinants from the H antigen ( $\alpha$ 3-GalNAc-transferase GTA and  $\alpha$ 3-Gal-transferase GTB, respectively) are homologous retaining enzymes within the GT6 family with a GT-A fold (Table 3). It is astounding that the critical difference in donor specificities determining blood group A or B lies in a difference of only four amino acids. While GTB that transfers Gal has Gly176, Ser235, Met266, and Ala268, the GTA protein that transfers the slightly larger GalNAc has mostly smaller amino acids Arg176, Gly235, Leu266, and Gly268.

In GTA and GTB, two domains are separated by a catalytic cleft containing the DxD motif (Asp211–Asp213) (39). However, a highly conserved Glu303 is likely to be the active nucleophile. UDP binds in the nucleotide sugar-binding domain at the N terminus and the Mn<sup>2+</sup> ion coordinates the  $\beta$ -phosphate of UDP. The H antigen acceptor binds to the C terminus. A disordered and flexible internal loop adjacent to the active site (40) becomes ordered when the nucleotide (sugar) is bound. This leads to a conformational change in the protein (43). Two amino acids are in contact with donor or acceptor in GTA and GTB (39) but only one of them determines the binding of the nucleotide-bound sugar moiety, i.e., either Gal or GalNAc. Leu266 in GTA has contact with the acetamido group that allows binding of UDP-GalNAc. Due to the larger Met in this position (Met266), GalNAc cannot be accommodated and, therefore, Gal binds. Ala/Gly268 has contact with the 3- and 4-hydroxyl groups of Gal and thus does not contribute to the difference in donor specificity.

Human beings have antibodies against the absent blood group (A or B), and it is possible that this is induced by bacteria displaying

this blood group. A number of bacterial GTA-like enzymes are also in the GT6 family and resemble the human counterpart with relatively high sequence identity of about 20%. The similarities between human and bacterial enzymes suggest a horizontal gene transfer between species and between bacteria. The bacterial enzymes have an NxN sequence instead of the eukaryotic DxD motif, and most of these enzymes do not have a metal ion requirement. Thus, bacterial enzymes may have altered catalytic mechanisms, although there is a strong conservation of mammalian-type of residues in the active sites (104).

*Helicobacter mustelae* (*Hm*) synthesize the blood group A determinant, which reacts with anti-human blood group A antibodies (105). The enzyme responsible, GTA-like  $\alpha$ 3-GalNAc-transferase (BgtA), has 20% sequence identity to its human counterpart GTA and can act on Fuc $\alpha$ 1-2Gal $\beta$ 1–3-R or Fuc $\alpha$ 1-2Gal $\beta$ 1–4-R substrates (Table 2). Thus, bacteria may have acquired the GTA gene from a mammalian host, enhancing their molecular mimicry, although it is not clear how the human blood group is giving them a selective advantage.

The GTA-like enzyme BoGT6a from *Bacteroides ovatus* (*Bo*) (44) and GTB-like  $\alpha$ 3-Gal-transferase WbnI from *E. coli* O86 (95) that synthesize blood group B are related to the human enzymes with significant sequence homology in the GT6 family. Both donor and acceptor substrates are the same as those for GTA and GTB from human beings. The crystal structure of BoGT6a revealed a disordered region, which becomes ordered when acceptor Fuc-lactose is bound. This is accompanied by a large conformational change from the open to a closed state. Isothermal titration calorimetry (ITC) experiments showed that BoGT6a binds UDP-GalNAc with high affinity.

In non-primate mammals and new world monkeys, the linear blood group B occurs (Gal $\beta$ 1–3Gal-), without the  $\alpha$ 2-linked Fuc residue. This structure is foreign to human beings who have anti-linear B antibodies, thus hindering xenotransplantation. The  $\alpha$ 3-Gal-transferase A3GALT that synthesizes the linear B determinant is a homolog of GTA and GTB and has been crystallized with UDP and Mn<sup>2+</sup> (41). The invariable Glu317 was identified as the catalytic base. The crystal structure in a complex with Gal $\beta$ -pnp suggests that Trp residues are critical for binding the natural substrate Gal $\beta$ 1-4GlcNAc (106). The disordered C terminal region is critical for allowing the substrate to bind (42). Bacterial analogs of this  $\alpha$ 3-Gal-transferase remain to be characterized.

#### SIALYLTRANSFERASES IN MAMMALS AND BACTERIA

Sialyltransferases are ubiquitous in eukaryotes and are also expressed in certain bacteria (107). These enzymes synthesize sialic acid linkages commonly found on the non-reducing termini of N- and O-glycans, and gangliosides as sialyl $\alpha$ 2–3Gal $\beta$ 1- or sialyl $\alpha$ 2–6Gal(NAc)-linkages. In addition, sialyl $\alpha$ 2–8 linkages are found, especially in polysialic acids (PSA), which are extremely large, linear polymers, expressed in a cell type specific, restricted fashion in embryonic, neuronal, and other selected cell types (108). Sialic acids contribute to the acidity and hydration of a glycoprotein, the metal ion binding, and epitope exposure. While sialic acid can mask the underlying epitope, certain lectins of the immune system (e.g., siglecs) directly recognize sialic acid in specific linkages. Metastatic cancer cells and leukemia cells are often hypersialylated,

which reduces further processing of glycans and causes glycan chains to be shorter (15). Sialylation significantly affects the adhesive properties of cells and has also been implicated in the functions of cell surface receptors (109).

Sialyltransferases are inverting GTs that usually lack a DxD motif and do not require divalent metal ions. Thus, general acids and bases identified in the crystal structures of  $\alpha$ 3- and  $\alpha$ 6-sialyltransferases that may interact closely with the substrates include His residues (28, 33). All known eukaryotic sialyltransferases have been classified as inverting GT29 with a GT-A fold, having at least four sialylmotifs (Table 5), a large (L), small (S), very small (VS), and motif III (65). The L motif contains the donor binding site while the S motif also binds the acceptor.

Bacterial sialyltransferases are inverting enzymes that bind CMP-sialic acid donor substrate and acceptors terminating in Gal or sialic acid but do not have these sialylmotifs and do not belong to the GT29 family. Instead, they are classified as GT42 (with a GT-A fold), GT52 or GT80 (with a GT-B fold), and GT38 (polysialyltransferases, PSTs). The 6-sialyltransferases (ST6GalNAc) acting on O-glycans do not appear to have a bacterial counterpart. Two highly conserved short motifs have been identified in bacterial PST and other bacterial sialyltransferases (GT52 and GT80), a C-terminally located HP sequence and a more N-terminally located D/E-D/E-G sequence (66). Certain bacterial sialyltransferases have multiple activities, including CMP-sialic acid hydrolase, transsialidase, and neuraminidase activities and are usually from the GT80 family. Thus, sialyltransferases can be promiscuous with respect to the linkages they form (or cleave) and the acceptor substrates they recognize. Bacterial sialyltransferases probably evolved separately from the eukaryotic enzymes, although their functions and mechanisms can be similar.

### ALPHA3-SIALYLTRANSFERASES

In human beings, 6  $\alpha$ 3-sialyltransferases (ST3GAL) form the sialyl $\alpha$ 2–3 linkage. The expression and activity of  $\alpha$ 3-sialyltransferase ST3GAL1 that synthesizes sialyl-T antigen are increased in breast cancer (110) and appear to promote survival of cancer cells in the blood (111). In keeping with its activity in adding a terminal structure, ST3GAL1 is localized to the medial and late Golgi compartments in human mammary cells (112). ST3GAL1 acts on glycopeptides with core 1 structure and also on Gal $\beta$ 1–3GalNAc $\alpha$ -R acceptors that have hydrophobic aglycone groups. In contrast, a bacterial equivalent, WbW from *E. coli* O104, responsible for the rare occurrence of the sialyl-T antigen in *E. coli*, is in the GT52 family with a GT-B fold. WbW, but not mammalian ST3GAL1, also has HP and D/E-D/E-G motifs (Table 5). The crystal structure of porcine ST3GAL1 with CMP and Gal $\beta$ 1–3GalNAc-acceptor substrate suggests that the essential His302 interacts with the phosphate of CMP-sialic acid. His319 is the catalytic base in motif III that is proposed to be positioned near carbon-2 of the sialic acid moiety (28). The conserved Tyr269 residue interacts with the 4-hydroxyl of GalNAc and thus determines the enzyme specificity for Gal $\beta$ 1–3GalNAc- over the Gal $\beta$ 1–3GlcNAc- acceptors.

An  $\alpha$ 3-sialyltransferase of the GT80 family from *Photobacterium Phosphoreum* (Pp) has been crystallized with CMP (29). The acceptor-binding site has a wide access explaining that a range of

possible disaccharides with Gal $\alpha$  and Gal $\beta$  linkages can form substrates. CMP binds in a cleft between the two domains of the GT-B fold. The main chain nitrogen of His317 in the HP motif is close to the nitrogen-4 of Cytidine and the side chain of His317 is near the oxygen of the phosphate. This suggests a critical role of these His residues in catalysis. Another  $\alpha$ 3-sialyltransferase with a GT-B fold in family GT52 from *Nm* was crystallized (31) with the donor analog CMP-3F-Neu5Ac. Asp258 could be a general base and His280 (within the HP motif) a general acid.

### ALPHA6-SIALYLTRANSFERASES

Human  $\alpha$ 6-sialyltransferases add sialic acid to the Gal termini of N-acetylglucosamine chains of N-glycans. ST6GAL1 is highly expressed in colon cancer and metastatic cells (113) and also resides in the trans-Golgi (114). A homolog with 48% sequence identity (ST6GAL2) is mainly expressed in the brain and has the additional ability to synthesize sialyl $\alpha$ 2–6GalNAc $\beta$ 1–4GlcNAc-structures (115). Human ST6GAL1 is a glycoprotein stabilized by three disulfide bonds (33). The catalytic residue, His370, deprotonates the 6-hydroxyl of Gal, generating an active nucleophile that attacks the carbon-2 of sialic acid. The reaction follows a random-order mechanism of substrate binding. Rat ST6GAL has three disulfide bonds and two N-glycans (32). As many GTs, the enzyme has a disordered loop, and His367 is the catalytic base within the sialyl motif VS.

The bifunctional bacterial  $\alpha$ 6-sialyltransferase PM0188 from *Pasteurella multocida* (Pm) of GT family 80 has 14.6% sequence identity to human ST6GAL1. The crystal structure showed the GT-B fold and that Asp141, His311 (within the HP motif), Glu338, Ser355, and Ser356 were important for catalysis (37). The *Photobacterium* sp. (Psp)  $\alpha$ 6-sialyltransferase was also crystallized with CMP and lactose (35). The enzyme is in the GT80 family with a GT-B fold and has three domains, with the donor and acceptor bound between domains 2 and 3. Asp232 (within the D/E-D/E-G motif) is near the 6-hydroxyl of Gal while the nitrogen of His405 (within the HP motif) is close to the phosphate-oxygen. Thus, Asp232 could be a catalytic base that deprotonates the 6-hydroxyl of Gal, and His405 could be a catalytic acid that protonates the donor substrate.

### MULTIFUNCTIONAL SIALYLTRANSFERASES

In bacteria, mimics of human sialyl $\alpha$ 2-3/6/8Gal $\beta$ 1- structures occur, e.g., in the lipooligosaccharides of Gram-negative bacteria such as *Cj* (116, 117). Cells of the human nervous system are rich in gangliosides as well as glycoproteins containing similar sialyl-linkages. Thus, after bacterial infections, cross reactivity of antibodies could cause the rare development of neurological disorders. Guillain-Barré syndrome is an example (118). *Cj* expresses a bifunctional  $\alpha$ 3/8-sialyltransferase CstII and an  $\alpha$ 3-sialyltransferase CstI, which are responsible for the molecular mimicry of *Cj* in their lipooligosaccharide structures. Both enzymes have a predicted GT-A fold within the GT42 family. The structure of CstI shows (30) that His 202 is the catalytic base. Similarly, His188 is likely the catalytic base in CstII that deprotonates the 3-hydroxyl of Gal, which then attacks carbon-2 of sialic acid of the donor (119). The flexible lid in the CstII protein becomes ordered in a closed form when CMP binds (36). The acceptors



lactose (for  $\alpha 3$ -sialyltransferase activity) or sialyl-lactose (for  $\alpha 8$ -sialyltransferase activity) bind in a cleft and Arg129, Asn51, and Tyr81 contribute to the binding of the sialylated acceptor. The role of His188 as a catalytic base in CstII has also been confirmed by NMR studies (120). The intrinsic  $pK_a$  values of His188 were measured in monomeric mutants by determining the pH-dependent chemical shifts of [ $^{13}\text{C}$ ]-labeled His188.

The monofunctional sialyltransferases function with similar mechanisms compared to the mammalian enzymes. Multifunctional enzymes, however, are primarily found in bacteria and include the  $\alpha 3$ -sialyltransferase PmST1 from *Pm*, which binds CMP-sialic acid as donor and lactose, Gal, GalNAc as well as sialic acid as acceptor. The crystal structure shows that binding of CMP-sialic acid donor substrate causes a change in conformation and opens the acceptor-binding site. The activities of PmST1 function optimally at different pH values. It has a GT-B fold within the GT80 family (Table 3). The crystal structure shows that Asp141 is the catalytic base (34) with His112 also being important for enzyme activity. Another multifunctional  $\alpha 3$ -sialyltransferases PdST from *Pasteurella dagmatis* (*Pd*) in the GT80 family with a GT-B fold is also a CMP-sialic acid hydrolase. At low pH, it can act as a trans-sialidase and a sialidase (121).

## POLYSIALYLATION

Important sialic acid structures are the PSA, found in human neuronal and other selected cell types (107). Only a selected number of proteins carry the PSA modification (122). For example, polysialylated neural cell adhesion molecule N-CAM is prominent in the developing nervous system but also occurs in leukocytes with roles in the regulation of cell adhesion. N-CAM becomes anti-adhesive when long polymers of  $\alpha 2$ -8-linked sialic acids are covalently attached to its N-glycans (108). The sialyl $\alpha 2$ -8 linkages of PSA are synthesized based on sialyl $\alpha 2$ -3/6Gal residues of N-glycans by developmentally regulated PSTs, which are highly expressed in the developing and embryonic brain (123). Neuropilin-2 (NRP2) is a glycoprotein containing multiple N-glycosylation sites, as well as O-glycans with sialylated core 1 and 2 structures. In cells lacking core 2, human PST (ST8SialV) was shown to assemble PSA on sialylated core 1 chains of neuropilin (124). The presence of these PSA polymers extends the half-life of proteins.

*E. coli* and *Nm* are examples of bacteria that carry sialyl $\alpha 2$ -8 polymeric PSA capsules, which help bacteria to resist phagocytosis. These PSA capsules mimic the eukaryotic chains, although they are linked to the membrane via a lipid anchor, and may have bacteria-specific modifications such as O-acetylation (125). The large, charged and hydrated polymeric enzyme product is assembled in the cytoplasmic compartment and then extruded through the membranes by ABC transporter and export systems (Figure 3). PSA confers a selective advantage to bacteria in the human nervous system and is associated with meningitis or other neurological conditions. Bacteria may also have PSA with  $\alpha 2$ -9 linkages or alternating  $\alpha 2$ -8 and  $\alpha 2$ -9 linkages.

In mammals, PSTs synthesize PSA by the addition of individual sialic acid residues in a processive fashion. Like the other sialyltransferases, mammalian PSTs are inverting enzymes of the GT29 family (Table 3). In addition to four sialyl motifs, ST8SIAII and ST8SIAIV (formerly STX and PST, respectively) have a unique,

conserved, polybasic PST domain (PSTD motif) (Table 5) (67), which is absent from the other types of sialyltransferases. Basic residues in the PSTD motif are responsible for acceptor substrate recognition (126, 127).

In bacteria, the PSA capsule is synthesized by GT38 family PST that are inverting enzymes. In *E. coli*, PST has only 5.4% sequence identity with the human enzyme (128). The human PST equivalent from *Nm* has <10% sequence identity with human PST and has a requirement for  $\text{Mg}^{2+}$  (129). Kinetic experiments of His and Pro mutants of the PST from *Nm* suggested that the HP motif contributes to CMP-sialic acid but not acceptor binding. The acceptors can be a glycolipid containing two sialic acid residues as a primer. Gal-terminating glycans of glycopeptides, including the T antigen linked to Ser, also served as acceptor substrates for the PST form *Nm*. Different PSTs synthesize either the  $\alpha 2$ -8 or  $\alpha 2$ -9 linked polymers of bacterial PSA capsules.

## METHODS TO STUDY GLYCOSYLTRANSFERASE PROTEIN STRUCTURES AND FUNCTIONS

It is often difficult to produce sufficient pure enzyme in order to analyze protein structure by X-ray crystallography. In addition, the protein may not show exactly the same properties in a crystal, compared to solution and body fluids. To approximate the protein structure present in the natural environment, protein NMR studies have been helpful (130). Enzyme substrate or inhibitor interactions have been determined by biochemical kinetics studies but can also be studied by MS and Saturation Transfer Difference (STD) NMR (131, 132). Conformational dynamics of proteins to understand molecular recognition can be achieved by molecular dynamics simulation and docking programs, which requires knowledge of protein structure. Theoretical modeling has been undertaken to predict protein structure, substrate binding, and dynamic properties of GTs. Thus, the three-dimensional interactions between substrates and enzyme protein, cofactor binding sites, ligand flexibility, and movements can be estimated by computational methods (133, 134). Multivariate data analysis of the amino acid property patterns also helps to predict a protein fold (135).

New enzymes can be designed based on knowledge of protein structure and substrate binding. For example, the blood group B GTB Gal-transferase has been re-designed with a model Epimer Propensity Index (EPI) to transfer Glc instead of Gal (136). The orientation of the sugar donor in the folded enzyme is highly conserved. The R228K mutant of  $\beta 4\text{GalT1}$  has higher Glc-transferase activity due to the inability to effectively bind the axial 4-hydroxyl of Gal (23). Similarly, GTB modeling correctly predicted a higher Glc-transferase activity of GTB in the presence of the unnatural UDP-Glc donor upon increasing the sizes of Ser185 to Asn and Cys (136).

One approach to developing good GT inhibitors is to obtain qualitative and quantitative information on the substrate binding sites from NMR spectroscopy. STD NMR measures the signals of the unbound substrate, which is then compared to those of the bound substrate. Saturation transferred from the enzyme to specific sites of the bound substrate is seen as an attenuation of resonance signals. The difference spectra at different ligand concentrations allow to identify the bound substrate and to determine



the binding affinity. In spin-lock filtering experiments, transverse relaxation of substrate signals is recorded, which is enhanced when the ligand is bound. Thus, signals are attenuated upon ligand binding. This process can be enhanced by using spin labels. The conformations and relative placements of bound GnT V substrates have been determined using transferred NOE and STD measurements (137).

In addition to these NMR experiments, surface plasmon resonance (Bioacore) experiments can be used to determine the binding affinities (132). The ligand binding ability of GTs with and without donor or a number of potential inhibitors can be assessed with biotinylated substrates bound to a streptavidin-coated chip. The binding of donor and acceptor analogs to the blood group B enzyme GTB has also been analyzed by ITC combined with STD NMR titration (138). The results show the binding stoichiometry and binding affinity of one donor and acceptor molecule per protein, the thermodynamics, enthalpy, and entropy changes upon binding as well as the dissociation constants. The study also emphasizes that there can be differences in binding substrate analogs that should be considered.

Electrospray-mass spectrometry (ES-MS) has been used to determine the thermodynamics and affinities of substrates (139). Association constants were measured from the relative abundance of ions in the EI-MS spectra for GTA and GTB in aqueous solution with native donor and acceptor substrates as well as substrate analogs, products, and metal ion cofactor. To confirm the retaining mechanism of the enzymes, a mutant of blood group A (GTA) GalNAc-transferase in solution containing UDP-GalNAc and  $Mn^{2+}$  was studied, as well as the similar mutant of GTB. The catalytic Glu303 was replaced with Cys. After Trypsin digestion, a covalent intermediate could be trapped. Thus, tandem MS using collision-induced dissociation confirmed that Cys303 in both GTA and GTB enzymes was responsible for forming the glycosyl-enzyme intermediate. The formation of trisaccharide product can also be proven by MS (68). Thus, the double displacement mechanism of these retaining GTs was supported by MS.

### GLYCOSYLTRANSFERASE INHIBITORS

Detailed knowledge of GT structure and function is the basis for the development of effective GT inhibitors that may re-direct glycan biosynthesis. GT substrates usually bind through a small number of essential hydrogen bonds or hydrophobic interactions. Thus, not all of the substituents of the donor sugar, the base of the nucleotide, or the sugars of the acceptor are critically involved in binding (140). Therefore, modifications of these residues can result in competitive inhibitors that still bind in the catalytic site but do not support catalysis. Inhibitors can be ligands that bind well to the enzyme but cannot be released easily, or interfere with catalysis either as donor substrate analogs, acceptor substrate analogs, transition state analogs, compounds that prevent conformational changes necessary for catalysis, or compounds that distort protein conformation (74, 77, 103). Small structural modifications of compounds can have a dramatic effect on their inhibitory activity. Inhibitors have been designed that interfere with conformational changes and flexible loop movements that are essential events for substrate binding and catalysis (141). Sugar

donor analogs for ABO transferases (GTA and GTB), carrying a substituent at the uracil moiety, block the stacking of amino acids required for the proper folding of the internal loop. A heterocyclic compound inhibited GTB by interfering with its ability to bind metal ion, as well as donor and acceptor substrates. The compound does not appear to be structurally related to the acceptor but partly binds in the acceptor-binding site (142). A combination of crystal structure, Biacore, STD NMR, and docking experiments suggested that the inhibitor competes with binding of Fuc of the acceptor and the  $Mn^{2+}$  ion. Non-competitive inhibitors have also been described that potentially alter the structure of the enzyme leading to inactive proteins (77, 78). Modified nucleotide sugars are often recognized by GTs leading to transfer of unnatural sugars (143). Fluorescent groups modifying the base of the sugar-nucleotide can be useful as indicators of binding (144).

### CHEMOENZYMATIC SYNTHESIS OF SHARED EPITOPES

The preparation of bacterial GTs that lack a transmembrane domain is relatively inexpensive and they can be used in chemoenzymatic synthesis not only of bacterial glycoconjugates but also for mammalian oligosaccharides and glycoproteins with specific epitopes (Table 1). Due to the variety of bacterial enzymes with different specificities, a diverse range of glycan structures can be synthesized and processed for use as vaccine, to prepare antibodies for passive immunity, and for further studies of glycan functions. Examples include the synthesis of the complete blood group Forssman antigen GalNAc $\alpha$ 3GalNAc $\beta$ 3Gal $\alpha$ 4Gal $\beta$ 4Glc-p-nitrophenyl by  $\beta$ 3-GalNAc-transferase and  $\alpha$ 4-Gal-transferase from *Cj*, followed by  $\alpha$ 3-GalNAc-transferase from *Pm* (81, 140). The assembly of the entire blood group B determinant was achieved using GTs from *E. coli* O86 (145). Bacterial enzymes  $\alpha$ 4-Gal-transferase LgtC,  $\beta$ 3-Gal(NAc)-transferase LgtD, and  $\alpha$ 2-Fuc-transferase WbsJ (146) efficiently synthesized the tumor-associated epitope Globo-H-hexasaccharide (Fuc $\alpha$ 2-Gal $\beta$ 1-3GalNAc $\beta$ 1-3Gal $\alpha$ 1-4Gal $\beta$ 1-4Glc $\beta$ -benzyl).

Knowing the amino acids and mechanisms involved in substrate binding and catalysis, bacterial enzymes or new mutant enzymes can be engineered for use in the production of new natural or unnatural glycan structures, or for more efficient synthesis of known structures (147). For example, new donor specificity can be engineered by mutating only one or two critical amino acids that convert the function of the enzyme (140, 148).

Phosphorylases can also reversibly form glycosidic linkages (149). They can have similarity to either inverting or retaining glycohydrolases or to GT-B-folded retaining GT (CAZy). Sugar-1-P can be used as a substrate for phosphorylases to produce a wealth of different glycans with regio-selectivity. An interesting combination of chemical and enzymatic synthesis of the T antigen and the Gal $\beta$ 1-3GlcNAc linkage has been achieved using a combination of galactokinase (GalK) from *E. coli* that synthesizes Gal-1-P, and a Gal $\beta$ 1-3 HexNAc phosphorylase from *Bifobacterium infantis* that has promiscuous acceptor specificity (150). These enzymes could add Gal in the presence of ATP to synthetic GalNAc- and GlcNAc- substrates with various aglycone groups. The phosphorylase has multiple DxD motifs and an Asp-rich domain at the C terminus. The T antigen was also synthesized from sucrose and GlcNAc using phosphorylase from *Bifidobacterium longum*

(151) together with sucrose phosphorylase, UDP-Glc-hexose-1-P uridyltransferase and UDP-Glc 4-epimerase.

Glycosidases catalyze reversible reactions and have also been used to form sugar linkages using high concentrations of reactants. Glycosidases act with an inverting or retaining mechanism, utilizing a catalytically active nucleophile in the active site such as Asp or Glu. Mutant glycohydrolases that lack the catalytic base as well as hydrolase activity can be used to efficiently transfer a sugar to an acceptor substrate and synthesize specific linkages (glycosynthases). For example, large N-glycan-type oligosaccharides can be transferred to the GlcNAc residue linked to Asn of glycoproteins by a mutant endo-glucosaminidase that normally cleaves the chitobiose and releases the N-glycan. Thus, engineered glycosidases can be stereo-selective and very useful in achieving high yields of complex glycans (152).

## CONCLUDING REMARKS

It is astounding that proteins can be so different in amino acid sequence and yet become similar specific and effective catalysts for the transfer of sugars to proteins, lipids, and sugars and only have two major protein folds. Many possibilities are there for binding of donor and acceptor substrates but the transfer only involves inversion or retention of the anomeric configuration of the sugar. Mechanisms common to eukaryotes and bacteria include a change in protein conformation upon nucleotide sugar binding facilitating acceptor binding, and the action of a base (Glu, Asp, His) that deprotonates the hydroxyl to be glycosylated, which then becomes a nucleophile that results in cleavage of the sugar from the donor substrate. Bacterial and mammalian enzymes are often comparable in their action so that mammalian epitopes can easily be synthesized with bacterial enzymes, for example, to produce vaccines for cancer. However, the bacterial world is much more complex, variable, and challenging. Knowledge of bacterial GTs can lead to the synthesis of glycans, enzyme substrates, and antigens to study their biological functions and role in disease and to synthesize vaccines against specific pathogenic strains of bacteria. Bacteria may have evolved to express the GTs that make human-like structures, giving them a selective advantage. Most of the time, bacteria and human beings are symbiotic or compatible but once in a while, the mimicry of bacteria can lead to infection and serious consequences. We speculate that bacterial and mammalian enzymes with similar functions may have evolved in parallel, or may be derived from an ancient common ancestor. There may have been exchange of genes between these species (horizontal gene transfer), or GTs may be derived by convergent evolution. The many similar genes of a particular family may have been derived by gene duplication from an ancestral gene.

Further detailed understanding of GT structures and mechanisms helps to visualize how amino acids cooperate in forming a catalytic site, predict their functions, and to gain valuable insight into the syntheses of complex glycans in mammals and in our close neighbors, bacteria. Both, the bacterial world and human beings can benefit from this relationship. In addition, inhibitors of bacterial GTs may help to eliminate virulence factors, and this is an urgently needed goal in light of growing antibiotic resistance.

## ACKNOWLEDGMENTS

The work by the author has been supported by the Canadian Institutes of Health Research and the Natural Science and Engineering Research Council of Canada.

## REFERENCES

- Petit D, Teppa RE, Petit JM, Harduin-Lepers A. A practical approach to reconstruct evolutionary history of animal sialyltransferases and gain insights into the sequence-function relationships of Golgi-glycosyltransferases. *Methods Mol Biol* (2013) **1022**:73–97. doi:10.1007/978-1-62703-465-4\_7
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* (2014) **42**(Database issue):D490–5. doi:10.1093/nar/gkt1178
- Baycin Hizal D, Wolozny D, Colao J, Jacobson E, Tian Y, Krag SS, et al. Glyco-proteomic and glycomic databases. *Clin Proteomics* (2014) **11**:15. doi:10.1186/1559-0275-11-15
- Yen TY, Dutta SM, Litsakos-Cheung C, Corona AA, Timpe LC, Macher BA. Overcoming challenges and opening new opportunities in glycoproteomics. *Biomolecules* (2013) **3**:270–86. doi:10.3390/biom3020270
- Lundborg M, Modhukur V, Widmalm G. Glycosyltransferase functions of *E. coli* O-antigens. *Glycobiology* (2010) **20**:366–8. doi:10.1093/glycob/cwp185
- Varki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* (1993) **3**:97–130. doi:10.1093/glycob/3.2.97
- Aebi M. N-linked protein glycosylation in the ER. *Biochim Biophys Acta* (2013) **1833**:2430–7. doi:10.1016/j.bbamcr.2013.04.001
- Brockhausen I. The role of galactosyltransferases in cell surface functions and in the immune system. *Drug News Perspect* (2006) **19**:401–9. doi:10.1358/dnp.2006.tg.j.t02t4g1
- Taniguchi N, Korekane H. Branched N-glycans and their implications for cell adhesion, signaling and clinical applications for cancer biomarkers and in therapeutics. *BMB Rep* (2011) **44**:772–81. doi:10.5483/BMBRep.2011.44.12.772
- Boscher C, Dennis JW, Nabi IR. Glycosylation, galectins and cellular signaling. *Curr Opin Cell Biol* (2011) **23**:383–92. doi:10.1016/j.ccb.2011.05.001
- Dalziel M, Crispin M, Scanlan CN, Zitzmann N, Dwek RA. Emerging principles for the therapeutic exploitation of glycosylation. *Science* (2014) **343**:1235681. doi:10.1126/science.1235681
- Brockhausen I, Schutzbach J, Kuhns W. Glycoproteins and their relationship to human disease. *Acta Anat* (1998) **161**:36–78. doi:10.1159/000046450
- Nothaft H, Szymanski CM. Bacterial protein N-glycosylation: new perspectives and applications. *J Biol Chem* (2013) **288**:6912–20. doi:10.1074/jbc.R112.417857
- Brockhausen I. Pathways of O-glycan biosynthesis in cancer cells. *Biochim Biophys Acta* (1999) **1473**:67–95. doi:10.1016/S0304-4165(99)00170-1
- Brockhausen I. Biosynthesis of complex mucin-type O-glycans. In: Mander L, Lui H-W, Wang PG editors. *Comprehensive Natural Products II Chemistry and Biology: Carbohydrates, Nucleosides and Nucleic Acids*. (Vol. 6), Oxford: Elsevier (2010). p. 315–50. doi:10.1016/B978-008045382-8.00643-2
- Tarp MA, Clausen H. Mucin-type O-glycosylation and its potential use in drug and vaccine development. *Biochim Biophys Acta* (2008) **1780**:546–63. doi:10.1016/j.bbagen.2007.09.010
- Fritz TA, Hurley JH, Trinh LB, Shiloach J, Tabak LA. The beginnings of mucin biosynthesis: the crystal structure of UDP-GalNAc: polypeptide alpha-N-acetylgalactosaminyltransferase-T1. *Proc Natl Acad Sci U S A* (2004) **101**:15307–12. doi:10.1073/pnas.0405657101
- Fritz TA, Raman J, Tabak LA. Dynamic association between the catalytic and lectin domains of human UDP-GalNAc: polypeptide-N-acetylgalactosaminyltransferase-2. *J Biol Chem* (2006) **281**:8613–9. doi:10.1074/jbc.M513590200
- Ünlügil UM, Rini JM. Glycosyltransferase structure and mechanism. *Curr Opin Struct Biol* (2000) **10**:510–7. doi:10.1016/S0959-440X(00)00124-X
- Pak JE, Arnoux P, Zhou S, Sivarajah P, Satkunarajah M, Xing X, et al. X-ray crystal structure of leukocyte type core 2 beta1,6-N-acetylglucosaminyltransferase. Evidence for a convergence of metal ion-independent glycosyltransferase mechanism. *J Biol Chem* (2006) **281**:26693–701. doi:10.1074/jbc.M603534200
- Pak JE, Satkunarajah M, Seetharaman J, Rini JM. Structural and mechanistic characterization of leukocyte-type core 2 beta1,6-N-acetylglucosaminyltransferase: a metal-ion-independent GT-A glycosyltransferase. *J Mol Biol* (2011) **414**:798–811. doi:10.1016/j.jmb.2011.10.039

22. Ramakrishnan B, Balaji PV, Qasba PK. Crystal structure of beta1,4-galactosyltransferase complex with UDP-gal reveals an oligosaccharide acceptor binding site. *J Mol Biol* (2002) **318**:491–502. doi:10.1016/S0022-2836(02)00020-7
23. Ramakrishnan B, Boeggeman E, Qasba PK. Mutation of arginine 228 to lysine enhances the glucosyltransferase activity of bovine beta-1,4-galactosyltransferase I. *Biochemistry* (2005) **44**:3202–10. doi:10.1021/bi0479454
24. Tsutsui Y, Ramakrishnan B, Qasba PK. Crystal structures of  $\beta$ -1,4-galactosyltransferase 7 enzyme reveal conformational changes and substrate binding. *J Biol Chem* (2013) **288**:31963–70. doi:10.1074/jbc.M113.509984
25. Sun HY, Lin SW, Ko TP, Pan JF, Liu CL, Lin C, et al. Structure and mechanism of *Helicobacter pylori* fucosyltransferase. A basis for lipopolysaccharide variation and inhibitor design. *J Biol Chem* (2007) **282**:9973–82. doi:10.1074/jbc.M610285200
26. Ihara H, Ikeda Y, Toma S, Wang X, Suzuki T, Gu J, et al. Crystal structure of mammalian alpha1,6-fucosyltransferase, FUT8. *Glycobiology* (2007) **17**:455–66. doi:10.1093/glycob/cwl079
27. Brzezinski K, Stepkowski T, Panjikar S, Bujacz G, Jaskolski M. High-resolution structure of NodZ fucosyltransferase involved in the biosynthesis of the nodulation factor. *Acta Biochim Pol* (2007) **54**:537–49.
28. Rakic B, Rao FV, Freimann K, Wakarchuk W, Strynadka NCJ, Withers SG. Structure-based mutagenic analysis of mechanism and substrate specificity in mammalian glycosyltransferases: porcine ST3Gal-I. *Glycobiology* (2013) **23**:536–45. doi:10.1093/glycob/cwt001
29. Iwatani T, Okino N, Sakakura M, Kajiura H, Takakura Y, Kimura M, et al. Crystal structure of alpha/beta-galactoside alpha2,3-sialyltransferase from a luminous marine bacterium, *Photobacterium phosphoreum*. *FEBS Lett* (2009) **583**:2083–7. doi:10.1016/j.febslet.2009.05.032
30. Chiu CPC, Lairson LL, Gilbert M, Wakarchuk WW, Withers SG, Strynadka NCJ. Structural analysis of the R-2,3-sialyltransferase Cst-I from *Campylobacter jejuni* in Apo and substrate-analogue bound forms. *Biochemistry* (2007) **46**:7196–204. doi:10.1021/bi602543d
31. Lin LYC, Rakic B, Chiu CPC, Lameignere E, Wakarchuk WW, Withers SG, et al. Structure and mechanism of the lipooligosaccharide sialyltransferase from *Neisseria meningitidis*. *J Biol Chem* (2011) **286**:37237–48. doi:10.1074/jbc.M111.249920
32. Meng L, Forouhar F, Thieker D, Gao Z, Ramiah A, Moniz H, et al. Enzymatic basis for N-glycan sialylation. Structure of rat alpha2,6-sialyltransferase (st6gal1) reveals conserved and unique features for glycan sialylation. *J Biol Chem* (2013) **288**:34680–98. doi:10.1074/jbc.M113.519041
33. Kuhn B, Benz J, Greif M, Engel AM, Sobek H, Rudolph MG. The structure of human alpha-2,6-sialyltransferase reveals the binding mode of complex glycans. *Acta Crystallogr D Biol Crystallogr* (2013) **D69**:1826–38. doi:10.1107/S0907444913015412
34. Ni L, Chokhawala HA, Cao H, Henning R, Ng L, Huang S, et al. Crystal structures of *Pasteurella multocida* sialyltransferase complexes with acceptor and donor analogues reveal substrate binding sites and catalytic mechanism. *Biochemistry* (2007) **46**:6288–98. doi:10.1021/bi700346w
35. Kakuta Y, Okino N, Kajiura H, Ichikawa M, Takakura Y, Ito M, et al. Crystal structure of *Vibrionaceae Photobacterium* sp. JT-ISH-224  $\alpha$ 2,6-sialyltransferase in a ternary complex with donor product CMP and acceptor substrate lactose: catalytic mechanism and substrate recognition. *Glycobiology* (2008) **18**:66–73. doi:10.1093/glycob/cwm119
36. Lee HJ, Lairson LL, Rich JR, Lameignere E, Wakarchuk WW, Withers SG, et al. Structural and kinetic analysis of substrate binding to the sialyltransferase Cst-II from *Campylobacter jejuni*. *J Biol Chem* (2011) **286**:35922–32. doi:10.1074/jbc.M111.261172
37. Kim DU, Yoo JH, Lee YJ, Kim KS, Cho HS. Structural analysis of sialyltransferase PM0188 from *Pasteurella multocida* complexed with donor analogue and acceptor sugar. *BMB Rep* (2008) **41**:48–54. doi:10.5483/BMBRep.2008.41.1.048
38. Kubota T, Shiba T, Sugioka S, Furukawa S, Sawaki H, Kato R, et al. Structural basis of carbohydrate transfer activity by human UDP-GalNAc: polypeptide alpha-N-acetylgalactosaminyltransferase (pp-GalNAc-T10). *J Mol Biol* (2006) **359**:708–27. doi:10.1016/j.jmb.2006.03.061
39. Patenaude SI, Seto NO, Borisova SN, Szpacenko A, Marcus SL, Palcic MM, et al. The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nat Struct Biol* (2002) **9**:685–90. doi:10.1038/nsb832
40. Yazer MH, Palcic MM. The importance of disordered loops in ABO glycosyltransferases. *Transfus Med Rev* (2005) **19**:210–6. doi:10.1016/j.tmr.2005.02.008
41. Gastinel LN, Bignon C, Misra AK, Hindsgaul O, Shaper JH, Joziase DH. Bovine alpha1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases. *EMBO J* (2001) **20**:638–49. doi:10.1093/emboj/20.4.638
42. Boix E, Zhang Y, Swaminathan GJ, Brew K, Acharya KR. Structural basis of ordered binding of donor and acceptor substrates to the retaining glycosyltransferase, alpha-1,3-galactosyltransferase. *J Biol Chem* (2002) **277**:28310–8. doi:10.1074/jbc.M202631200
43. Alfaro JA, Zheng RB, Persson M, Letts JA, Polakowski R, Bai Y, et al. ABO(H) blood group A and B glycosyltransferases recognize substrate via specific conformational changes. *J Biol Chem* (2008) **283**:10097–108. doi:10.1074/jbc.M708669200
44. Thiagarajan N, Pham TTK, Stinson B, Sundriyal A, Tumbale P, Lizotte-Waniewski M, et al. Structure of a metal-independent bacterial glycosyltransferase that catalyzes the synthesis of histo-blood group A antigen. *Sci Rep* (2012) **2**:940. doi:10.1038/srep00940
45. Persson K, Ly HD, Dieckmann M, Wakarchuk WW, Withers SG, Strynadka NCJ. Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat Struct Biol* (2001) **8**:166–75. doi:10.1038/84168
46. Musumeci MA, Hug I, Scott NE, Ielmini MV, Foster LJ, Wang PG, et al. In vitro activity of *Neisseria meningitidis* PglL O-oligosaccharyltransferase with diverse synthetic lipid donors and a UDP-activated sugar. *J Biol Chem* (2013) **288**:10578–87. doi:10.1074/jbc.M112.432815
47. Musumeci MA, Faridmoayer A, Watanabe Y, Feldman MF. Evaluating the role of selected amino acids in bacterial O-oligosaccharyltransferases by in vivo, in vitro and limited proteolysis assays. *Glycobiology* (2014) **24**:39–50. doi:10.1093/glycob/cwt087
48. Whitfield C, Trent MS. Biosynthesis and export of bacterial lipopolysaccharides. *Annu Rev Biochem* (2014) **83**:99–128. doi:10.1146/annurev-biochem-060713-035600
49. Rush JS, Alaimo C, Robbani R, Wacker M, Waechter CJ. A novel epimerase that converts GlcNAc-P-P-undecaprenol to GalNAc-P-P-undecaprenol in *Escherichia coli* O157. *J Biol Chem* (2010) **285**:1671–80. doi:10.1074/jbc.M109.061630
50. Freinkman E, Chng SS, Kahne D. The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. *Proc Natl Acad Sci USA* (2011) **108**:2486–91. doi:10.1073/pnas.1015617108
51. Greenfield LK, Whitfield C. Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr Res* (2012) **356**:12–24. doi:10.1016/j.carres.2012.02.027
52. Coutinho PM, Deleury E, Davies GJ, Henrissat B. An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* (2003) **328**:307–17. doi:10.1016/S0022-2836(03)00307-3
53. Li T, Simonds L, Kovrigin EL, Noel KD. In vitro biosynthesis and chemical identification of UDP-N-acetyl-D-quinovosamine (UDP-d-QuiNAc). *J Biol Chem* (2014) **289**:18110–20. doi:10.1074/jbc.M114.555862
54. Wang S, Tanaka H, Hindsgaul O, Lam JS, Brockhausen I. A convenient synthesis of GDP-D-rhamnose: the donor substrate for D-rhamnosyltransferase WbpZ from *Pseudomonas aeruginosa* PAO1. *Bioorg Med Chem Lett* (2013) **23**:3491–5. doi:10.1016/j.bmcl.2013.04.051
55. Montoya-Peleaz PJ, Riley JG, Szarek WA, Valvano MA, Schutzbach JS, Brockhausen I. Identification of a UDP-Gal: GlcNAc-R galactosyltransferase activity in *Escherichia coli* VW187. *Bioorg Med Chem Lett* (2005) **15**:1205–11. doi:10.1016/j.bmcl.2004.11.077
56. Albesa-Jové D, Giganti D, Jackson M, Alalzari P, Guerin ME. Structure-function relationships of membrane-associated GT-B glycosyltransferases. *Glycobiology* (2014) **24**:108–24. doi:10.1093/glycob/cwt101
57. Breton C, Fournel-Gigleux S, Palcic MM. Recent structures, evolution and mechanisms of glycosyltransferases. *Curr Opin Struct Biol* (2012) **22**:540–9. doi:10.1016/j.sbi.2012.06.007
58. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* (2008) **77**:521–55. doi:10.1146/annurev.biochem.76.061005.092322

59. Namdjou DJ, Chen HM, Vinogradov E, Brochu D, Withers SG, Wakarchuk WW. A beta-1,4-galactosyltransferase from *Helicobacter pylori* is an efficient and versatile biocatalyst displaying a novel activity for thioglycoside synthesis. *Chembiochem* (2008) **9**:1632–40. doi:10.1002/cbic.200700775
60. Breton C, Oriol R, Imberty A. Conserved structural features in eukaryotic and prokaryotic fucosyltransferases. *Glycobiology* (1998) **8**:87–94. doi:10.1093/glycob/8.1.87
61. Narimatsu H. Construction of a human glycogene library and comprehensive functional analysis. *Glycoconj J* (2004) **21**:17–24. doi:10.1023/B:GLYC.0000043742.99482.01
62. Oriol R, Mollicone R, Cailleau A, Balanzino L, Breton C. Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. *Glycobiology* (1999) **9**:323–34. doi:10.1093/glycob/9.4.323
63. Martinez-Duncker I, Mollicone R, Candelier JJ, Breton C, Oriol R. A new superfamily of protein-O-fucosyltransferases, alpha2-fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and identification of conserved peptide motifs. *Glycobiology* (2003) **13**:1C–5C. doi:10.1093/glycob/cwg113
64. Li M, Liu XW, Shao J, Shen J, Jia Q, Yi W, et al. Characterization of a novel alpha1,2-fucosyltransferase of *Escherichia coli* O128: b12 and functional investigation of its common motif. *Biochemistry* (2008) **47**:378–87. doi:10.1021/bi701345v
65. Patel RY, Balaji PV. Identification of linkage-specific sequence motifs in sialyltransferases. *Glycobiology* (2006) **16**:108–16. doi:10.1093/glycob/cwj046
66. Freiburger F, Claus H, Gunzel A, Oltmann-Norden I, Vionnet J, Muhlenhoff M, et al. Biochemical characterization of a *Neisseria meningitidis* polysialyltransferase reveals novel functional motifs in bacterial sialyltransferases. *Mol Microbiol* (2007) **65**:1258–75. doi:10.1111/j.1365-2958.2007.05862.x
67. Nakata D, Zhang L, Troy IIEA. Molecular basis for polysialylation: a novel polybasic polysialyltransferase domain (PSTD) of 32 amino acids unique to the alpha2,8-polysialyltransferases is essential for polysialylation. *Glycoconj J* (2006) **23**:423–36. doi:10.1007/s10719-006-6356-5
68. Soya N, Fang Y, Palcic MM, Klassen JS. Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases. *Glycobiology* (2011) **21**:547–52. doi:10.1093/glycob/cwq190
69. Amado M, Almeida R, Schwientek T, Clausen H. Identification and characterization of large galactosyltransferase gene families: galactosyltransferases for all functions. *Biochim Biophys Acta* (1999) **1473**:35–53. doi:10.1016/S0304-4165(99)00168-3
70. Ramakrishnan B, Boeggeman E, Qasba PK. Effect of the Met344His mutation on the conformational dynamics of bovine beta-1,4-galactosyltransferase: crystal structure of the Met344His mutant in complex with chitobiose. *Biochemistry* (2004) **43**:12513–22. doi:10.1021/bi049007
71. Ramasamy V, Ramakrishnan B, Boeggeman E, Ratner DM, Seeberger PH, Qasba PK. Oligosaccharide preferences of beta1,4-galactosyltransferase-I: crystal structures of Met340His mutant of human beta1,4-galactosyltransferase-I with a pentasaccharide and trisaccharides of the N-glycan moiety. *J Mol Biol* (2005) **353**:53–67. doi:10.1016/j.jmb.2005.07.050
72. Qasba PK, Ramakrishnan B, Boeggeman E. Structure and function of beta-1,4-galactosyltransferase. *Curr Drug Targets* (2008) **9**:292–309. doi:10.2174/138945008783954943
73. Xu C, Liu B, Hu B, Han Y, Feng L, Allingham J, et al. Biochemical characterization of UDP-Gal: GlcNAc-pyrophosphate-lipid beta1,4-galactosyltransferase WfeD, a new enzyme from *Shigella boydii* type 14 that catalyzes the second step in O-antigen repeating-unit synthesis. *J Bacteriol* (2011) **193**:449–59. doi:10.1128/JB.00737-10
74. Gao Y, Lazar C, Szarek WA, Brockhausen I. Specificity of beta4galactosyltransferase inhibitor 2-naphthyl 2-butanamido-2-deoxy-1-thio-beta-D-glucopyranoside. *Glycoconj J* (2010) **27**:673–84. doi:10.1007/s10719-010-9312-3
75. Riley JG, Menggad M, Montoya-Peleaz P, Szarek WA, Marolda CL, Valvano MA, et al. The wbbD gene of *E. coli* strain VW187 (O7: K1) encodes a UDP-Gal: GlcNAc(alpha)-pyrophosphate-R (beta)1,3-galactosyltransferase involved in the biosynthesis of O7-specific lipopolysaccharide. *Glycobiology* (2005) **15**:605–13. doi:10.1093/glycob/cwi038
76. Ju T, Cummings RD. A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta3-galactosyltransferase. *Proc Natl Acad Sci U S A* (2002) **99**:16613–8. doi:10.1073/pnas.262438199
77. Gao Y, Aryal RP, Ju T, Cummings RD, Gahlay G, Jarvis DL, et al. Acceptor specificities and selective inhibition of recombinant human Gal- and GlcNAc-transferases that synthesize O-glycan core structures 1, 2, 3 and 4 of O-glycans. *Biochim Biophys Acta* (2013) **1830**:4274–81. doi:10.1016/j.bbagen.2013.04.001
78. Wang S, Czuchry D, Liu B, Vinnikova AN, Gao Y, Vlahakis JZ, et al. Characterization of Two UDP-Gal: GalNAc-diphosphate-lipid beta1,3-galactosyltransferases WbwC from *Escherichia coli* serotypes O104 and O5. *J Bacteriol* (2014) **196**(17):3122–33. doi:10.1128/JB.01698-14
79. Steffensen R, Carlier K, Wiels J, Levery SB, Stroud M, Cedergren B, et al. Cloning and expression of the histo-blood group Pk UDP-galactose: Galbeta-4G1cbeta1-cer alpha1, 4-galactosyltransferase. Molecular genetic basis of the p phenotype. *J Biol Chem* (2000) **275**:16723–9. doi:10.1074/jbc.M000728200
80. Suzuki N, Yamamoto K. Molecular cloning of pigeon UDP-galactose: beta-D-galactoside alpha1,4-galactosyltransferase and UDP-galactose: beta-D-galactoside beta1,4-galactosyltransferase, two novel enzymes catalyzing the formation of Gal alpha1-4Gal beta1-4Gal beta1-4GlcNAc sequence. *J Biol Chem* (2010) **285**:5178–87. doi:10.1074/jbc.M109.018663
81. Houliston RS, Bernatchez S, Karwaski MF, Mandrell RE, Jarrell HC, Wakarchuk WW, et al. Complete chemoenzymatic synthesis of the Forssman antigen using novel glycosyltransferases identified in *Campylobacter jejuni* and *Pasteurella multocida*. *Glycobiology* (2009) **19**:153–9. doi:10.1093/glycob/cwn117
82. Chan PH, Cheung AH, Okon M, Chen HM, Withers SG, McIntosh LP. Investigating the structural dynamics of alpha-1,4-galactosyltransferase C from *Neisseria meningitidis* by nuclear magnetic resonance spectroscopy. *Biochemistry* (2013) **52**:320–32. doi:10.1021/bi301317d
83. Yi W, Perali RS, Eguchi H, Motari E, Woodward R, Wang PG. Characterization of a bacterial beta1,3-galactosyltransferase with application in the synthesis of tumor-associated t-antigen mimics. *Biochemistry* (2008) **47**:1241–8. doi:10.1021/bi7020712
84. Brockhausen I, Hu B, Liu B, Lau K, Szarek WA, Wang L, et al. Characterization of two beta1,3-glucosyltransferases from the *Escherichia coli* serotypes O56 and O152. *J Bacteriol* (2008) **190**:4922–32. doi:10.1128/JB.00160-08
85. Gao Y, Liu B, Strum S, Schutzbach J, Druzhinina TN, Utkina NS, et al. Biochemical characterization of WbdN, a beta1,3-glucosyltransferase involved in O-antigen synthesis in enterohemorrhagic *Escherichia coli* O157. *Glycobiology* (2012) **22**:1092–102. doi:10.1093/glycob/cws081
86. Bernatchez S, Gilbert M, Blanchard MC, Karwaski MF, Li J, DeFrees S, et al. Variants of the beta1,3-galactosyltransferase CgtB from the bacterium *Campylobacter jejuni* have distinct acceptor specificities. *Glycobiology* (2007) **17**:1333–43. doi:10.1093/glycob/cwm090
87. Liu XW, Xia C, Li L, Guan WY, Pettit N, Zhang HC, et al. Characterization and synthetic application of a novel beta1,3-galactosyltransferase from *Escherichia coli* O55: H7. *Bioorg Med Chem* (2009) **17**:4910–5. doi:10.1016/j.bmc.2009.06.005
88. Seko A, Yamashita K. Characterization of a novel galactose beta1,3-N-acetylglucosaminyltransferase (beta3Gn-T8): the complex formation of beta3Gn-T2 and beta3Gn-T8 enhances enzymatic activity. *Glycobiology* (2005) **15**:943–51. doi:10.1093/glycob/cwi082
89. Togayachi A, Kozono Y, Kuno A, Ohkura T, Sato T, Hirabayashi J, et al. Beta3GnT2 (B3GNT2), a major polyactosamine synthase: analysis of B3GNT2-deficient mice. *Methods Enzymol* (2010) **479**:185–204. doi:10.1016/S0076-6879(10)79011-X
90. Yang X, Qin W, Lehotay M, Toki D, Dennis P, Schutzbach JS, et al. Soluble human core 2 beta6-N-acetylglucosaminyltransferase C2GnT1 requires its conserved cysteine residues for full activity. *Biochim Biophys Acta* (2003) **1648**:62–74. doi:10.1016/S1570-9639(03)00105-5
91. Toki D, Sarkar M, Yip B, Reck F, Joiasse D, Fukuda M, et al. Expression of stable human O-glycan core 2 beta-1,6-N-acetylglucosaminyltransferase in Sf9 insect cells. *Biochem J* (1997) **325**:63–9.
92. Peng W, Pranskevich J, Nycholat C, Gilbert M, Wakarchuk W, Paulson JC, et al. *Helicobacter pylori* beta1,3-N-acetylglucosaminyltransferase for versatile synthesis of type 1 and type 2 poly-LacNAcs on N-linked, O-linked and I-antigen glycans. *Glycobiology* (2012) **22**:1453–64. doi:10.1093/glycob/cws101
93. Guan W, Ban L, Cai L, Li L, Chen W, Liu X, et al. Combining carbochips and mass spectrometry to study the donor specificity for the *Neisseria meningitidis* beta1,3-N-acetylglucosaminyltransferase LgtA. *Bioorg Med Chem Lett* (2011) **21**:5025–8. doi:10.1016/j.bmcl.2011.04.100
94. Wang G, Boulton PG, Chan NW, Palcic MM, Taylor DE. Novel *Helicobacter pylori* alpha1,2-fucosyltransferase, a key enzyme in the synthesis of Lewis antigens. *Microbiology* (1999) **145**(Pt 11):3245–53.

95. Li M, Shen J, Liu X, Shao J, Yi W, Chow CS, et al. Identification of a new  $\alpha$ 1,2-fucosyltransferase involved in O-antigen biosynthesis of *Escherichia coli* O86: B7 and formation of H-type 3 blood group antigen. *Biochemistry* (2008) **47**:11590–7. doi:10.1021/bi801067s
96. Pettit N, Styslingera T, Meia Z, Hana W, Zhaoa G, Wang PG. Characterization of WbiQ: an  $\alpha$ 1,2-fucosyltransferase from *Escherichia coli* O127: K63(B8), and synthesis of H-type 3 blood group antigen. *Biochem Biophys Res Commun* (2010) **402**:190–5. doi:10.1016/j.bbrc.2010.08.087
97. Sperandio M, Gleissner CA, Ley K. Glycosylation in immune cell trafficking. *Immunol Rev* (2009) **230**:97–113. doi:10.1111/j.1600-065X.2009.00795.x
98. Mollicone R, Moore SE, Bovin N, Garcia-Rosasco M, Candelier JJ, Martinez-Duncker I, et al. Activity, splice variants, conserved peptide motifs, and phylogeny of two new  $\alpha$ 1,3-fucosyltransferase families (FUT10 and FUT11). *J Biol Chem* (2009) **284**:4723–38. doi:10.1074/jbc.M809312200
99. Dupuy F, Germot A, Julien R, Maftah A. Structure/function study of Lewis  $\alpha$ 3- and  $\alpha$ 3/4-fucosyltransferases: the  $\alpha$ 1,4 fucosylation requires an aromatic residue in the acceptor-binding domain. *Glycobiology* (2004) **14**:347–56. doi:10.1093/glycob/cwh053
100. Ma B, Lau LH, Palcic MM, Hazes B, Taylor DEA. Single aromatic amino acid at the carboxyl terminus of *Helicobacter pylori*  $\alpha$ 1,3/4 fucosyltransferase determines substrate specificity. *J Biol Chem* (2005) **280**:36848–56. doi:10.1074/jbc.M504415200
101. Lin SW, Yuan TM, Li JR, Lin CH. Carboxyl terminus of *Helicobacter pylori*  $\alpha$ 1,3-fucosyltransferase determines the structure and stability. *Biochemistry* (2006) **45**:8108–16. doi:10.1021/bi0601297
102. Zhang L, Lau K, Cheng J, Yu H, Li Y, Sugianto G, et al. *Helicobacter hepaticus* Hh0072 gene encodes a novel  $\alpha$ 1-3-fucosyltransferase belonging to CAZY GT11 family. *Glycobiology* (2010) **20**:1077–88. doi:10.1093/glycob/cwq068
103. Bastida A, Fernandez-Mayoralas A, Arrayas RG, Iradier F, Carretero JC, Garcia-Junceda E. Heterologous over-expression of  $\alpha$ -1,6-fucosyltransferase from *Rhizobium* sp.: application to the synthesis of the trisaccharide  $\beta$ -D-GlcNAc(1-4)-[ $\alpha$ -1-Fuc(1-6)]-D-GlcNAc, study of the acceptor specificity and evaluation of polyhydroxylated indolizidines as inhibitors. *Chemistry* (2001) **7**:2390–7. doi:10.1002/1521-3765(20010601)7:11<2390::AID-CHEM23900>3.0.CO;2-O
104. Brew K, Tumbale P, Acharya KR. Family 6 glycosyltransferases in vertebrates and bacteria: inactivation and horizontal gene transfer may enhance mutualism between vertebrates and bacteria. *J Biol Chem* (2010) **285**:37121–7. doi:10.1074/jbc.R110.176248
105. Yi W, Shen J, Zhou G, Li J, Wang PG. Bacterial homologue of human blood group A transferase. *J Am Chem Soc* (2008) **130**:14420–1. doi:10.1021/ja805844y
106. Jamaluddin H, Tumbale P, Ferns TA, Thiyagarajan N, Brew K, Acharya KR. Crystal structure of  $\alpha$ -1,3-galactosyltransferase ( $\alpha$ 3GT) in a complex with p-nitrophenyl- $\beta$ -galactoside (pNPGal). *Biochem Biophys Res Commun* (2009) **385**:601–4. doi:10.1016/j.bbrc.2009.05.111
107. Harduin-Lepers A, Mollicone R, Delannoy P, Oriol R. The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach. *Glycobiology* (2005) **15**:805–17. doi:10.1093/glycob/cwi063
108. Troy FA. Polysialylation: from bacteria to brains. *Glycobiology* (1992) **2**:5–23. doi:10.1093/glycob/2.1.5
109. Amith SR, Jayanth P, Franchuk S, Finlay T, Seyrantepe V, Beyaert R, et al. Neu1 desialylation of sialyl  $\alpha$ -2,3-linked  $\beta$ -galactosyl residues of TOLL-like receptor 4 is essential for receptor activation and cellular signaling. *Cell Signal* (2010) **22**:314–24. doi:10.1016/j.cellsig.2009.09.038
110. Burchell J, Poulson R, Hanby A, Whitehouse C, Cooper L, Clausen H, et al. An  $\alpha$ 2,3 sialyltransferase (ST3Gal I) is elevated in primary breast carcinomas. *Glycobiology* (1999) **9**:1307–11. doi:10.1093/glycob/9.12.1307
111. Picco G, Julien S, Brockhausen I, Beaton R, Antonopoulos A, Haslam S, et al. Over-expression of ST3Gal-I promotes mammary tumorigenesis. *Glycobiology* (2010) **20**:1241–50. doi:10.1093/glycob/cwq085
112. Whitehouse C, Burchell J, Gschmeissner S, Brockhausen I, Lloyd KO, Taylor-Papadimitriou J. A transfected sialyltransferase that is elevated in breast cancer and localizes to the medial/trans-Golgi apparatus inhibits the development of core-2-based O-glycans. *J Cell Biol* (1997) **137**:1229–41. doi:10.1083/jcb.137.6.1229
113. Dall'Olio F, Malagolini N, Serafini-Cessi F. The expression of soluble and cell-bound  $\alpha$ 2,6 sialyltransferase in human colonic carcinoma CaCo-2 cells correlates with the degree of enterocytic differentiation. *Biochem Biophys Res Commun* (1992) **184**:1405–10. doi:10.1016/S0006-291X(05)80039-7
114. Taatjes DJ, Roth J, Weinstein J, Paulson JC. Post-Golgi apparatus localization and regional expression of rat intestinal sialyltransferase detected by immunoelectron microscopy with polypeptide epitope-purified antibody. *J Biol Chem* (1988) **263**:6302–9.
115. Lehoux S, Groux-Degroote S, Cazet A, Dhaenens CM, Maurage CA, Caillet-Boudin ML, et al. Transcriptional regulation of the human ST6GAL2 gene in cerebral cortex and neuronal cells. *Glycoconj J* (2010) **27**:99–114. doi:10.1007/s10719-009-9260-y
116. Houlston RS, Vinogradov E, Dzieciatkowska M, Li J, St. Michael F, Karwaski MF, et al. Lipooligosaccharide of *Campylobacter jejuni*. Similarity with multiple types of mammalian glycans beyond gangliosides. *J Biol Chem* (2011) **286**:12361–70. doi:10.1074/jbc.M110.181750
117. Yuki N, Odaka M. Ganglioside mimicry as a cause of Guillain-Barré syndrome. *Curr Opin Neurol* (2005) **18**:557–61. doi:10.1097/01.wco.0000174604.42272.2d
118. Wakerley BR, Yuki N. Infectious and noninfectious triggers in Guillain-Barré syndrome. *Expert Rev Clin Immunol* (2013) **9**:627–39. doi:10.1586/1744666X.2013.811119
119. Chiu CPC, Watts AG, Lairson LL, Gilbert M, Lim D, Wakarchuk WW, et al. Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog. *Nat Struct Mol Biol* (2004) **11**:163–70. doi:10.1038/nsmb720
120. Chan PHW, Lairson LL, Lee HJ, Wakarchuk WW, Strynadka NCJ, Withers SG, et al. NMR spectroscopic characterization of the sialyltransferase CstII from *Campylobacter jejuni*: histidine 188 is the general base. *Biochemistry* (2009) **48**:11220–30. doi:10.1021/bi901606n
121. Schmöler K, Ribitsch D, Czabany T, Luley-Goedl C, Kokot D, Lyskowski A, et al. Characterization of a multifunctional  $\alpha$ 2,3-sialyltransferase from *Pasteurella dagmatis*. *Glycobiology* (2013) **23**:1293–304. doi:10.1093/glycob/cwt066
122. Drake PM, Nathan JK, Stock CM, Chang PV, Muench MO, Nakata D, et al. Polysialic acid, a glycan with highly restricted expression, is found on human and murine leukocytes and modulates immune responses. *J Immunol* (2008) **181**:6850–8. doi:10.4049/jimmunol.181.10.6850
123. Nakayama J, Angata K, Ong E, Katsuyama T, Fukuda M. Polysialic acid, a unique glycan that is developmentally regulated by two polysialyltransferases, PST and STX, in the central nervous system: from biosynthesis to function. *Pathol Int* (1998) **48**:665–77. doi:10.1111/j.1440-1827.1998.tb03967.x
124. Rollenhagen M, Buettner FF, Reissmann M, Jirmo AC, Grove M, Behrens GM, et al. Polysialic acid on neuropilin-2 is exclusively synthesized by the polysialyltransferase ST8SiaIV and attached to mucin-type o-glycans located between the b2 and c domain. *J Biol Chem* (2013) **288**:22880–92. doi:10.1074/jbc.M113.463927
125. Cress BF, Englaender JA, He W, Kasper D, Linhardt RJ, Koffas MA. Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS Microbiol Rev* (2014) **38**:660–97. doi:10.1111/1574-6976.12056
126. Zapater JL, Colley KJ. Sequences prior to conserved catalytic motifs of polysialyltransferase ST8Sia IV are required for substrate recognition. *J Biol Chem* (2012) **287**:6441–53. doi:10.1074/jbc.M111.322024
127. Keys TG, Fuchs HL, Ehrig J, Alves J, Freiburger F, Gerardy-Schahn R. Engineering the product profile of a polysialyltransferase. *Nat Chem Biol* (2014) **10**:437–42. doi:10.1038/nchembio.1501.13
128. Lindhout T, Bainbridge CR, Costain WJ, Gilbert M, Wakarchuk WW. Biochemical characterization of a polysialyltransferase from *Mannheimia haemolytica* A2 and comparison to other bacterial polysialyltransferases. *PLoS ONE* (2013) **8**:e69888. doi:10.1371/journal.pone.0069888
129. Willis LM, Gilbert M, Karwaski MF, Blanchard MC, Wakarchuk WW. Characterization of the  $\alpha$ -2,8-polysialyltransferase from *Neisseria meningitidis* with synthetic acceptors, and the development of a self-priming polysialyltransferase fusion enzyme. *Glycobiology* (2008) **18**:177–86. doi:10.1093/glycob/cwm126
130. Jayalakshmi V, Biet T, Peters T, Krishna NR. Refinement of the conformation of UDP-galactose bound to galactosyltransferase using the STD NMR intensity-restrained CORCEMA optimization. *J Am Chem Soc* (2004) **126**:8610–1. doi:10.1021/ja048703u
131. Angulo J, Langpap B, Blume A, Biet T, Meyer B, Rama Krishna N, et al. Blood group B galactosyltransferase: insights into substrate binding from



- NMR experiments. *J Am Chem Soc* (2006) **128**:13529–38. doi:10.1021/ja063550r
132. Rademacher C, Landström J, Sindhuwinata N, Palcic MM, Widmalm G, Peters T. NMR-based exploration of the acceptor binding site of human blood group B galactosyltransferase with molecular fragments. *Glycoconj J* (2010) **27**:349–58. doi:10.1007/s10719-010-9282-5
  133. Woods RJ, Tessier MB. Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan–protein complexes. *Curr Opin Struct Biol* (2010) **20**:575–83. doi:10.1016/j.sbi.2010.07.005
  134. Patel RY, Balaji PV. Fold-recognition and comparative modeling of human beta3GalT I, II, IV, V and VI and beta3GalNAcT I: prediction of residues conferring acceptor substrate specificity. *J Mol Graph Model* (2007) **26**:255–68. doi:10.1016/j.jmgm.2006.12.003
  135. Rosén ML, Edman M, Sjöström M, Wieslander A. Recognition of fold and sugar linkage for glycosyltransferases by multivariate sequence analysis. *J Biol Chem* (2004) **279**:38683–92. doi:10.1074/jbc.M402925200
  136. Nakahara T, Hindsgaul O, Palcic MM, Nishimura SI. Computational design and experimental evaluation of glycosyltransferase mutants: engineering of a blood type B galactosyltransferase with enhanced glucosyltransferase activity. *Protein Eng Des Sel* (2006) **19**:571–8. doi:10.1093/protein/gzl046
  137. Macnaughtan MA, Alvarez-Manilla MKG, Venot A, Glushka J, Pierce JM, Prestegard JH. NMR structural characterization of substrates bound to N-acetylglucosaminyltransferase V. *J Mol Biol* (2007) **366**:1266–81. doi:10.1016/j.jmb.2006.12.015
  138. Sindhuwinata N, Grimm LL, Weißbach S, Zinn S, Munoz E, Palcic MM, et al. Thermodynamic signature of substrates and substrate analogs binding to human blood group B galactosyltransferase from isothermal titration calorimetry experiments. *Biopolymers* (2013) **99**:784–95. doi:10.1002/bip.22297
  139. Soya N, Shoemaker GK, Palcic MM, Klassen JS. Comparative study of substrate and product binding to the human ABO(H) blood group glycosyltransferases. *Glycobiology* (2009) **19**:1224–34. doi:10.1093/glycob/cwp114
  140. Palcic MM. Glycosyltransferases as biocatalysts. *Curr Opin Chem Biol* (2011) **15**:226–33. doi:10.1016/j.cbpa.2010.11.022
  141. Jørgensen R, Pesnot T, Lee HJ, Palcic MM, Wagner GK. Base-modified donor analogues reveal novel dynamic features of a glycosyltransferase. *J Biol Chem* (2013) **288**:26201–8. doi:10.1074/jbc.M113.465963
  142. Jørgensen R, Grimm LL, Sindhuwinata N, Peters T, Palcic MM. A glycosyltransferase inhibitor from a molecular fragment library simultaneously interferes with metal ion and substrate binding. *Angew Chem Int Ed* (2012) **51**:4171–5. doi:10.1002/anie.201108345
  143. Schaefer K, Albers J, Sindhuwinata N, Peters T, Meyer BA. New concept for glycosyltransferase inhibitors: nonionic mimics of the nucleotide donor of the human blood group B galactosyltransferase. *Chembiochem* (2012) **13**:443–50. doi:10.1002/cbic.201100642
  144. Pesnot T, Palcic MM, Wagner GK. A novel fluorescent probe for retaining galactosyltransferases. *Chembiochem* (2010) **11**:1392–8. doi:10.1002/cbic.201000013
  145. Yi W, Shao J, Zhu L, Li M, Singh M, Lu Y, et al. *Escherichia coli* O86 O-antigen biosynthetic gene cluster and stepwise enzymatic synthesis of human blood group B antigen tetrasaccharide. *J Am Chem Soc* (2005) **127**:2040–1. doi:10.1021/ja045021y
  146. Su DM, Eguchi H, Yi W, Li L, Wang PG, Xia C. Enzymatic synthesis of tumor-associated carbohydrate antigen Globo-H hexasaccharide. *Org Lett* (2008) **10**:1009–12. doi:10.1021/ol703121h
  147. Chang A, Singh S, Phillips GN Jr., Thorson JS. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Curr Opin Biotechnol* (2011) **22**:800–8. doi:10.1016/j.copbio.2011.04.013
  148. Hancock SM, Vaughan MD, Withers SG. Engineering of glycosidases and glycosyltransferases. *Curr Opin Chem Biol* (2006) **10**:509–19. doi:10.1016/j.cbpa.2006.07.015
  149. Nakai H, Kitaoka M, Svensson B, Ohtsubo K. Recent development of phosphorylases possessing large potential for oligosaccharide synthesis. *Curr Opin Chem Biol* (2013) **17**:301–9. doi:10.1016/j.cbpa.2013.01.006
  150. Yu H, Thon V, Lau K, Cai L, Chen Y, Mu S, et al. Highly efficient chemoenzymatic synthesis of  $\beta$ 1–3-linked galactosides. *Chem Commun (Camb)* (2010) **46**:7507–9. doi:10.1039/c0cc02850a
  151. Nishimoto M, Kitaoka M. One-pot enzymatic production of beta-D-galactopyranosyl-(1 → 3)-2-acetamido-2-deoxy-D-galactose (galacto-N-biose) from sucrose and 2-acetamido-2-deoxy-D-galactose (N-acetylgalactosamine). *Carbohydr Res* (2009) **344**:2573–6. doi:10.1016/j.carres.2009.09.031
  152. Armstrong Z, Withers SG. Synthesis of glycans and glycopolymers through engineered enzymes. *Biopolymers* (2013) **99**:666–74. doi:10.1002/bip.22335

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 July 2014; accepted: 23 September 2014; published online: 20 October 2014.

Citation: Brockhausen I (2014) Crossroads between bacterial and mammalian glycosyltransferases. *Front. Immunol.* 5:492. doi: 10.3389/fimmu.2014.00492

This article was submitted to *Immunotherapies and Vaccines*, a section of the journal *Frontiers in Immunology*.

Copyright © 2014 Brockhausen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# MCL and Mincle: C-type lectin receptors that sense damaged self and pathogen-associated molecular patterns

Mark B. Richardson and Spencer J. Williams\*

School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, VIC, Australia

## Edited by:

Mark Agostino, Curtin University, Australia

## Reviewed by:

Katsumi Maenaka, Hokkaido University, Japan

Norberto Walter Zwirner, Institute of Biology and Experimental Medicine, Argentina

## \*Correspondence:

Spencer J. Williams, School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, 30 Flemington Road, Parkville, VIC 3010, Australia  
e-mail: sjwill@unimelb.edu.au

Macrophage C-type lectin (MCL) and macrophage inducible C-type lectin (Mincle) comprise part of an extensive repertoire of pattern recognition receptors with the ability to sense damage-associated and pathogen-associated molecular patterns. In this review, we cover the discovery and molecular characterization of these C-type lectin receptors, and highlight recent advances in the understanding of their roles in orchestrating the response of the immune system to bacterial and fungal infection, and damaged self. We also discuss the identification and structure–activity relationships of activating ligands, particularly trehalose dimycolate and related mycobacterial glycolipids, which have significant potential in the development of T<sub>H</sub>1/T<sub>H</sub>17 vaccination strategies.

**Keywords:** cord factor, C-type lectin receptors, glycolipids, PAMP, DAMP, T cell

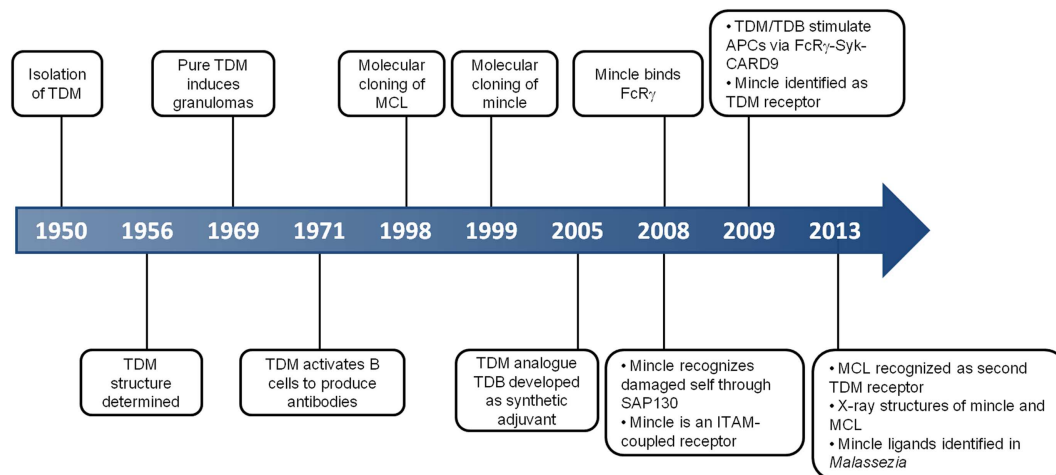
## INTRODUCTION

Macrophage C-type lectin (MCL; Clec4d, ClecSf8) and macrophage inducible C-type lectin (Mincle; Clec4e, ClecSf9) are transmembrane germline-encoded pattern recognition receptors (PRRs) that form part of the innate immune system. These C-type lectin receptors (CLRs) recognize damage-associated molecular patterns (DAMPs) and enable immune sensing of damaged self, and pathogen-associated molecular patterns (PAMPs) from a growing list of bacteria and fungi (**Figure 1**) (1, 2). The PAMPs notably include mycobacterial trehalose dimycolate (TDM, cord factor), and appear to play a significant roles in the immune response to certain bacterial and fungal infections. In the case of Mincle, recognition of PAMPs is mediated through the carbohydrate binding part of the carbohydrate recognition domain (CRD) in the extracellular region of the CLR, whereas recognition of DAMPs occurs through a distinct region of the CRD (3). For both CLRs, signal transduction occurs through the immunoreceptor tyrosine-based activation motif (ITAM)-containing adaptor molecule Fc receptor  $\gamma$ -chain (FcR $\gamma$ ). Ligand binding to Mincle leads to phosphorylation of the ITAM of FcR $\gamma$  and recruitment of spleen tyrosine kinase (Syk) (3). Syk recruitment by FcR $\gamma$  leads to nuclear factor kappa-light-chain-enhancer of activated B cells (NF- $\kappa$ B) activation through Card9–Bcl10–MALT1 signalosomes, pivotal regulators that link innate and adaptive immune responses (4). DNA transcription leads to the production of cytokines and chemokines that shape the development of naïve T cells into effector T helper (T<sub>H</sub>) cell T<sub>H</sub>1 and T<sub>H</sub>17 subtypes (5). There is a growing interest in the immunological roles of MCL and Mincle for the development of defined synthetic adjuvants for T<sub>H</sub>1/T<sub>H</sub>17 vaccination, and due to the involvement of these CLRs in the recognition of bacterial and fungal pathogens, and the response to dysregulated cell death.

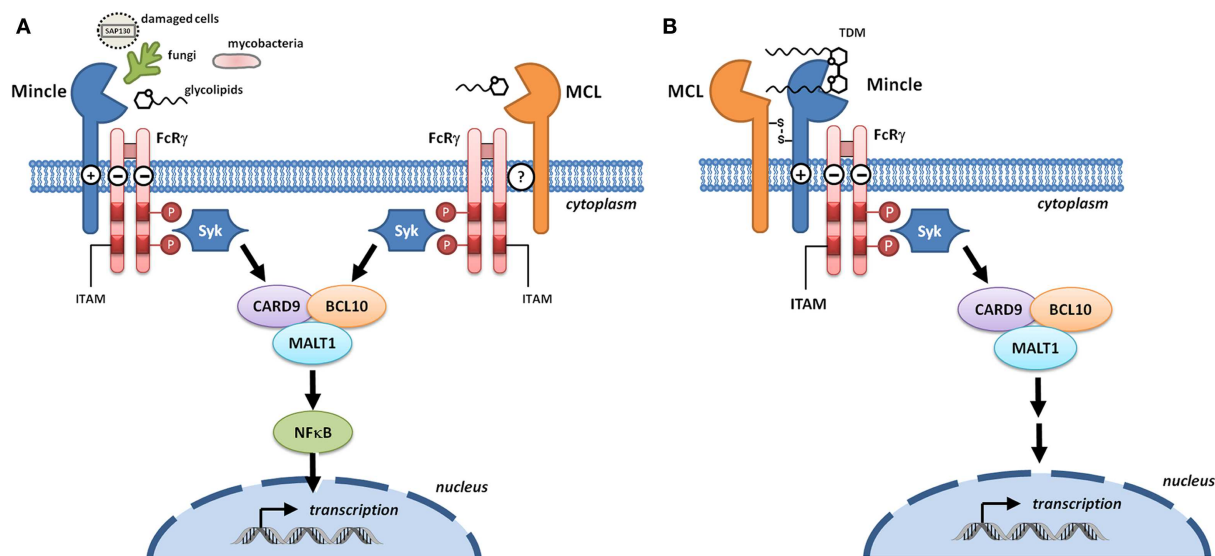
## MCL AND MINCLE: C-TYPE LECTIN RECEPTORS

The genetic and molecular analysis of MCL and Mincle preceded the understanding of their function. MCL was originally cloned as a mouse macrophage-restricted C-type lectin (6). Mincle was cloned as a transcriptional target of the nuclear factor NF-IL-6, which binds the interleukin-1 (IL-1) responsive element of the IL-6 gene (7). Mincle RNA was induced upon exposure to inflammatory stimuli including lipopolysaccharide (LPS), tumor necrosis factor- $\alpha$  (TNF $\alpha$ ), IL-6, and interferon- $\gamma$  (IFN $\gamma$ ). Cloning/identification of the human and rat genes followed shortly thereafter (8, 9). According to the HUGO Gene Nomenclature Committee, the gene encoding MCL is CLEC4D (formerly CLECSF8) and that encoding Mincle is CLEC4E (formerly CLECSF9). Both CLRs are located on mouse chromosome 6 and human chromosome 12, clustered with closely related CLRs of the Dectin-2 family (Dectin-2, DCIR, DCAR, and BDCA-2) within the telomeric region of the natural killer complex (10). Sequence analyses predict type 2 transmembrane proteins with an N-terminal cytoplasmic tail, a transmembrane region, a membrane proximal stalk, followed by a single C-terminal C-type lectin domain (11). The transmembrane region of Mincle contains a positively charged arginine that mediates interaction with ITAM-bearing adaptors; a direct interaction of Mincle with FcR $\gamma$  was obtained by immunoprecipitation of Mincle using an anti-FcR $\gamma$  antibody in transfected human embryonic kidney cells, with the precipitation abolished upon substitution of arginine by leucine (**Figure 2A**) (3). This interaction is critical for the ability of Mincle to signal as Mincle-dependent signaling is lost in the leucine mutant (3).

Macrophage C-type lectin is an FcR $\gamma$  coupled receptor (12) that signals through Syk (13); however the nature of the adaptor protein interaction is unclear (**Figure 2A**). MCL does not contain an



**FIGURE 1 | Key events in the history of MCL and Mincle.** This timeline depicts some of the critical discoveries that have shaped our understanding of the function and roles of MCL and Mincle.



**FIGURE 2 | Signaling through Mincle and MCL. (A)** Mincle and MCL pair with the signaling adaptor molecule Fc receptor  $\gamma$ -chain (Fc $\gamma$ ). With Mincle this association is driven by the presence of a positively charged arginine in the transmembrane region whereas the corresponding residue is not present in MCL. Upon binding to DAMPs or fungal and bacterial PAMPs, phosphorylation of the immunoreceptor tyrosine-based activation motifs (ITAMs) of Fc $\gamma$  recruits spleen tyrosine kinase (Syk)

and induces signaling through the Card9–Bcl10–MALT1 complex. Transcription results in the formation of immunoregulatory cytokines and chemokines. **(B)** Heterodimers of MCL and Mincle have the capacity to sense trehalose dimycolate (TDM). In these complexes, Mincle recognizes the carbohydrate headgroup while MCL recognizes the lipid tail, and Mincle acts as a bridge to enable formation of a functional MCL–Mincle–Fc $\gamma$  complex.

arginine residue within the transmembrane domain that is typically required for association with Fc $\gamma$  and so a direct interaction with this adaptor seems unlikely. Immunoprecipitation of MCL in a rat myeloid cell line led to co-precipitation with Fc $\gamma$  (14) but this could not be replicated in a transfected non-myeloid cell line, indicating that additional cellular components are required for this interaction (15). Co-transfection of non-myeloid cells with MCL, Mincle, and Fc $\gamma$  followed by immunoprecipitation gave

evidence for a covalently linked heterodimer of MCL and Mincle that associated with Fc $\gamma$ , suggesting that Mincle acts as a bridge for the interaction between MCL and Fc $\gamma$  (Figure 2B) (15, 16). Phagocytic internalization of anti-Mincle and anti-MCL coated beads provided strong evidence for a functional complex of MCL–Mincle–Fc $\gamma$  on the cell surface. In primary rat peritoneal macrophages, expression of Mincle and MCL is tightly coupled, suggesting that Mincle–MCL heterodimers are formed.

It was proposed that in Mincle–MCL heterodimers, TDM binding is mediated through carbohydrate binding to Mincle and lipid binding to MCL, a model that explains the ability of MCL to recognize soluble TDM but not TDM associated with mycobacterial cells.

In humans and rodents, Mincle is expressed on monocytes, macrophages, neutrophils, dendritic cells (DCs), and some subsets of B cells (3, 7, 17–19), and has been additionally detected on T cells and concanavalin A blasts in rats (9). Mincle is inducible on mouse macrophages upon activation of TLR4 (18), and is also inducible on human B cells through activation of TLR9 (17). In guinea pig, Mincle is expressed in the spleen, lymph nodes, and peritoneal macrophages and is up-regulated upon stimulation by zymosan and TDM (20).

Macrophage C-type lectin expression was first characterized as macrophage-restricted in mice (6). Human MCL is expressed on macrophages (synovial, peritoneal, and blood monocyte-derived) and Langerhans cells (8), and on neutrophils, monocytes, and immature and mature DCs (13). In rats, expression of MCL has been detected on macrophages, neutrophils, DCs, B cells, and T cells (9, 14). Upon treatment with various stimuli including PAM3CSK4, TNF- $\alpha$ , and IFN $\gamma$ , MCL expression is inducible on monocytes and neutrophils (8, 13), but considerable variation is found between studies. For example, using monocytes from human donors, Arce et al. qualitatively demonstrated that IL-6, TNF- $\alpha$ , IL-10, and IFN $\gamma$  caused up-regulation of MCL expression, whereas LPS caused down-regulation (8). In contrast, using similar protocols and cell lines, Graham et al. obtained different results (13). LPS caused up-regulation; and TNF- $\alpha$ , IFN $\gamma$ , IL-6, and IL-10 did not change expression levels, although a less than twofold change in expression levels was observed. Expression of MCL on rat macrophages is inducible with IFN $\gamma$  and pro-inflammatory stimuli from Gram-negative bacteria (14).

Carbohydrate recognition by lectins is frequently associated with conserved glutamic acid–proline–asparagine (EPN) or glutamine–proline–aspartate (QPD) motifs within the CRDs of C-type lectins (see **Box 1**) (21). Mincle contains an EPN motif within its CRD, leading to the suggestion of mannose/fucose/*N*-acetylglucosamine/glucose specificity, which is commonly seen for such motifs. While MCL contains a conserved Ca<sup>2+</sup> binding site, it lacks a canonical EPN or QPD motif, instead possessing an EPX motif in rat (X = K) and human (X = D),

whereas mouse possesses an ESN sequence. Insight into carbohydrate recognition by Mincle and MCL was assessed for six hexoses: mannose, fucose, *N*-acetylglucosamine, glucose, galactose, and *N*-acetylgalactosamine (22). Relative affinities for Mincle were mannose ~ fucose > glucose > *N*-acetylglucosamine > galactose ~ *N*-acetylgalactosamine. These relative affinities were roughly paralleled by MCL: mannose ~ fucose > glucose ~ *N*-acetylglucosamine > galactose ~ *N*-acetylgalactosamine, although this CLR bound only very weakly to all hexoses examined. A screen of a 326-member carbohydrate microarray for ligands for a soluble MCL–Fc fusion protein failed to identify any carbohydrate ligands for this receptor (13), suggesting that MCL may not in fact be a carbohydrate binding lectin.

Functional studies of CLRs have been greatly accelerated through the development of reporter cell lines. In the case of Mincle, a useful reporter is a T cell hybridoma that expresses Mincle and FcR $\gamma$ , as well as green fluorescent protein (GFP) under the control of the transcription factor nuclear factor of activated T cells (NFAT) to detect ITAM-mediated signals (3). An MCL reporter strain in which LacZ  $\beta$ -galactosidase is expressed under the control of NFAT in a T cell hybrid has been reported (15).

### MINCLE IN THE DAMAGED CELL RESPONSE

Dead cells activate Mincle-expressing cells. The factor causing activation was identified to be spliceosome-associated protein 130 (SAP130) (3). SAP130 binds to Mincle in a Ca<sup>2+</sup> independent manner and mutation of the EPN motif of Mincle did not affect binding. Conversely, a mutant in the region recognized by a blocking antibody, VEGQW, was not activated by dead cells, suggesting that SAP130 binds to the CRD but at a distinct site to that of carbohydrate binding. SAP130 derived from either living or dead cells has a similar ability to activate through Mincle. As SAP130 is located in the nucleus in live cells, Mincle recognition of dead cells must occur after translocation to the external milieu. SAP130 therefore acts as a DAMP, providing an alarm signal for dysregulated cell death. Gamma-ray irradiation causes cell death in the thymus and induces neutrophil infiltration. Mincle RNA is up-regulated after irradiation, but use of a Mincle-blocking antibody suppressed neutrophil infiltration into the thymus after irradiation. The observation that macrophage inflammatory protein 2 (MIP-2), a specific signal produced by thymic macrophages, was inhibited by the Mincle-blocking antibody suggests that Mincle

#### Box 1 | C-type lectins: structure and classification.

C-type lectins are a subclass of lectins that are distinguished by a requirement for Ca<sup>2+</sup> for binding (21). Crystallographic studies reveal that C-type lectins contain a compact globular structure that comprises the carbohydrate recognition domain (CRD). The CRD contains conserved amino acid residue motifs, and which allows the prediction of new C-type lectins on the basis of sequence data. Somewhat confusingly, it has since been found that many predicted C-type lectin CRDs do not bind either carbohydrates or Ca<sup>2+</sup>. C-type lectins can be soluble or transmembrane proteins. An early classification of C-type lectins was introduced on the basis of (1) the identification of mannose- or galactose-specific binding motifs [glutamic acid–proline–asparagine (EPN) or glutamine–proline–aspartate (QPD) motifs, respectively] in well-characterized mannose- and galactose-specific lectins known at the time, and (2) conversion of the ligand specificity from mannose to galactose by mutagenesis (23). Although based on compelling studies at the time, this further adds to the confusion as some predicted mannose or galactose binding C-type lectins do not in fact bind to these carbohydrates. C-type lectins are classified into 17 groups; Mincle and MCL fall into Group II of the CTL family and are typically grouped in with a subset of CTLRs termed the Dectin-2 cluster that contains other PRRs including Dectin-2, DCIR, DCAR, and BDCA-2 (10).

activation by ligands induces the production of inflammatory cytokines and/or chemokines.

Ischemia results in dysregulated cell death and the exit of cellular components, suggesting the possible involvement of Mincle-mediated inflammation. Mincle knockout mice show a better outcome after stroke (24). Cerebral ischemia results in induction of Mincle expression in immune, neuronal, and endothelial cells, which paralleled increases in SAP130 expression (25). Levels of phosphorylated-Syk (p-Syk) were raised following ischemia suggesting that Mincle activation leads to increased levels of p-Syk. Application of the Syk inhibitor piceatannol reduced infarct volume and swelling, suggesting that signaling through the Syk-Card9-Bcl10-Malt1 axis is an important factor in the response to ischemia.

### IDENTIFICATION OF MYCOBACTERIAL GLYCOLIPIDS AS MCL AND MINCLE ANTIGENS AND THEIR ROLE IN MYCOBACTERIAL INFECTION

Research in the areas of mycobacterial immunogenicity and C-type lectins began to merge with the report that TDM activates macrophages and DCs via Syk-Card9-Bcl10-Malt1 signaling to produce innate activation that was distinct from that produced by Toll-like receptor ligands (4). Activation of antigen presenting cells was independent of Dectin-1 but required the ITAM-bearing signaling adaptor FcR $\gamma$ , leading to the suggestion that a range of C-type lectins, including Mincle, were possible TDM receptors (4). Independently, Mincle had been shown to be associated with FcR $\gamma$  (3). Two contemporaneous reports identified Mincle as a TDM receptor. Using Mincle-expressing reporter cells, Yamasaki and co-workers showed that while heat-killed mycobacteria could activate Mincle-expressing cells, delipidated cells could not, and the activity was located within the lipid extract (26, 27). Sub-fractionation of this extract identified TDM as the activating species. Independently, Lang and co-workers mined a gene expression array database for genes expressed and up-regulated in bone marrow macrophages treated with TDB (28). The candidate Mincle was expressed as an Fc fusion protein and was shown by ELISA to bind to TDB and TDM in a dose-dependent manner. TDM is an important glycolipid produced by all mycobacteria that possesses potent immunostimulatory properties, in particular the ability to cause granulomas. Using Mincle<sup>-/-</sup> mice, Mincle was shown to be essential for the granulomatous response to TDM, providing compelling evidence that Mincle is a major TDM receptor (26). While initial reports demonstrated that TDM activates mouse Mincle, a recent report using a reporter cell has shown that TDM also activates human Mincle (29).

The discovery that MCL is a receptor for TDM arose from the initial observation that MCL is expressed on neutrophils and monocytes and triggers cellular activation through Syk (13). The observation that resting macrophages barely express Mincle, yet addition of TDM drives Mincle expression, suggested the existence of another TDM receptor (12). Innate immune responses were impaired in MCL-deficient mice, including the TDM-induced acquired immune response, experimental autoimmune encephalomyelitis (EAE) (12). Further, MCL was shown to be required to drive Mincle expression in a Clec4e-GFP fusion reporter mouse (12). Daws reported that MCL reporter strains

were not responsive to intact mycobacteria and argued that this observation is consistent with MCL recognizing the lipid portion of TDM, which is exposed in the purified antigen but embedded in the bacterial cell wall in intact bacteria (15). Recent studies of guinea pig homologs of Mincle (gpMincle) and MCL (gpMCL) revealed that only gpMincle binds TDM and that gpMincle is constitutively expressed (20). gpMCL lacks the hydrophobic region proposed to be involved in TDM binding, although it does bind FcR $\gamma$ . This work suggested that of these two receptors, only gpMincle is involved in guinea pig immune responses against mycobacteria and that the functions of MCL and Mincle in recognition of mycobacteria are not conserved between humans and guinea pig.

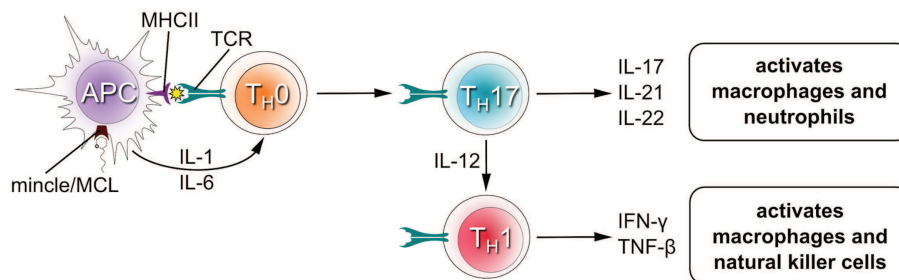
The identification of Mincle and MCL as TDM receptors and establishing that Mincle is required for the TDM-induced granulomatous response has resolved longstanding questions in the field. However, the connection of Mincle and MCL to the anti-mycobacterial response in the context of infection has been less clearly answered (30). An early role for the cytosolic adaptor caspase recruitment domain family, member 9 (Card9) in pulmonary tuberculosis was demonstrated when Card9<sup>-/-</sup> mice were shown to succumb rapidly to aerosol infection by *Mycobacterium tuberculosis* H37Rv (31). Comparison of bone marrow derived macrophages of wild-type and Card9<sup>-/-</sup> mice when infected with *M. tuberculosis* revealed a significant reduction in the homozygous mutant of the pro-inflammatory cytokines, TNF, IL-1 $\beta$ , and IL-6, and reduced IL-12 and CCL5, compared to wild-type. Heat-killed *Mycobacterium bovis*, *Mycobacterium smegmatis*, and *M. tuberculosis* H37Rv all activated an NFAT-driven GFP reporter strain that expresses Mincle and FcR $\gamma$ , with activation ablated upon mutation of the Mincle CRD EPN motif to QPD (26). The effect of Mincle deletion upon *M. tuberculosis* infection has been studied in Mincle<sup>-/-</sup> mice. While in the absence of Mincle, macrophages did not produce the reactive nitrogen species NO<sub>2</sub><sup>-</sup> upon TDM stimulation, Mincle-deficient macrophages were responsive to *M. tuberculosis* infection, and co-stimulation with IFN $\gamma$  resulted in normal levels of granulocyte colony-stimulating factor (G-CSF), TNF and NO<sub>2</sub><sup>-</sup>. Mincle<sup>-/-</sup> mice did not form granulomas upon stimulation with TDM, the same mice forms granulomas that were indistinguishable from wild-type upon infection with *M. tuberculosis*. Furthermore, while TDM resulted in induction of T<sub>H</sub>1 and T<sub>H</sub>17 cells, Mincle<sup>-/-</sup> mice infected with *M. tuberculosis* developed normal T<sub>H</sub>1 and T<sub>H</sub>17 immune responses, indicating that in the context of *M. tuberculosis* infection, Mincle is not required for instructing maturation of naïve T cells. This observation is consistent with studies with *Fonsecaea pedrosoi* that suggests that while specific signals are transmitted by Mincle during vaccination with purified TDM, PRR co-stimulation during infection leads to a cocktail of cytokines and chemokines that shape T cell development (32) (see Box 2).

MCL-deficient mice have defective immune responses to mycobacterial infection. Induction of TNF and MIP-2 RNA was impaired in bone marrow derived macrophages from MCL<sup>-/-</sup> mice (12). Further, MCL<sup>-/-</sup> mice gave impaired IFN $\gamma$  levels in Mantoux test responses when stimulated with purified protein derivative compared to wild-type.



### Box 2 | The role of MCL and Mincle in T cell development.

Upon recognition of their cognate antigen, naïve CD4<sup>+</sup> T cells (T<sub>H</sub>0 cells) can differentiate into novel effector CD4<sup>+</sup> T cell lineages that can regulate or assist active immune responses (**Figure 3**). The fate of the transforming T<sub>H</sub>0 cell is determined by the pattern of cytokines it receives at the moment of antigen recognition by the T cell receptor. Activation of Mincle and MCL induces expression of interleukins-1 (IL-1) and IL-6 in antigen presenting cells (4, 19, 26, 33, 34), which flood the microenvironment of the transforming T<sub>H</sub>0 cell and shape the development of T<sub>H</sub>1 and T<sub>H</sub>17 phenotypes in both humans and mice (12, 28). The T<sub>H</sub>17 lineage is defined by the production of IL-17, which induces and mediates pro-inflammatory responses leading to the recruitment of monocytes and neutrophils, which can clear infections (35). T<sub>H</sub>1 cells characteristically produce IFN $\gamma$ , an activator of natural killer cells (which provide direct killing of pathogens) and macrophages (leading to phagocytosis).

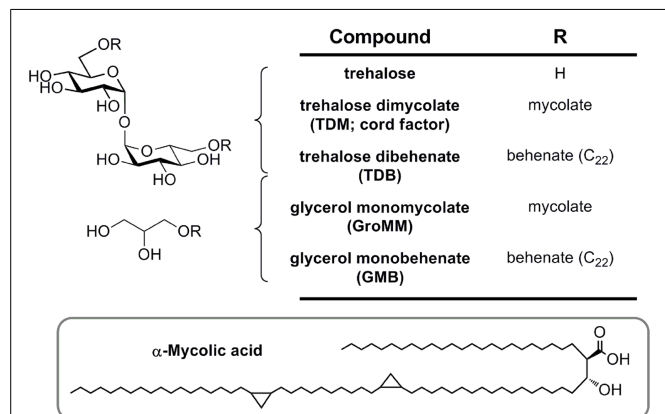


**FIGURE 3 | T cell differentiation upon stimulation of Mincle/MCL.**

Glycerol monomycolate (GroMM) has been identified as a Mincle-activating lipid (29). Using NFAT–GFP reporter strains, it was shown that human Mincle-expressing cells could be activated by GroMM, although less potently than for TDM. While mouse Mincle-expressing cells were not activated by GroMM, transgenic mice expressing human Mincle gained the ability to recognize GroMM, and injection of GroMM liposomes into the skin of these mice resulted in infiltration of macrophages and eosinophils. In primary human monocyte-derived macrophages, GroMM produced TNF $\alpha$  in response to GroMM that could be blocked by an anti-human Mincle antibody, demonstrating that this glycolipid is a ligand for human, but not mouse Mincle.

### STRUCTURE–ACTIVITY RELATIONSHIPS OF TREHALOSE AND GLYCEROL-BASED ANTIGENS FOR MINCLE

Mincle is potently activated by TDM (**Figure 4**). Treatment of TDM with trehalase, which apparently cleaves this molecule into glucose monomycolate (GMM), abolished Mincle binding (26). This result suggests that Mincle specifically recognizes the two glucose residues within TDM and is particularly interesting as GMM is a glycolipid produced upon infection by *M. tuberculosis* and which itself is a potent antigen when presented to T cells by CD1b (36). Cells are activated to similar degrees by the TDM analog trehalose dibehenate, which suggests that complex mycolate structures are not necessary for Mincle activation. Measurement of the direct interaction of Mincle with TDM analogs is limited by their poor solubility and so accurate data has only been obtained with shorter acyl groups. For bovine Mincle, affinities of trehalose diesters generally increase with increasing chain length (37). Trehalose monoesters are also effective ligands for human and bovine Mincle with affinities increasing up to 6-*O*-lauryltrehalose



**FIGURE 4 | Mycobacterial mycolic acid-based lipids and synthetic analogs (TDB and GMB), structurally defined glycolipid ligands for MCL and Mincle.**

(C<sub>12</sub>) and 6-*O*-octanoyltrehalose, respectively (37, 38). Trehalose itself is a weak ligand for Mincle, and methyl  $\alpha$ -glucoside, representing a monomer of trehalose, possesses 36-fold weaker affinity for Mincle (37). The C<sub>22</sub> and C<sub>26</sub> trehalose monoesters were able to activate mouse macrophages in a Mincle-dependent manner (34).

Glycerol monomycolate is an antigenic ligand for human, but not mouse Mincle (**Figure 4**) (29). Plate-bound GroMM is less potent than TDM in an NFAT–GFP reporter cell assay. Glycerol monobehenate (GMB) possesses similar activity to GroMM toward human Mincle and gave only marginal responses for mouse Mincle reporter cells.

## STRUCTURES OF CRDs OF MCL AND MINCLE

The classification of Mincle and MCL as CLRs leads to the prediction that the CRD domain will conform to the typical domains seen for this class of proteins. Two recent papers have independently reported structures of Mincle and MCL C-type lectin domains. Maenaka and co-workers reported the structures of human Mincle and MCL, recombinantly expressed in *Escherichia coli* (38). While wild-type human Mincle failed to crystallize, the I99K mutant (which matches the equivalent residue in human MCL) provided diffracting crystals. Drickamer and co-workers have reported the structure of bovine Mincle, expressed using a similar approach and crystallized without the need for mutagenesis (37).

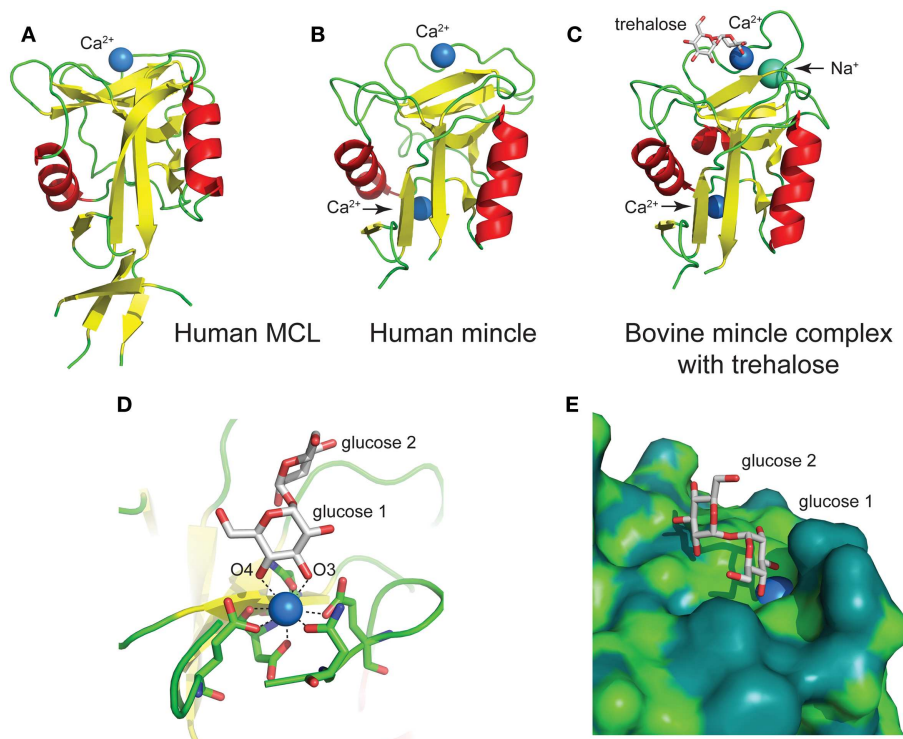
The human MCL C-type lectin domain comprises a globular fold containing two alpha helices around a beta-strand core, with a single  $\text{Ca}^{2+}$  bound (Figure 5A) (38). The human and bovine Mincle C-type lectin domains reveal largely identical folds, and both contain two  $\text{Ca}^{2+}$  ions, with the bovine structure containing an additional  $\text{Na}^+$  ion (Figure 5B) (38). Structural insight into carbohydrate recognition by Mincle was obtained through a complex of bovine Mincle with trehalose (Figure 5C) (37). In this complex, only one pyranose ring of trehalose binds to calcium through O3 and O4 of glucose (Figure 5D). Insight into a possible molecular basis for signal transduction was obtained through a conformational change in a loop between Asn-170 and Asp-177 near the conserved  $\text{Ca}^{2+}$  upon binding of trehalose. Examination

of the surface of Mincle near the trehalose binding site identified a hydrophobic channel lined with lipophilic amino acids that is adjacent to the 6-OH of the  $\text{Ca}^{2+}$ -bound glucose ring, which was proposed to comprise a lipid binding channel (Figure 5E).

## MCL AND MINCLE IN FUNGAL DISEASE

Effective protection of host from pathogenic fungi and clearance of infection requires a coordinated immune response from both  $\text{T}_\text{H}1$  and  $\text{T}_\text{H}17$  cells that restrict fungal cell growth and drive phagocytic clearance (2, 5). The Syk–Card9–Bcl10–MALT1 signaling axis leads to NF- $\kappa$ B activation, which drives the expression of multiple cytokines including IL-1, IL-6, and IL-23 that facilitate  $\text{T}_\text{H}17$  cell differentiation, and IL-12, which is essential for  $\text{T}_\text{H}1$  differentiation (see Box 2).

Yamasaki and co-workers screened 50 species of pathogenic fungi as Mincle activators using a NFAT–GFP reporter cell line (39). Of these only *Malassezia* species, including *Malassezia pachydermatis*, *Malassezia dermatis*, *Malassezia japonica*, *Malassezia nana*, *Malassezia slooffiae*, *Malassezia sympodialis*, and *Malassezia furfur*, induced strong NFAT–GFP activation. Interestingly, particularly given results from other laboratories, *Candida albicans* did not result in NFAT–GFP expression. In normal skin, *Malassezia* spp. are commensals, however in atopic/eczema and psoriasis, these fungi can elicit inflammatory responses in skin lesions and can cause diseases such as tinea versicolor, atopic dermatitis, and lethal sepsis. Reporter cells with mutant Mincle mutated in the



**FIGURE 5 |** Structures of the C-type lectin domains of (A) human MCL (PDB: 3whd); (B) human Mincle (PDB: 3wh3); (C) bovine Mincle complex with trehalose (PDB: 4kzv); (D) interactions with  $\text{Ca}^{2+}$  at the primary glucose binding site of bovine Mincle; and (E) putative

hydrophobic channel adjacent to the primary glucose binding site of bovine Mincle complex with trehalose (the protein surface is colored according to the underlying atoms: light green, nitrogen/oxygen; dark green, carbon).

CRD were not activated by *Malassezia* indicating that the likely ligand is a carbohydrate, and as discussed below, subsequent work led to the isolation of glycolipids including gentiobiosyl diacylglycerides and a complex mannosyloxylstearyl mannitol glycolipid (40). Co-culture of *M. pachydermatis* with wild-type murine bone marrow macrophages resulted in production of MIP-2, TNF $\alpha$ , keratinocyte-derived chemokine (CXCL1), and IL-10 cytokines, which were reduced but not ablated in Mincle<sup>-/-</sup> macrophages. When *M. pachydermatis* was injected into the peritoneal cavity, wild-type mice produced IL-6 and TNF, which was reduced in the Mincle<sup>-/-</sup> mutant, suggesting that Mincle is important in immune responses to these fungi.

Chromoblastomycosis is a chronic skin infection caused by fungi including *F. pedrosoi*. *F. pedrosoi* is recognized by Mincle, resulting in production of high levels of IL-10 and low levels of TNF and IFN $\gamma$  (32). Under these conditions, *F. pedrosoi* establishes a chronic infection which is unable to be cleared. Co-stimulation of macrophages and DCs infected with *F. pedrosoi* with the TLR2 agonist PAM3CSK4, the TLR7 agonist Imiquimod, or the TLR4 agonist LPS gave robust levels of TNF, suggesting that this fungus fails to cause co-stimulation of PRRs. Indeed, Mincle<sup>-/-</sup> bone marrow DCs lacked the ability to be co-stimulated by LPS and *F. pedrosoi*, and *F. pedrosoi* infected mice effectively cleared disseminated infections with this fungus when treated with a single dose of LPS. As well, topical application of Imiquimod significantly decreased fungal burden in skin after subcutaneous infection. To gain further insight into the mechanisms IL-6 of inflammatory responses, Wevers and co-workers studied stimulation of human DCs with the related fungus *Fonsecaea monophora*, also a causative agent of chromoblastomycosis (41, 42). *F. monophora* triggered the maturation of DCs and production of IL-6, IL-1 $\beta$ , and IL-23, but not IL-12p70. In this case, *F. monophora* is recognized by both Dectin-1 and Mincle, and while activation through Dectin-1 resulted in cytokine production, Mincle activation resulted in suppression of IL-12p70 through suppression of IL-12p35 transcription. This in turn was achieved by Mincle signals targeting the proteasomal degradation of nuclear IRF1 via the ubiquitin E3 ligase Mdm2. Mdm2 activation and translocation was a result of Mincle triggering of the signaling kinase PKB, which was not triggered by Dectin-1 or TLR4. Overall, this process leads to impairment of IL-12 and is important for shaping the development of CD4<sup>+</sup> T cells toward T<sub>H</sub>17 cells. This appears to be a general phenomenon for Mincle activation by other fungi including *Fonsecaea compacta* and *Cladophialophora carrionii*, and indeed by the TDM mimic TDB. Naïve CD4<sup>+</sup> T cells co-cultured with DCs primed with the TLR4 agonist LPS differentiate to T<sub>H</sub>1 polarized cells, however in the presence of *F. pedrosoi* and other fungi, these are skewed to a T<sub>H</sub>2 response.

Mincle appears to play crucial roles in *C. albicans* infection. Mincle-deficient mice are more susceptible to systemic candidiasis, and production of TNF- $\alpha$  by macrophages was reduced *in vivo* and *in vitro* (43). The soluble CRD of human and mouse Mincle was found to bind whole *C. albicans* cells (44). However, a subsequent screening study of different *C. albicans* strains did not activate a Mincle-expressing NFAT-GFP reporter strain, leading to the suggestion that strain-specific features are required for Mincle activation (39). The identity of the *C. albicans* ligands is not known.

A direct role for MCL in fungal disease is yet to be demonstrated. No significant differences were observed in the ability of MCL-deficient mice to resist infection with *C. albicans* (13).

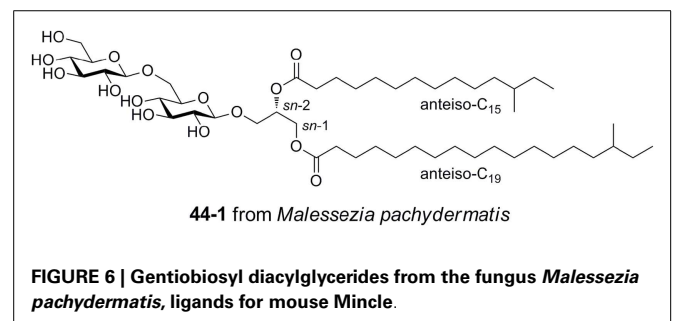
## IDENTIFICATION OF MINCLE ANTIGENS FROM FUNGI

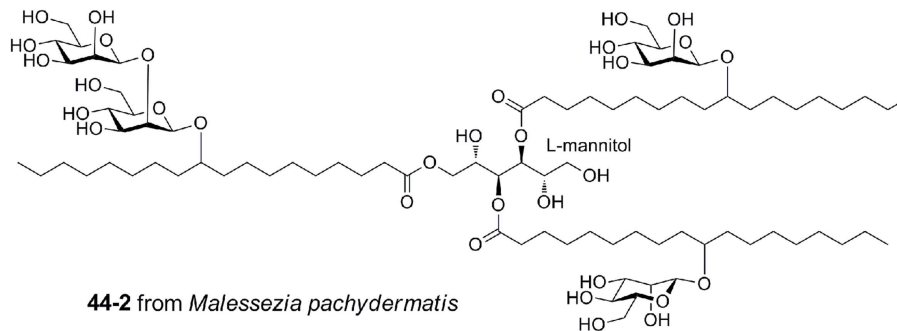
Gentiobiosyl diacylglycerides from *M. pachydermatis* have been identified as mouse Mincle ligands (**Figure 6**) (40). These glycolipids bear anteiso fatty acyl groups (anteiso-C<sub>15</sub>, C<sub>17</sub>, C<sub>19</sub>, or C<sub>20</sub>) at the *sn*-1 and *sn*-2 positions of the glycerol. Four lipofoms were identified with the following substituent permutations (*sn*-1/*sn*-2): anteiso-C<sub>19</sub>/anteiso-C<sub>15</sub>, anteiso-C<sub>17</sub>/anteiso-C<sub>15</sub>, anteiso-C<sub>20</sub>/anteiso-C<sub>15</sub>, and anteiso-C<sub>19</sub>/anteiso-C<sub>17</sub>. All isomers activated Mincle to similar degrees, but were significantly less potent than TDM. While these gentiobiosyl diacylglycerides bear significant resemblance to the glycolipid membrane anchor of lipoteichoic acid, it did not activate via Mincle.

*Malassezia pachydermatis* produces a complex mannosyloxylstearyl mannitol glycolipid that is a potent activator via mouse Mincle, with a potency approaching that of TDM (**Figure 7**) (40). This glycolipid comprises  $\beta$ -linked mannose residues attached to 10-hydroxystearic acid and esterified onto an L-mannitol core.

## CONCLUSION

T cell mediated immune responses induced by PRRs are important in recognition of damaged self and pathogens. The identification of DAMPs and PAMPs that activate Mincle has elevated the importance of this CLR, and the involvement of MCL as an auxiliary CLR that forms heterodimers with Mincle provides the potential for an expansion in the range of ligands recognized by Mincle. MCL and Mincle appear to be an important part of a larger repertoire of PRRs and recent studies of the interplay of Mincle with Dectin-1 and TLRs demonstrate the potential for modulation of immune signals through PRR co-stimulation. The availability of three dimensional X-ray structures of the CRDs of these receptors has unveiled a molecular picture of ligand recognition that may inform development of novel T<sub>H</sub>1/T<sub>H</sub>17 vaccines. However, at this stage it is not clear how Mincle can recognize a diverse range of structurally dissimilar antigens, or the structural changes that lead to signal transduction and transcription. Further, it is likely that additional pathogens that activate MCL and Mincle remain to be identified as well as new ligands from existing pathogens such as *Klebsiella pneumonia* (45) and *F. pedrosoi* that activate through Mincle. It is noteworthy that the majority of studies with MCL and Mincle has been performed on mice and it will require additional work to translate these findings to the human system.





**FIGURE 7 | Mannosyloxylstearyl mannitol glycolipid from *Malessezia pachydermatis*, a ligand for mouse Mincle.**

The species differences noted in the expression and functional properties of MCL and Mincle, and the selective activation of human, but not mouse Mincle by GroMM, suggest that the development of humanized animal models and cell lines will be essential for understanding the role of these CLR in human health and disease.

## ACKNOWLEDGMENTS

We thank the Australian Research Council for financial support. Spencer J. Williams is an ARC funded Future Fellow.

## REFERENCES

- Hardison SE, Brown GD. C-type lectin receptors orchestrate antifungal immunity. *Nat Immunol* (2012) **13**:817–22. doi:10.1038/ni.2369
- Wuthrich M, Deepe GS Jr., Klein B. Adaptive immunity to fungi. *Annu Rev Immunol* (2012) **30**:115–48. doi:10.1146/annurev-immunol-020711-074958
- Yamasaki S, Ishikawa E, Sakuma M, Hara H, Ogata K, Saito T. Mincle is an ITAM-coupled activating receptor that senses damaged cells. *Nat Immunol* (2008) **9**:1179–88. doi:10.1038/ni.1651
- Werninghaus K, Babiak A, Gross O, Holscher C, Dietrich H, Agger EM, et al. Adjuvanticity of a synthetic cord factor analogue for subunit *Mycobacterium tuberculosis* vaccination requires FcRgamma-Syk-Card9-dependent innate immune activation. *J Exp Med* (2009) **206**:89–97. doi:10.1084/jem.20081445
- Geijtenbeek TB, Gringhuis SI. Signalling through C-type lectin receptors: shaping immune responses. *Nat Rev Immunol* (2009) **9**:465–79. doi:10.1038/nri2569
- Balch SG, McKnight AJ, Seldin MF, Gordon S. Cloning of a novel C-type lectin expressed by murine macrophages. *J Biol Chem* (1998) **273**:18656–64. doi:10.1074/jbc.273.29.18656
- Matsumoto M, Tanaka T, Kaisho T, Sanjo H, Copeland NG, Gilbert DJ, et al. A novel LPS-inducible C-type lectin is a transcriptional target of NF-IL6 in macrophages. *J Immunol* (1999) **163**:5039–48.
- Arce I, Martinez-Munoz L, Roda-Navarro P, Fernandez-Ruiz E. The human C-type lectin CLECSF8 is a novel monocyte/macrophage endocytic receptor. *Eur J Immunol* (2004) **34**:210–20. doi:10.1002/eji.200324230
- Flornes LM, Bryceson YT, Spurkland A, Lorentzen JC, Dissen E, Fossum S. Identification of lectin-like receptors expressed by antigen presenting cells and neutrophils and their mapping to a novel gene complex. *Immunogenetics* (2004) **56**:506–17. doi:10.1007/s00251-004-0714-x
- Graham LM, Brown GD. The dectin-2 family of C-type lectins in immunity and homeostasis. *Cytokine* (2009) **48**:148–55. doi:10.1016/j.cyt.2009.07.010
- Balch SG, Greaves DR, Gordon S, McKnight AJ. Organization of the mouse macrophage C-type lectin (Mcl) gene and identification of a subgroup of related lectin molecules. *Eur J Immunogenet* (2002) **29**:61–4. doi:10.1046/j.1365-2370.2002.00266.x
- Miyake Y, Toyonaga K, Mori D, Kakuta S, Hoshino Y, Oyama A, et al. C-type lectin MCL is an FcRgamma-coupled receptor that mediates the adjuvanticity of mycobacterial cord factor. *Immunity* (2013) **38**:1050–62. doi:10.1016/j.immuni.2013.03.010
- Graham LM, Gupta V, Schafer G, Reid DM, Kimberg M, Dennehy KM, et al. The C-type lectin receptor CLECSF8 (CLEC4D) is expressed by myeloid cells and triggers cellular activation through Syk kinase. *J Biol Chem* (2012) **287**:25964–74. doi:10.1074/jbc.M112.384164
- Lobato-Pascual A, Saether PC, Dahle MK, Gaustad P, Dissen E, Fossum S, et al. Rat macrophage C-type lectin is an activating receptor expressed by phagocytic cells. *PLoS One* (2013) **8**:e57406. doi:10.1371/journal.pone.0057406
- Lobato-Pascual A, Saether PC, Fossum S, Dissen E, Daws MR. Mincle, the receptor for mycobacterial cord factor, forms a functional receptor complex with MCL and FcepsilonRI-gamma. *Eur J Immunol* (2013) **43**:3167–74. doi:10.1002/eji.201343752
- Yamasaki S. Signaling while eating: MCL is coupled with Mincle. *Eur J Immunol* (2013) **43**:3156–8. doi:10.1002/eji.201344131
- Kawata K, Illarionov P, Yang GX, Kenny TP, Zhang W, Tsuda M, et al. Mincle and human B cell function. *J Autoimmun* (2012) **39**:315–22. doi:10.1016/j.jaut.2012.04.004
- Lee W-B, Kang J-S, Yan J-J, Lee MS, Jeon B-Y, Cho S-N, et al. Neutrophils promote mycobacterial trehalose dimycolate-induced lung inflammation via the Mincle pathway. *PLoS Pathog* (2012) **8**:e1002614. doi:10.1371/journal.ppat.1002614
- Vijayan D, Radford KJ, Beckhouse AG, Ashman RB, Wells CA. Mincle polarizes human monocyte and neutrophil responses to *Candida albicans*. *Immunol Cell Biol* (2012) **90**:889–95. doi:10.1038/icb.2012.24
- Toyonaga K, Miyake Y, Yamasaki S. Characterization of the receptors for mycobacterial cord factor in guinea pig. *PLoS One* (2014) **9**:e88747. doi:10.1371/journal.pone.0088747
- Zelensky AN, Gready JE. The C-type lectin-like domain superfamily. *FEBS J* (2005) **272**:6179–217. doi:10.1111/j.1742-4658.2005.05031.x
- Lee RT, Hsu TL, Huang SK, Hsieh SL, Wong CH, Lee YC. Survey of immune-related, mannose/fucose-binding C-type lectin receptors reveals widely divergent sugar-binding specificities. *Glycobiology* (2011) **21**:512–20. doi:10.1093/glycob/cwq193
- Drickamer K. Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature* (1992) **360**:183–6. doi:10.1038/360183a0
- Manzanero S, Hsieh YH, Gelderblom M, Wells C, Arumugam TV. The immune receptor Mincle mediates ischemic stroke-induced injury. *Int. J. Stroke* (2012) **7**:19. doi:10.1111/j.1747-4930.2012.00907.x
- Suzuki Y, Nakano Y, Mishiro K, Takagi T, Tsuruma K, Nakamura M, et al. Involvement of Mincle and Syk in the changes to innate immunity after ischemic stroke. *Sci Rep* (2013) **3**:3177. doi:10.1038/srep03177
- Ishikawa E, Ishikawa T, Morita YS, Toyonaga K, Yamada H, Takeuchi O, et al. Direct recognition of the mycobacterial glycolipid, trehalose dimycolate, by C-type lectin Mincle. *J Exp Med* (2009) **206**:2879–88. doi:10.1084/jem.20091750
- Matsunaga I, Moody DB. Mincle is a long sought receptor for mycobacterial cord factor. *J Exp Med* (2009) **206**:2865–8. doi:10.1084/jem.20092533
- Schoenen H, Bodendorfer B, Hitchens K, Manzanero S, Werninghaus K, Nimmerjahn F, et al. Cutting edge: Mincle is essential for recognition and adjuvanticity of the mycobacterial cord factor and its synthetic analog trehalose-dibehenate. *J Immunol* (2010) **184**:2756–60. doi:10.4049/jimmunol.0904013

29. Hattori Y, Morita D, Fujiwara N, Mori D, Nakamura T, Harashima H, et al. Glyc-erol monomycolate is a novel ligand for the human, but not mouse macrophage inducible C-type lectin, Mincle. *J Biol Chem* (2014) **289**(22):15405–12. doi:10.1074/jbc.M114.566489
30. Lang R. Recognition of the mycobacterial cord factor by Mincle: relevance for granuloma formation and resistance to tuberculosis. *Front Immunol* (2013) **4**:5. doi:10.3389/fimmu.2013.00005
31. Dorhoi A, Desel C, Yermeev V, Pradl L, Brinkmann V, Mollenkopf HJ, et al. The adaptor molecule CARD9 is essential for tuberculosis control. *J Exp Med* (2010) **207**:777–92. doi:10.1084/jem.20090067
32. Sousa MG, Reid DM, Schweighoffer E, Tybulewicz V, Ruland J, Langhorne J, et al. Restoration of pattern recognition receptor costimulation to treat chromoblastomycosis, a chronic fungal infection of the skin. *Cell Host Microbe* (2011) **9**:436–43. doi:10.1016/j.chom.2011.04.005
33. Heitmann L, Schoenen H, Ehlers S, Lang R, Holscher C. Mincle is not essential for controlling *Mycobacterium tuberculosis* infection. *Immunobiology* (2013) **218**:506–16. doi:10.1016/j.imbio.2012.06.005
34. Stocker BL, Khan AA, Chee SH, Kamena F, Timmer MS. On one leg: trehalose monoesters activate macrophages in a Mincle-dependant manner. *Chembiochem* (2014) **15**:382–8. doi:10.1002/cbic.201300674
35. Dubin PJ, Kolls JK. Th17 cytokines and mucosal immunity. *Immunol Rev* (2008) **226**:160–71. doi:10.1111/j.1600-065X.2008.00703.x
36. Moody DB, Reinhold BB, Guy MR, Beckman EM, Frederique DE, Fur-long ST, et al. Structural requirements for glycolipid antigen recognition by CD1b-restricted T cells. *Science* (1997) **278**:283–6. doi:10.1126/science.278.5336.283
37. Feinberg H, Jegouzo SA, Rowntree TJ, Guan Y, Brash MA, Taylor ME, et al. Mechanism for recognition of an unusual mycobacterial glycolipid by the macrophage receptor Mincle. *J Biol Chem* (2013) **288**:28457–65. doi:10.1074/jbc.M113.497149
38. Furukawa A, Kamishikiryō J, Mori D, Toyonaga K, Okabe Y, Toji A, et al. Structural analysis for glycolipid recognition by the C-type lectins Mincle and MCL. *Proc Natl Acad Sci U S A* (2013) **110**:17438–43. doi:10.1073/pnas.1312649110
39. Yamasaki S, Matsumoto M, Takeuchi O, Matsuzawa T, Ishikawa E, Sakuma M, et al. C-type lectin Mincle is an activating receptor for pathogenic fungus, *Malassezia*. *Proc Natl Acad Sci U S A* (2009) **106**:1897–902. doi:10.1073/pnas.0805177106
40. Ishikawa T, Itoh E, Yoshida S, Saijo S, Matsuzawa T, Gono T, et al. Identification of distinct ligands for the C-type lectin receptors Mincle and dectin-2 in the pathogenic fungus *Malassezia*. *Cell Host Microbe* (2013) **13**:477–88. doi:10.1016/j.chom.2013.03.008
41. Rivera A. When PRRs collide: Mincle meddles with dectin and toll. *Cell Host Microbe* (2014) **15**:397–9. doi:10.1016/j.chom.2014.03.013
42. Wevers BA, Kaptein TM, Zijlstra-Willems EM, Theelen B, Boekhout T, Geijtenbeek TB, et al. Fungal engagement of the C-type lectin Mincle suppresses dectin-1-induced antifungal immunity. *Cell Host Microbe* (2014) **15**:494–505. doi:10.1016/j.chom.2014.03.008
43. Wells CA, Salvage-Jones JA, Li X, Hitchens K, Butcher S, Murray RZ, et al. The macrophage-inducible C-type lectin, Mincle, is an essential component of the innate immune response to *Candida albicans*. *J Immunol* (2008) **180**:7404–13. doi:10.4049/jimmunol.180.11.7404
44. Bugarcic A, Hitchens K, Beckhouse AG, Wells CA, Ashman RB, Blanchard H. Human and mouse macrophage-inducible C-type lectin (Mincle) bind *Candida albicans*. *Glycobiology* (2008) **18**:679–85. doi:10.1093/glycob/cwn046
45. Steichen AL, Binstock BJ, Mishra BB, Sharma J. C-type lectin receptor Clec4d plays a protective role in resolution of Gram-negative pneumonia. *J Leukoc Biol* (2013) **94**:393–8. doi:10.1189/jlb.1212622

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 April 2014; paper pending published: 19 May 2014; accepted: 03 June 2014; published online: 23 June 2014.

Citation: Richardson MB and Williams SJ (2014) MCL and Mincle: C-type lectin receptors that sense damaged self and pathogen-associated molecular patterns. *Front. Immunol.* **5**:288. doi: 10.3389/fimmu.2014.00288

This article was submitted to Immunotherapies and Vaccines, a section of the journal *Frontiers in Immunology*.

Copyright © 2014 Richardson and Williams. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Computational and experimental prediction of human C-type lectin receptor druggability

Jonas Aretz<sup>1,2</sup>, Eike-Christian Wamhoff<sup>1,2</sup>, Jonas Hanske<sup>1,2</sup>, Dario Heymann<sup>1</sup> and Christoph Rademacher<sup>1,2\*</sup>

<sup>1</sup> Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany

<sup>2</sup> Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Berlin, Germany

## Edited by:

Elizabeth Yuriev, Monash University, Australia

## Reviewed by:

Anthony G. Coyne, University of Cambridge, UK

Stephen James Headey, Monash University, Australia

## \*Correspondence:

Christoph Rademacher, Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, Potsdam 14424, Germany  
e-mail: christoph.rademacher@mpikg.mpg.de

Mammalian C-type lectin receptors (CTLRs) are involved in many aspects of immune cell regulation such as pathogen recognition, clearance of apoptotic bodies, and lymphocyte homing. Despite a great interest in modulating CTLR recognition of carbohydrates, the number of specific molecular probes is limited. To this end, we predicted the druggability of a panel of 22 CTLRs using DoGSiteScorer. The computed druggability scores of most structures were low, characterizing this family as either challenging or even undruggable. To further explore these findings, we employed a fluorine-based nuclear magnetic resonance screening of fragment mixtures against DC-SIGN, a receptor of pharmacological interest. To our surprise, we found many fragment hits associated with the carbohydrate recognition site (hit rate = 13.5%). A surface plasmon resonance-based follow-up assay confirmed 18 of these fragments (47%) and equilibrium dissociation constants were determined. Encouraged by these findings we expanded our experimental druggability prediction to Langerin and MCL and found medium to high hit rates as well, being 15.7 and 10.0%, respectively. Our results highlight limitations of current *in silico* approaches to druggability assessment, in particular, with regard to carbohydrate-binding proteins. In sum, our data indicate that small molecule ligands for a larger panel of CTLRs can be developed.

**Keywords:** C-type lectin receptors, druggability, inhibitor, DC-SIGN, langerin, MCL, fragment screening, NMR screening

## INTRODUCTION

Glycans are present in a large diversity on cell surfaces and are essential in many aspects of life such as embryonic development, cell–cell communication, and regulation of the immune system (1). In particular, our understanding of the role of glycans in immunobiology has grown significantly during the last decades. Three major families of secreted or membrane-bound lectins recognize carbohydrates. Complementary to other receptors of the innate and adaptive immune system, Galectins, Siglecs, and C-type lectins shape the response to incoming signals (2, 3). Among many other processes, they are involved in pathogen recognition and killing, antigen processing, and tumor progression (2, 4, 5).

Mammalian C-type lectin receptors (CTLRs) represent a large family of lectins, which is subdivided into 17 groups based on their phylogenetic relationships and domain structure (6). CTLRs are present in a variety of tissues and the glycan specificity of receptors present on cells of the innate immune system has been studied extensively. For example, they function as homing receptors on leukocytes as well as pattern recognition receptors (2, 3, 7). A particularly well-studied pattern recognition receptor is the dendritic cell-specific intercellular adhesion molecules-3-grabbing non-integrin (DC-SIGN) (8, 9). This CTLR is expressed on dendritic cells and macrophages and is involved in the recognition of a large array of pathogens such as *Mycobacterium tuberculosis*, *Leishmania*, HCV, Ebola, and HIV (3, 10–15). It was demonstrated that DC-SIGN promotes HIV *trans*-infection of T cells and has since

then drawn attention as a therapeutic target in anti-viral therapy (10, 16, 17).

Aside from interference with pathogen recognition, leukocyte homing has been a target for small molecule inhibition of CTLR function. To this end, Selectins, a group of three CTLRs, have been in the focus as anti-inflammatory drug targets since the mid-90s (18). Only recently, the glycomimetic GMI-1070 has entered clinical trials for the treatment of sickle cell anemia (19). Likewise, agonistic CTLR ligands hold promise to serve as adjuvants for immune stimulation (20). However, despite increasing interest in CTLRs as pharmacological targets, there is only a limited set of small molecule agonists or antagonists available (17). Partially, this can be attributed to the limited success of previous attempts to find lead structures from classical drug discovery campaigns.

All CTLRs share a C-type lectin domain (CTLD) that has a conserved fold with a characteristic double-loop stabilized by two disulfide bridges (7, 21). This domain is often referred to as carbohydrate recognition domain (CRD) for those CTLRs involved in glycan binding. Additional domains are frequently present and in particular, heptad-repeats and collagen-like neck domains promote oligomerization, resulting in high avidity glycan binding. In transmembrane CTLRs, CRD, and neck domain are referred to as extracellular domain (ECD). Canonical carbohydrate recognition is mediated by a calcium ion and although there are four Ca<sup>2+</sup> binding sites, only the second site (Ca<sup>2+</sup>-2) is described to be involved in coordinating glycans (21). While Ca<sup>2+</sup>-4 has

not been associated with carbohydrate binding, positive cooperative effects are observed between the other sites (22, 23). Not all potential  $\text{Ca}^{2+}$  sites are occupied in every CTLD, which reflects the fine-tuned physiological role of this interaction. For endocytic CTLRs the pH sensitivity of the heptad-repeat neck formation and  $\text{Ca}^{2+}$  coordination as well as active  $\text{Ca}^{2+}$  export from the endosome are major contributors to endosomal ligand release (23, 24). Some CTLRs bind carbohydrates in a  $\text{Ca}^{2+}$ -independent, non-canonical binding site with Dectin-1 being the prime example (25). All CRDs share a carbohydrate recognition site that is largely flat and hydrophilic. This is a consequence of glycans being highly hydrophilic themselves (17, 26). Hence, binders are also often hydrophilic and do not suffice the requirements for orally available drugs (27).

Whether a protein is a suitable candidate for drug development is of major concern during the drug discovery process. Considering the expenses involved in the development of a pharmacologically active small molecule, target selection has to be done carefully (28). The modulation of a suitable drug target with a rule of five compliant molecules should result in a therapeutic effect (29). The term druggability, however, refers to the ability of a protein to bind a drug-like ligand with high affinity and specificity (29–31). Furthermore, this interaction has to result in a modulation of the protein function. Importantly, a high druggability does not infer the protein being a good drug target. The latter definition includes a therapeutic effect induced by small molecule binding (32, 33). Methods to assess the druggability of a target protein have become good predictors prior to starting a drug discovery campaign, as low scores are indicators for a high failure rate during later stages of the project (30, 33).

The availability of structural information enables computational assessment of druggability. Limited resources are required and many computational tools have been developed to deduce druggability scores from crystallographic information (34, 35). In a two-step process, pockets on the protein surface are first identified and then scored (28, 32, 34, 36). Large sets of proteins can be analyzed and predictions have been found to correlate well with experimental data (31, 34, 37, 38). To the best of our knowledge, there are only two studies on the computational druggability assessments of glycan-binding proteins, both reporting low scores (39, 40).

Experimental assessment of target druggability can be pursued even in the absence of structural information. For this, screening of drug-like molecules in a high-throughput screening format can be performed. Previous reports on micromolar inhibitors of DC-SIGN resulting from a screening campaign highlight the success of this approach (41, 42). Alternatively, a diverse library of fragments of drug-like molecules is screened against the target. The molecular weight of these fragments ranges between 150 and 300 Da. Estimates propose that 1000 fragments can cover a similar chemical space as 10 trillion drug-sized molecules would (33). This in turn allows applying smaller libraries to test the druggability of a candidate protein (31, 33). The low complexity of fragments increases their likelihood of binding a receptor and consequently hit rates of 5–15% are regularly observed for druggable targets (31, 37, 43).

Small molecule fragments have low affinities with dissociation constants in the upper micro- to lower millimolar range. Hence, sensitive biophysical techniques are necessary to monitor this interaction and nuclear magnetic resonance (NMR) spectroscopy has established itself as one of the major techniques used for fragment screening (31, 33, 37–39, 44, 45). In particular, hit rates from NMR-based screenings have proven to be reliable measures of druggability (31, 37, 44). In ligand-observed NMR, mixtures of fragments are screened against a target and changes in NMR observables such as chemical shift, line width, and signal intensity upon binding allow hit identification. Notably, deconvolution of the fragment mixtures is not necessary. The use of fluorine atoms in drug-like fragments has proven to be instrumental (38, 46). As fluorine is rare in biological samples,  $^{19}\text{F}$  NMR spectra of fragment cocktails are not perturbed by background resonances. Moreover, the fluorine spin is highly susceptible to changes in its chemical environment and allows sensitive identification of hits.

To predict the druggability of human CTLRs, we compiled a set of 22 crystal structures and analyzed it by applying computational methods. We then chose DC-SIGN and conducted experimental fragment screening to compare these findings. Low druggability scores derived *in silico* did not match the moderate to high fragment hit rates during experimental evaluation. Hence, we expanded our screening by two additional CTLRs, namely Langerin and MCL and discovered similarly high experimental druggability estimates. Taken together, our results highlight the limitations of *in silico* druggability prediction for CTLRs while our fragment screening present promising grounds for inhibitor design against this family.

## MATERIALS AND METHODS

### STRUCTURE-BASED MULTIPLE SEQUENCE ALIGNMENT AND CONSENSUS STRUCTURE

The scope of structural data on human CTLRs was assessed using the protein family (Pfam) database (accession code: PF00059) (47). Natural killer (NK) cell lectin-like receptors were treated as a closely related, yet physiologically distinct subfamily according to the Pfam annotation and were not included in the analysis. Furthermore, CTLRs crystallized as a domain swap dimer, namely blood dendritic cell antigen 2 (BDCA-2) and mannose receptor (MR), were omitted (48, 49). Murine Dectin-1 was included in the selection as it has an unusual  $\text{Ca}^{2+}$ -independent carbohydrate-binding mode and no structural information of the human ortholog is available (25). All structures considered for analysis are listed (Table 1). If available, a structure in complex with a carbohydrate ligand was selected. Prior to the calculations, all structures were trimmed down to the respective CRD domain as inferred from the Pfam domain definition. A structure-based multiple sequence alignment was performed in molecular operating environment (MOE) (50). Pairwise root mean square deviation (RMSD) values were determined for all pairs of  $\text{C}_\alpha$  atoms unless a gap was found in one of the compared sequences. Next, a phylogenetic analysis based on the pairwise sequence similarities was conducted in R (51, 52). Hierarchical clustering was performed based on the Manhattan metric and via the complete linkage criterion. To complement the phylogenetic analysis, MOE was used

to predict a consensus structure of all CRDs. During model construction, up to 20 gaps and RMSD values of  $C_{\alpha}$  up to 10 Å were allowed for a single position in the multiple sequence alignment.

**BINDING SITE PREDICTION AND *IN SILICO* DRUGGABILITY ASSESSMENT**

Initially, CTLR structures were superposed in MOE. For superposition and the subsequent druggability assessment, physiologically relevant oligomerization states were assumed (Table 1). The EGF domains of Selectin structures were removed. The resulting files served as input data for binding site prediction with DoGSite (72). The predicted binding sites were mapped on the structure and classified into four categories following the reported nomenclature of secondary structure elements and  $Ca^{2+}$  binding sites (21): (i)  $Ca^{2+}$ -2-binding sites, (ii)  $Ca^{2+}$ -associated binding sites in long loop, (iii)  $Ca^{2+}$ -independent carbohydrate-binding sites, and (iv) other binding sites. A binding site was assigned to category (i) if the  $Ca^{2+}$ -2 ion was part of the predicted binding site. For category (ii), the criteria were less restrictive and all binding sites with residues within a 6 Å radius of either  $Ca^{2+}$ -1, 2, or 3 were included (Figure S1 in Supplementary Material). Binding sites in category (iii) are located in close proximity to the experimentally determined  $Ca^{2+}$ -independent carbohydrate-binding site. The druggability of all binding sites was scored with DoGSiteScorer (73). Finally, category (i), (ii), or (iii) binding sites that displayed the highest score for a receptor were selected and this selection served to determine a mean druggability score for the analyzed CTLRs.

**CLONING**

Codon optimized genes for DC-SIGN and human Langerin for expression in *E. coli* were purchased from Life Technologies (Carlsbad, CA, USA) and GenScript (Piscataway, NJ, USA), respectively. The DC-SIGN gene included a C-terminal TEV (tobacco etch virus) cleavage site and a Strep-tag II for affinity purification. The ECD and CRD, ranging from amino acids 62 to 404 and 250 to 404 (Figure S3 in Supplementary Material), respectively, were cloned into a pUC19 vector using primers including a T7 promoter and ribosomal binding site (RBS) upstream of the gene (Table 2). Human Langerin truncated ECD, ranging from amino acids 148 to 328, was cloned with a C-terminal TEV cleavage site and a Strep-tag II into a pET32a expression vector (EMD Millipore, Billerica, MA, USA). The MCL gene was obtained from the DNASU Plasmid Repository (HsCD00507041, Arizona State University, Phoenix, AZ, USA) and the ECD was cloned into a pUC19 vector already carrying a Strep-tag II, a T7 promoter and an RBS. For MCL ECD, amino acids ranging from 61 to 215 were used (65).

**PROTEIN EXPRESSION AND PURIFICATION**

All growth media or chemicals used for protein expression and purification were purchased from Carl Roth (Karlsruhe, Germany) if not stated otherwise. Proteins were expressed insoluble in *E. coli* BL21(DE3) (New England Biolabs, Ipswich, MA, USA) or KRX (Promega, Fitchburg, WI, USA). Precultures were grown in 50 mL Luria–Bertani (LB) medium supplemented with 100 mg L<sup>-1</sup> carbenicillin for DC-SIGN and MCL expression or 35 mg L<sup>-1</sup> kanamycin for Langerin expression at 37°C in 250 mL baffled shaking flasks at 220 rpm shaking frequency. The precultures of DC-SIGN and MCL were centrifuged (3,000 × g, 10 min,

**Table 1 | List of analyzed CTLR structures.**

Receptor	PDB code	Domain	Oligo-merization	Reference
ASGPR	1DV8	CRD	Monomer	Meier et al. (53)
CD23a	2H2T	CRD	Monomer	Wurzburg et al. (54)
Clec1b	2C6U	CRD	Monomer	Watson et al. (55)
Clec5a	2YHF	CRD	Monomer	Watson et al. (56)
Clec9a	3VPP	CRD	Monomer	Zhang et al. (57)
DC-SIGN	2XR6	CRD	Monomer	Unpublished
DC-SIGNR	1K9J	CRD	Monomer	Feinberg et al. (58)
mDectin-1	2CL8	CRD	Monomer	Brown et al. (25)
E-Selectin	1ESL	CRD–EGF	Monomer	Graves et al. (59)
EMBP	1H8U	CRD	Monomer	Swaminathan et al. (60)
Langerin	3P5F	CRD	Monomer	Feinberg et al. (61)
	3KQG	ECD	Trimer	Feinberg et al. (62)
L-Selectin	3CFW	CRD–EGF	Monomer	Unpublished
LOX-1	1YPO	ECD	Dimer	Park et al. (63)
MBP-C	1HUP	ECD	Trimer	Sheriff et al. (64)
MCL	3WHD	ECD	Monomer	Furukawa et al. (65)
Mincle	3WH3	ECD	Monomer	Furukawa et al. (65)
P-Selectin	1G1S	CRD–EGF	Monomer	Somers et al. (66)
Reg3a	1UV0	CRD	Monomer	Abergel et al. (67)
Reg1a	1QDD	CRD	Monomer	Gerbaud et al. (68)
SCARA4	2OX8	CRD	Monomer	Feinberg et al. (48)
SP-D	3IKN	ECD	Trimer	Shrive et al. (69)
Tetranectin	1HTN	CRD	Monomer	Nielsen et al. (70)
	1TN3	ECD	Trimer	Kastrup et al. (71)

**Table 2 | Primer sequences used for cloning.**

Primer	Sequence
DC-SIGN ECD F	GCCGCCTCTAGAGAGTAATACGACTCACTATAGGGACT AGAGAAAGAGGAGAGAAACTAGATGGC- CAAAGTTCGAGCAGCATT
DC-SIGN CRD F	GCCGCCTCTAGAGGAGTAATACGACTCACTATAGGGAC TAGAGAAAGAGGAGAGAAACTAGATGGCT- GAACGTCTGTGTCATCCGTG
Langerin F	GGTGGTCATATGGCC- TCGACGCTGAATGCCAGATTCCGG
Langerin R	ACCACC-AAGCTTTTATTTTCAAACGCGGATG
MCL F	GGCGGCGCCGGC-CATGCAAAGCTCAAATGCAT
MCL R	GCCGCCCTGCAG-GTTCAATGTTGTTCCAGGTA

4°C), the supernatant was discarded, and the sediment was resuspended in 500 mL LB medium supplemented with 50 mg L<sup>-1</sup> carbenicillin. The cells were afterwards cultivated at 37°C in 2.5 L baffled shaking flasks at 220 rpm shaking frequency. Protein expression was induced with 1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) at OD<sub>600</sub> = 0.4–0.6 for additional 4 h at 37°C. The preculture of Langerin trECD was diluted directly to OD<sub>600</sub> = 0.1 into 500 mL of LB medium supplemented with 35 mg

L<sup>-1</sup> kanamycin, 0.01% D-glucose, and 0.05% L-rhamnose for autoinduction of expression. Bacteria were harvested (4,000 × g, 20 min, 4°C), frozen, and resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 10 mM magnesium chloride, 0.1% Triton X-100, 4 mg lysozyme [Sigma-Aldrich, St. Louis, MO, USA] and 500 U DNaseI (Applichem, Darmstadt, Germany) per gram of wet biomass] and incubated on ice for 4 h. Inclusion bodies were harvested by centrifugation (10,000 × g, 10 min, 4°C) and washed thrice with 20 mL washing buffer (50 mM Tris-HCl, pH 8.0, 4 M urea, 500 mM sodium chloride, 1 mM EDTA) to remove soluble proteins.

For DC-SIGN ECD and Langerin ECD purification, the washed inclusion bodies were resuspended and denatured in 40 mL denaturation buffer (6 M guanidine hydrochloride, 100 mM Tris-HCl, pH 8.0, 1 mM DTT) and incubated at 30°C for 1 h or at 4°C over night, following a centrifugation (42,000 × g, 1 h, 4°C). The denatured inclusion bodies were slowly diluted threefold with cold binding buffer (TBS, pH 7.8 with 25 mM calcium chloride), supplemented with 1 mM reduced glutathione (GSH, Applichem) and 0.1 mM oxidized glutathione (GSSG, Applichem), and afterwards dialyzed twice against 2 L of this buffer for 24 h at 4°C. After another 2 L dialysis against binding buffer, proteins were purified according to previously published protocols using a mannan agarose affinity chromatography (Sigma-Aldrich) (74).

The washed inclusion bodies of DC-SIGN CRD were resuspended and denatured in 10 mL denaturation buffer and incubated at 30°C for 1 h or at 4°C over night, following a centrifugation (42,000 × g, 1 h, 4°C). The solubilized inclusion bodies in the supernatant were refolded by rapid dilution into 50 mL of cold refolding buffer (100 mM Tris-HCl, pH 8.0, 1 M L-arginine, 150 mM sodium chloride, 120 mM sucrose) while stirring at 4°C. After 2 days, protein solution was dialyzed against 2 L of cold buffer W (100 mM Tris-HCl, pH 8.0, 150 mM sodium chloride, 1 mM EDTA) and aggregated protein was removed by centrifugation (42,000 × g, 1.5 h, 4°C). The protein was purified using a Streptactin affinity chromatography (IBA, Goettingen, Germany) according to the manufacturer's instructions.

MCL refolding and purification was performed according to Furukawa and coworkers introducing minor changes in the protocol. Briefly, purification was performed via Streptactin affinity chromatography after dialysis against 2 L of buffer W.

## FRAGMENT LIBRARY

Fragments were selected from a pool of commercially available compounds from different manufacturers (Sigma-Aldrich, St. Louis, MO, USA; KeyOrganics, Camelford, UK; ACB Blocks, Toronto, ON, Canada; Santa Cruz Biotechnology, Santa Cruz, CA, USA; Vistas-MLab, Moscow, Russia; LifeChemicals, Kyiv, Ukraine; Alfa Aesar, Ward Hill, MA, USA; TCI, Tokyo, Japan; Apollo Scientific, Stockport, UK) using chemoinformatic tools as implemented in MOE and KNIME (75). Only compounds with <23 non-hydrogen atoms and at least one ring were PAINS-filtered and consecutively included in the diversity selection (76). Fragment selection was based on normalized moments of inertia for shape diversity, Tanimoto coefficient (<0.8) using MACCS fingerprint for chemical diversity and scaffold diversity was ensured following definitions given by Murcko and coworkers (77, 78).

Maximum pairwise similarities were calculated in MOE using three-point pharmacophore-based fingerprints (GpiDAPH3) as descriptors and Tanimoto coefficient as similarity metric. The same descriptor was used to assess the chemical complexity of the fragments (31).

Fragments were dissolved in d<sub>6</sub>-DMSO (Euriso-Top, Saint-Aubin, France) to 100 mM stock solutions under a nitrogen atmosphere in Matrix plates (Thermo Scientific, Waltham, MA, USA) followed by shaking at room temperature for 18 h at 140 rpm. Fragments were stored at -20°C. Next, each fragment was dissolved under nitrogen atmosphere at 1 mM in 500 µL 10 mM deuterated phosphate buffer, pH 7.0, containing 50 µM d<sub>4</sub>-TSP [(3-(trimethylsilyl)-2,2',3,3'-tetra-deuteriopropionic acid, Sigma-Aldrich], 50 µM TFA (trifluoroacetic acid, Sigma-Aldrich), and 0.01% sodium azide (Carl Roth). A <sup>19</sup>F and <sup>1</sup>H NMR spectrum of each fragment was recorded for quality control. All NMR studies were measured at 298 K in Norell SP5000-7 5 mm tubes (Norell, Landisville, NJ, USA) on a Varian PremiumCOMPACT 600 MHz spectrometer equipped with an oneNMR probe (Agilent, Santa Clara, CA, USA) with TSP and TFA as internal references. All spectra were analyzed in MestReNova 9.0.0 (Mestrelab Research, Santiago de Compostela, Spain) for identity and for solubility in D<sub>2</sub>O of at least 200 µM. Substances, that did not fulfill these quality criteria (17%), were removed from the library. Chemical shifts were used to design 8 screening mixtures consisting of 36 compounds each. A genetic algorithm was used to solve the optimization problem of mixture prediction (unpublished data). Prior to screening, all mixtures were analyzed in <sup>19</sup>F NMR spectra after 18–24 h incubation at room temperature to ensure stability of the mixtures. Compounds experiencing precipitation or changes in chemical shift were removed from the following experiments. The quality control left 281 compounds (83%) to be prepared in mixtures of 100 µM compound each, 100 µM TFA, 150 mM sodium chloride in 20 mM Tris-HCl, pH 7.8, in 20% D<sub>2</sub>O (Euriso-Top) that were stored at -20°C as aliquots until used.

## NMR SCREENING

All protein samples were prepared at 20 µM of final concentration in 20 mM Tris-HCl, pH 7.8, with 150 mM sodium chloride and 1 mM EDTA and mixed 1:1 with the screening mixture aliquots resulting in a final protein and compound concentration of 10 and 50 µM, respectively, in 500 µL final volume. Fluorine spectra were recorded with a spectral width of 140 ppm and a transmitter offset at -120 ppm, acquiring 128 scans, with an acquisition time of 0.8 and 2 s relaxation time. T<sub>2</sub>-filtered spectra were recorded using a CPMG pulse sequence with a 180° pulse repetition rate of 50 Hz and duration of 1.0 s using same acquisition and relaxation times (79, 80). Two CPMG spectra were recorded per mixture to cover the full spectral width. A spectrum ranging from -50 to -100 ppm and from -100 to -150 ppm was recorded with 96 and 256 scans, respectively. Screening was performed first in the presence and absence of protein including 0.5 mM EDTA. Next, calcium chloride was added to a final concentration of 10 mM and measurements were repeated. All spectra were analyzed for changes in peak intensity and chemical shift. As an additional quality control, frequent hitters identified during unrelated screening campaigns were removed.

## SPR FOLLOW-UP SCREENING

All surface plasmon resonance (SPR) measurements were performed on a Biacore® T100 (GE Healthcare, Chalfont St. Giles, UK) with a flow-rate of  $10 \mu\text{L min}^{-1}$  using HBS-P buffer [10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.6, 150 mM sodium chloride, 0.05% Tween-20] at 298 K. DC-SIGN ECD was immobilized on a CM7 Series S sensor chip in a density of 3317 RU using 0.2 M EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, Sigma-Aldrich) and 0.05 M NHS (*N*-hydroxysuccinimide, Merck, Hohenbrunn, Germany) as coupling reagents. The activated surface was saturated with 1 M ethanolamine (Sigma-Aldrich), pH 8.5, after immobilization. The reference flow cell was treated in the same manner without immobilizing protein. Prior to measurements, the solubility of each compound in SPR buffer was determined by recording absorption spectra at different concentrations between 400 and 800 nm in clear 96-well plates (Nalge Nunc International, Penfield, NY, USA) in a SpectraMax M5 plate reader (Molecular Devices, Sunnyvale, CA, USA). During SPR measurements, fragments were injected for 30 s following a dissociation time of 120 s at  $10 \mu\text{L min}^{-1}$  flow-rate omitting regeneration as fast off-rates were observed for all ligands. To estimate the apparent affinity of a compound, at least three dilutions between 0.1 and 1 mM depending on the solubility were run in triplicates, blanking the data against a corresponding DMSO control. A positive control was included during screening to ensure stability of the sensorgrams. A 1:1 binding model was applied for data fitting:

$$RU = RU_{\max} \frac{[L]}{K_{D, \text{app}} + [L]} \quad (1)$$

with the fragment concentration  $[L]$ , the measured relative response units RU, the apparent dissociation constant  $K_{D, \text{app}}$ , and the maximal relative response units  $RU_{\max}$  using Origin8.6Gpro (OriginLab, Northampton, MA, USA). The maximal relative response units were estimated using:

$$RU_{\max} = A \cdot RU_{\text{immobilized}} \frac{MW_{\text{compound}}}{MW_{\text{protein}}} \quad (2)$$

with the immobilization level of protein  $RU_{\text{immobilized}}$ , the molecular weight of the compound and protein  $MW_{\text{compound}}$  and  $MW_{\text{protein}}$ , respectively, and the remaining activity of the protein on the chip A. The latter was determined to be 0.6 using 4 as positive control (Figure S4 in Supplementary Material). The apparent affinity constant for each compound was determined under two conditions, either in the presence of 0.5 mM EDTA or 2 mM calcium chloride included in the running and sample buffer. Ligand efficiencies (LE) were calculated applying

$$LE = - \frac{RT \ln (K_{D, \text{app}})}{HA} \quad (3)$$

using the apparent dissociation constant  $K_{D, \text{app}}$ , the temperature T, the gas constant R, and the number of non-hydrogen atoms HA (81).

## RESULTS

### STRUCTURE-BASED SEQUENCE ALIGNMENT IDENTIFIES CANONICAL CARBOHYDRATE-BINDING SITES

A comparative framework between the CTLRs served as the starting point of our druggability prediction. To this end, a structure-based sequence alignment was performed for 22 CRDs (Figure S3 in Supplementary Material). With an average of 41%, the global sequence similarity within the set of receptors is low. It spans a range from 26 to 86% (Figure 1A). A phylogenetic analysis based on this alignment yields a dendrogram that resembles the canonical classifications of CTLRs, in particular with respect to the correct assignment of members of the groups II, III, IV, V, and VII (1). Collectin-12 deviates from this classification, as it is part of the group II cluster. Moreover, Tetranectin and eosinophil major basic protein (EMBP) are the only representatives of group IX and XII used in this study. Both display elevated distances to other branches. EMBP and Tetranectin as well as Clec9a, Lox-1, Clec1b, and Reg1a have been reported to interact with non-carbohydrate ligands and all of these CTLRs were assigned to cluster B. Strikingly, CRDs known to recognize carbohydrates via the  $\text{Ca}^{2+}$ -2-binding site are exclusively present in cluster A (Figure 1A).

### CONSENSUS STRUCTURE PREDICTION REVEALS ELEVATED STRUCTURAL VARIABILITY IN THE LONG LOOP

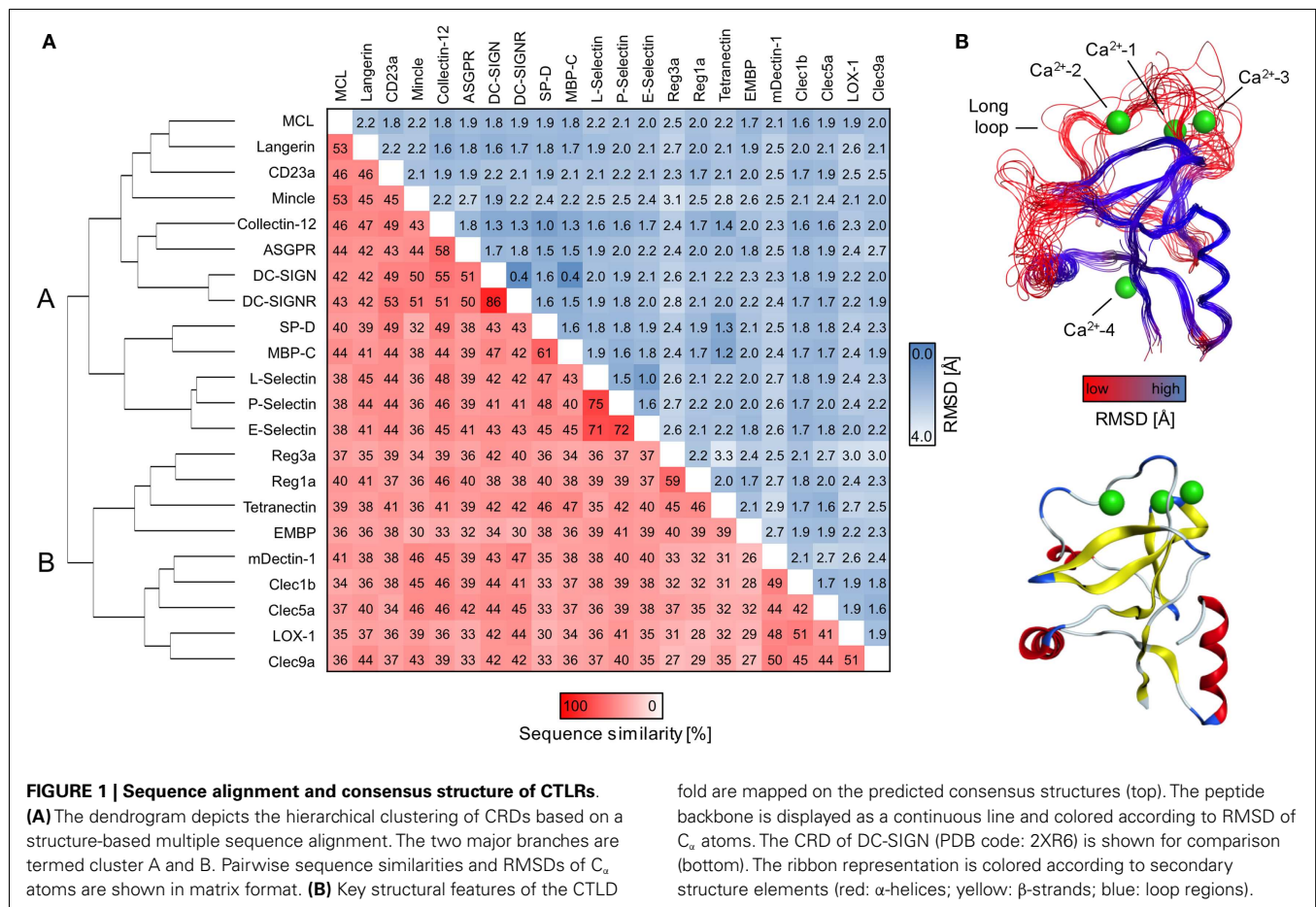
Contrasting the low global sequence similarity, the overall structure of the CTLD is highly conserved. RMSD values of  $C_{\alpha}$  atoms obtained from the structure-based multiple sequence alignment are uniformly low and do not exceed 3.2 Å (Figure 1A). To visualize the conservation of the domain architecture, we calculated a consensus structure (Figure 1B). While the core of the CTLD displays only minor deviations, a higher level of structural variability characterizes the two loop regions. The long loop is of particular interest as it harbors the  $\text{Ca}^{2+}$ -1, -2, and -3 sites and thus plays a fundamental role in  $\text{Ca}^{2+}$ -dependent carbohydrate recognition (21).

### COMPUTATIONAL ANALYSIS PREDICTS LOW DRUGGABILITY FOR THE MAJORITY OF CTLRs

The initial identification of binding sites with DoGSite yielded between three and nine sites for CRDs and 9–19 for ECDs. Next, DoGSiteScorer was applied to calculate druggability scores. In the scoring scheme of this program, scores over 0.5 are indicative of a druggable binding site (73). At least one site that meets this criterion is found for the majority of the analyzed CTLRs (Figure 2A). However, targeting these sites with drug-like molecules will not necessarily exert an effect on the physiological function of the respective CTLR.

We propose that binding sites in proximity to  $\text{Ca}^{2+}$  ions located in the long loop region are relevant to carbohydrate recognition. Therefore, we assumed that small molecule-binding to these sites potentially modulates CTLR function. To this end, binding sites were assigned to four categories: (i)  $\text{Ca}^{2+}$ -2-dependent, (ii)  $\text{Ca}^{2+}$ -associated binding sites, (iii)  $\text{Ca}^{2+}$ -independent carbohydrate-binding sites, and (iv) other binding sites (Figure 2A).  $\text{Ca}^{2+}$ -associated binding sites (i, ii) were identified by DoGSite for all CTLRs coordinating a  $\text{Ca}^{2+}$ -2 ion except for Mincle and the Langerin trimer. Experimentally determined  $\text{Ca}^{2+}$ -independent





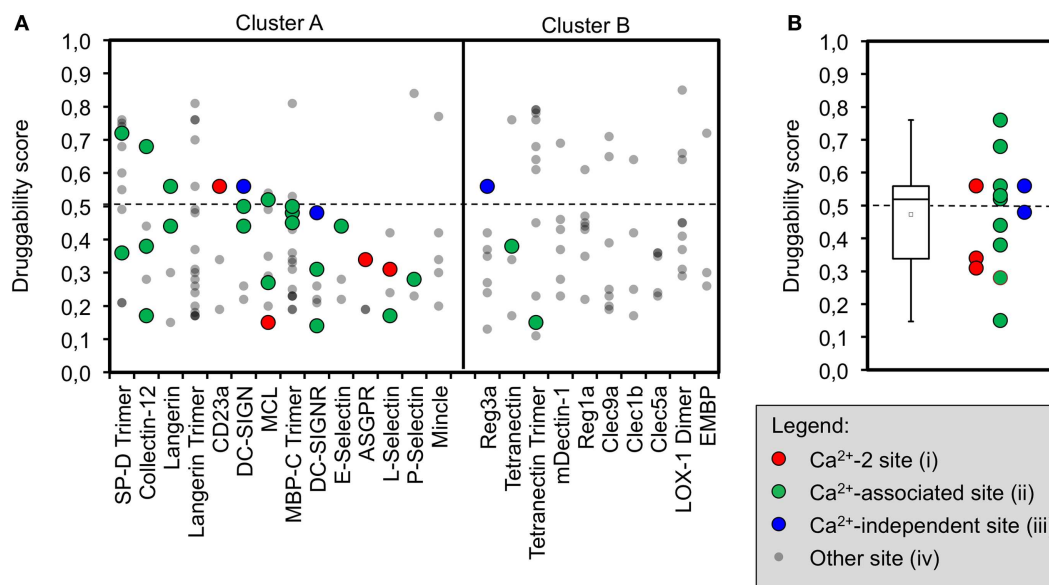
carbohydrate-binding sites (iii) were identified for DC-SIGN, DC-SIGNR, and Reg3a (58, 82). The existence of a single druggable site is sufficient to render a target druggable. Accordingly, for each CTLR, sites assigned to categories (i) and (ii) displaying the highest score were selected for statistical analysis and a mean druggability score of 0.47 was calculated (**Figure 2B**). This classifies CTLRs as “difficult” or even “undruggable” targets (73). Notably, individual receptors such as SP-D and Collectin-12 possess favorable pockets in the long loop region. Other targets such as E-Selectin display druggability values well below the mean.

#### FRAGMENT SCREENING REVEALS HIGH HIT RATES AGAINST DC-SIGN, LANGERIN, AND MCL

The existence of pockets on the surface of a receptor that are suitable to accommodate drug-like ligands can be assessed experimentally using fragment screening. The resulting hit rate serves as a predictor for druggability. Therefore, we composed a chemical library of fragments to be used in a homogeneous, label-free NMR-based screening assay. All fragments carry a fluorine atom, which allows for  $^{19}\text{F}$  NMR spectroscopy-based assessment of fragment binding. After quality control, 281 fragments were available for screening in 8 mixtures of maximum 36 fragments. The fragment library displays high shape and chemical diversity (**Figures 3A,B**).

DC-SIGN CRD and ECD were screened against the fragment library using  $^{19}\text{F}$  and  $T_2$ -filtered  $^{19}\text{F}$  NMR spectra. Fragment binding to DC-SIGN was observed monitoring changes in chemical shift, line broadening, and  $T_2$  relaxation. Moreover, three spectra were recorded per fragment mixture. First, a spectrum was recorded in the absence of protein to exclude false positives such as  $\text{Ca}^{2+}$  chelators. The second spectrum was acquired in the presence of  $10\ \mu\text{M}$  protein to monitor fragment binding. Finally,  $\text{Ca}^{2+}$  was added in excess to the protein-fragment mixture, hypothesizing that metal binding to DC-SIGN modulates interaction of those fragments that are good candidates for inhibition of carbohydrate recognition (*vide supra*). Hits for DC-SIGN CRD and ECD were combined and frequent hitters were removed. Consequently, we identified 38 hits (13.5%) from mixtures binding to DC-SIGN in a  $\text{Ca}^{2+}$ -dependent manner (**Figure 3C**). Out of these hits, 16 were found in both screenings and 21 hits were identified only during the CRD screening. Only one fragment was found while the ECD was used for screening.

To further validate these hits, SPR spectroscopy was employed as an orthogonal biophysical assay. This method not only detects binding of small molecules to macromolecules, but also allows for the determination of equilibrium dissociation constants. DC-SIGN ECD was immobilized on the chip surface and two experimental setups were utilized to differentiate  $\text{Ca}^{2+}$ -mediated fragment binding from  $\text{Ca}^{2+}$ -fragment competition. Fragments



**FIGURE 2 | Computational druggability prediction of CTLRs.**

**(A)** A comprehensive account of all binding sites predicted by DoGSite is given. CTLRs were grouped according to the structure-based multiple sequence alignment. A druggability score of <0.5 is indicative of an undruggable binding site (dotted line). Binding sites are colored according to

the corresponding category (red: Ca<sup>2+</sup>-2-binding sites (i), green: Ca<sup>2+</sup>-associated binding sites in long loop region (ii), blue: previously reported Ca<sup>2+</sup>-independent carbohydrate-binding sites (iii), gray: other binding sites). **(B)** A boxplot of the highest druggability scores of categories (i) to (iii) of each CTLR is shown. Tukey-style representation was chosen.

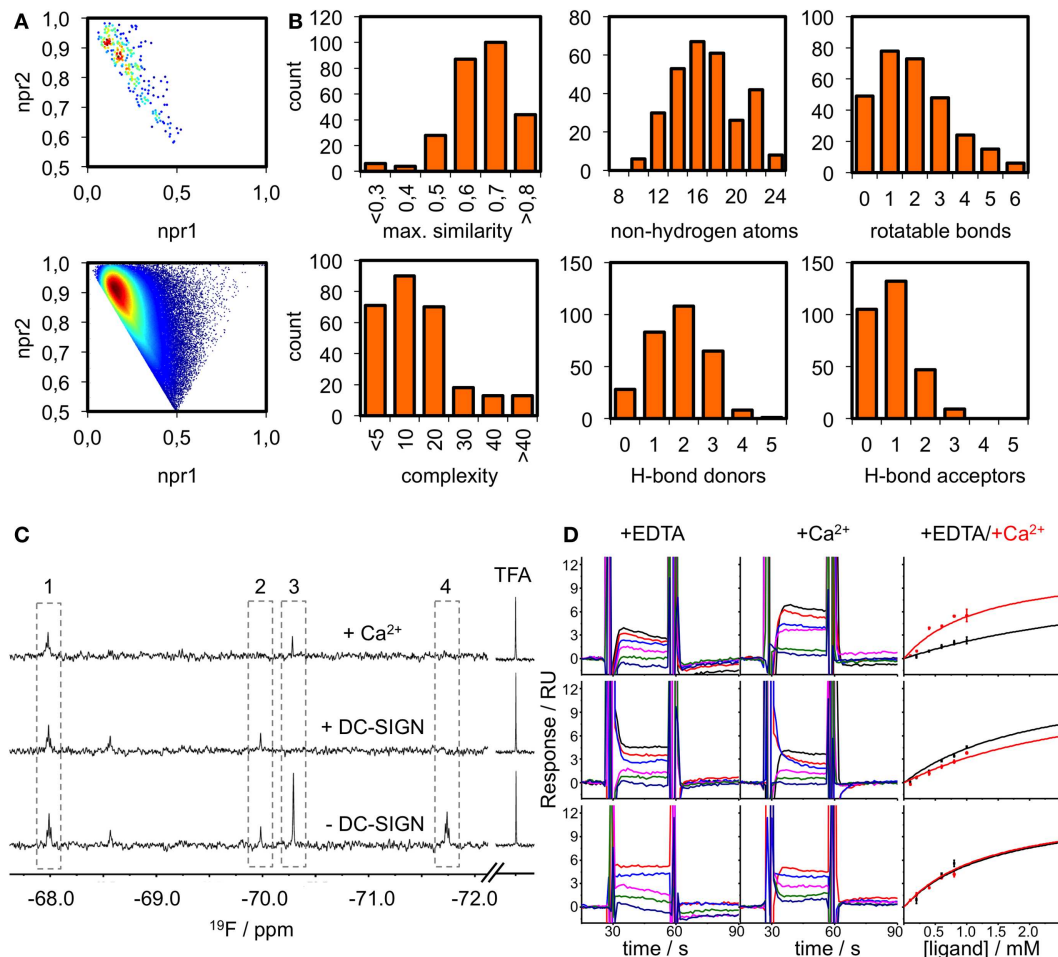
were injected either in the presence of 0.5 mM EDTA or 2 mM calcium chloride (**Figure 3D**), confirming a 1:1 binding model for 18 fragments (47%). Five fragments (13%) bound with a higher stoichiometry, 3 experienced no change in response in presence or absence of Ca<sup>2+</sup> (8%), and 12 fragments (32%) did not give rise to detectable signals. The highest affinities measured were in the upper micromolar to lower millimolar range (0.6 mM <  $K_{D,app}$  > 1.3 mM). Of the 18 fragments confirmed by SPR, 9 showed increase in affinity upon Ca<sup>2+</sup> addition and 9 displayed competitive behavior. Moreover, fragments similar to substructures of an already published submicromolar DC-SIGN inhibitor were identified (41, 42) (**Figure 4**). While fragments 1 and 2 bound competitive with the polysaccharide mannan in a <sup>19</sup>F NMR competition assay, fragment 3 showed no such behavior upon addition of the natural carbohydrate ligand of DC-SIGN (data not shown).

In light of our computational analysis, we were surprised to find such a high fragment hit rate for DC-SIGN, and decided to expand our <sup>19</sup>F NMR-based druggability prediction against the ECDs of two further CTLRs. We decided to screen our fragment library against Langerin being sufficiently distant to DC-SIGN in our structural sequence alignment (**Figure 1A**). To compare these findings to a CTLR more closely related to Langerin, we also included MCL in our analysis. Both proteins were expressed in *E. coli* and screened following the same protocol as for DC-SIGN. Again, Ca<sup>2+</sup> was utilized as a competitor (Figures S5A,B in Supplementary Material) and several hits associated with Ca<sup>2+</sup> binding were identified (**Table 3**). The pairwise overlap between the three CTLRs was low and none of the fragment hits bound to all CTLRs (Figure S5C in Supplementary Material).

## DISCUSSION

In this report, we assessed the potential of human CTLRs to be targeted with drug-like molecules. Therefore, we explored the ability of a set of CTLRs to accommodate inhibitors to modulate the receptor–carbohydrate interaction. This druggability prediction is an important part of the decision on whether a drug discovery campaign should be pursued (28–30). Despite a large body of recent research highlighting the importance of CTLRs in immune cell regulation, pathogen uptake, and as targets for adjuvants, only a few drug-like molecules have been developed for the CTLR family (2, 16, 17, 20). Herein, we aimed to rationalize why these receptors are considered challenging targets.

To start our investigations, CTLR druggability was predicted by computational methods. No data focusing on CTLRs are available and more general reports on glycan-binding proteins presented low druggability scores (39, 40). Unfortunately, the exact structures were either not disclosed or highly redundant and no CTLR was explicitly included. We assembled a set of 21 human CTLRs, and the murine Dectin-1. The latter was included as a reference as it is a well-studied CTLR and harbors a potential non-canonical, calcium-independent carbohydrate recognition site. The druggability prediction was performed using DoGSiteScorer, recently released software to predict the druggability of protein targets based on structural and physicochemical properties (73). Here, potential pockets on the protein surface were identified first, and then scored according to their physicochemical properties. Major determinants of druggability are depth, volume, and amino acid composition of the pocket (28, 32, 34, 36, 73). Generally, highly hydrophilic binding sites are considered undruggable (36).



**FIGURE 3 | Fragment library and screening against DC-SIGN.**

(A) Molecular shape distribution of the 281 fluorinated fragments was assessed using normalized moments of inertia plots of the fragment library (top) and molecules from the ZINC database fragment set (bottom). (B) The diversity of the library is displayed for a selection of molecular descriptors. Histograms of the maximum pairwise similarity, number of non-hydrogen atoms, number of rotatable bonds, molecular complexity, and number of hydrogen bond donors/acceptors are shown. (C) Examples from the  $^{19}\text{F}$  NMR screen using  $T_2$ -filtered spectra ( $T = 1.0$  s,  $\nu_{\text{CPMG}} = 50$  Hz) showing

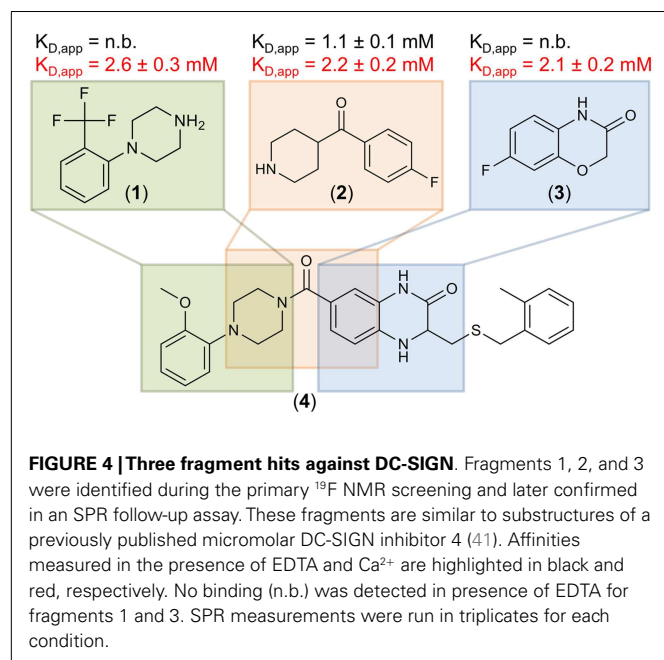
compounds that do not bind (1), bind  $\text{Ca}^{2+}$ -dependently (2), are competed by  $\text{Ca}^{2+}$  (3) and are binding at another binding site (4). (D) Sensorgrams for three examples from the SPR follow-up are shown. DC-SIGN ECD was immobilized on a CM7 chip. A one-site-binding model was used for fitting the data. This exemplifies compounds that experience increased (top), decreased (middle), or no alteration of affinity (bottom) in the presence of  $\text{Ca}^{2+}$ . Data were extracted from regions of the sensorgram not perturbed by injection peaks. SPR sensorgrams are representatives of three independent measurements.

Between three and nine binding sites were identified for CRDs, which is in accordance with values reported for other protein families (32). For Langerin, MBP-C, and Tetranectin, data on the homo-trimeric form were available. Here, the algorithm identified more potential sites, which is not surprising due to the larger surface area and symmetry of the assemblies (Figure 2A). Yet, targeting this initial set of binding sites does not necessarily interfere with carbohydrate recognition. Therefore, we categorized pockets according to their potential to modulate glycan binding. We argue that a druggable pocket located in close proximity of the long loop renders it a potential binding site for an inhibitor. The loop exhibits considerable movement in the absence of calcium as observed for other CRDs (65, 67, 83, 84) and adjacent sites have been proposed to communicate with the primary carbohydrate

recognition site (22, 23). Four categories of sites were defined out of which only two, namely categories (i) and (ii), are either directly or indirectly associated with calcium ion binding.

The success-rate of detecting the canonical  $\text{Ca}^{2+}$ -2 site (i) was low. Only 4 out of the 14 structures known to harbor such a site were identified (Figure 2A). This low number reflects a limitation of the employed pocket prediction, potentially due to shallow architecture of the  $\text{Ca}^{2+}$ -2 sites. The low druggability score of the successfully identified  $\text{Ca}^{2+}$ -2 sites corroborates this finding. Overall, these findings suggest that identification of carbohydrate recognition sites with computational algorithms such as DoGSite is challenging (*vide infra*).

Moreover, we analyzed a larger panel of sites associated with either the  $\text{Ca}^{2+}$ -1, -2, or -3 site, summarized in category (ii)



**Table 3 | Hit rates for three C-type lectin receptors from  $^{19}\text{F}$  NMR screening against a library of 281 compounds and hits confirmed by SPR for DC-SIGN.**

C-type lectin receptor	$\text{Ca}^{2+}$ -associated hits	Validated hits by SPR
DC-SIGN	38 (13.5%)	18 (6.4%)
Langerin	44 (15.7%)	
MCL	28 (10.0%)	

(Figure 2A). The criteria of this category were less stringent and based on an extended definition of sites potentially interfering with carbohydrate binding. Again, druggable sites were sparse. Collectin-12 and SP-D, both members of the Collectin group (CTLR group III), represent notable exceptions. Furthermore, our data on Langerin, for which monomer and trimer were analyzed side by side, highlight that subtle changes in the long loop region upon oligomerization abrogate the recognition of these sites by DoGSite (62).

Low scores for category (ii) sites are also found for members of cluster B of the sequence alignment. This cluster exclusively contains CTLRs not known to bind carbohydrates with their  $\text{Ca}^{2+}$ -2 site (Figure 2A). The  $\text{Ca}^{2+}$ -independent carbohydrate-binding sites of category (iii) found for Reg3a (group VII) is located in other regions of the CRD fold and has druggability scores of 0.56, predicting this CTLR to be challenging (82). Overall, only a few members of the CTLR family were predicted to be druggable (Figure 2B), which is in line with previous reports on glycan-binding proteins (39, 40).

To substantiate the computational studies, a  $^{19}\text{F}$  NMR-based fragment screening against one of the analyzed CTLRs was conducted. We chose DC-SIGN because as a viral uptake receptor it is of pharmacological interest and has been targeted in a high-throughput screening (41). While the successful HTS was

already an indicator of DC-SIGN being amenable to fragment binding, the low druggability assessment by our computational analysis predicted a low hit rate of fragments interfering with any of the three DC-SIGN calcium sites. To our surprise, a high hit rate of 13.5% of the fragments from our library bound to DC-SIGN in  $\text{Ca}^{2+}$ -associated sites during the NMR screening. The follow-up screening via SPR validated 18 (47%) of these fragments, a value not unusual for these two assay systems (85). Hits that were not validated by the SPR screening were either superstoichiometric binders (13%), not competitive with  $\text{Ca}^{2+}$  (8%), or had affinities below the detection limit of the SPR assay. The latter can be attributed to the high sensitivity of  $^{19}\text{F}$  NMR as a primary screen (38). Together, NMR and SPR result in a hit rate of 6.4%, which is in the expected range for fragment-based screenings and does not suggest a low likelihood to bind drug-like molecules (31, 37, 43, 86).

We performed the primary NMR screen against the CRD and the tetrameric ECD of DC-SIGN. Notably, only one fragment was uniquely identified during the screening of the ECD compared to 21 in the CRD screening. Conversely, many fragments binding to the ECD were later discovered to be false positives, such as frequent hitters from unrelated screening campaigns against non-CTLR targets. Hence, we conclude that screening for inhibitors has a lower false positive rate in absence of the neck region of DC-SIGN.

Another indicator for the validity of our screen to discover fragments inhibiting carbohydrate binding to DC-SIGN was the identification of the three fragments 1, 2, and 3. These hits are similar to substructures of the previously reported micromolar DC-SIGN inhibitor 4 (Figure 4) (41). In this respect, four has been shown to compete with carbohydrate binding and antagonized the DC-SIGN-mediated cell adhesion and particle uptake (41, 42). Direct competition between four and the three fragments was hampered by direct interaction of the fragments with four in absence of DC-SIGN (data not shown). Thus, mannan was employed to compete with fragments 1–3 and resulted in reproducible competition with fragments 1 and 2 (data not shown). Although, fragment 3 did not experience competition with the natural ligand, it can be speculated that it is associated with the binding site, as recognition was detected in SPR only in presence of  $\text{Ca}^{2+}$  (Figure 4). Moreover, other fragment hits showed even higher LE ranging from 0.30 to 0.37, which is a good starting point for further fragment evolution. A subsequent expansion of our  $^{19}\text{F}$  NMR-based screening to Langerin and MCL, also revealed similarly high hit rates (Table 3). Following up on these initial hits is subject of current research in the laboratory.

These encouraging experimental results are in contrast to our computational predictions. We attribute this conflict to the limitations of the DoGSiteScorer algorithm, which on the one hand is not parameterized for carbohydrate or metal binding sites (72) and on the other does not account for protein flexibility. Currently, there is no single software for druggability prediction available that is able to overcome these limitations.

Throughout the experimental evaluation, we employed competition with calcium ions as an indicator for the inhibition of carbohydrate recognition. We assumed the existence of allosteric



sites originating from the flexibility of the long loop and cooperativity between the adjacent sites as previously described for other CTLRs (22, 23, 65, 67, 83, 84). In this context, it should be noted that accounting for conformational dynamics is recognized as a particular challenge for the development of improved algorithms (34).

To summarize, we report high *in silico* druggability scores for group III and V CTLRs as well as high experimental hit rates from fragment screenings against group II CTLRs. These data stand alongside with a successful drug design campaign that has already been launched against group IV CTLRs (19). Hence, we conclude that our data, while highlighting the limitations of current computational methods, support the assessment of CTLRs as suitable targets for drug-like molecules.

## ACKNOWLEDGMENTS

The authors thank the Max Planck Society, the German Research Foundation (DFG, Emmy Noether program RA1944/2-1), and the Chemical Industry Fund for financial support. Dr. Andrea Volkamer is gratefully acknowledged for support using DoGSiteScorer and Olaf Niemeyer for NMR technical support. Jonas Hanske is supported by a fellowship from the Chemical Industry Fund and Eike Christian Wamhoff acknowledges funding from the Collaborative Research Centre 765. We thank Prof. Dr. Peter H. Seeberger for support and helpful discussions.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/Journal/10.3389/fimmu.2014.00323/abstract>

## REFERENCES

- Varki A. *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (2009).
- van Kooyk Y, Rabinovich GA. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat Immunol* (2008) 9:593–601. doi:10.1038/ni.f.203
- Geijtenbeek TB, Gringhuis SI. Signalling through C-type lectin receptors: shaping immune responses. *Nat Rev Immunol* (2009) 9:465–79. doi:10.1038/nri2569
- Fuster MM, Esko JD. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat Rev Cancer* (2005) 5:526–42. doi:10.1038/nrc1649
- Stowell SR, Arthur CM, Dias-Baruffi M, Rodrigues LC, Gourdine JP, Heimbürg-Molinari J, et al. Innate immune lectins kill bacteria expressing blood group antigen. *Nat Med* (2010) 16:295–301. doi:10.1038/nm.2103
- Cummings RD, McEver RP. C-type lectins. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (2009). p. 439–58.
- Weis WI, Taylor ME, Drickamer K. The C-type lectin superfamily in the immune system. *Immunol Rev* (1998) 163:19–34. doi:10.1111/j.1600-065X.1998.tb01185.x
- Curtis BM, Scharnrowske S, Watson AJ. Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. *Proc Natl Acad Sci U S A* (1992) 89:8356–60. doi:10.1073/pnas.89.17.8356
- Geijtenbeek TB, Torensma R, Van Vliet SJ, Van Duijnhoven GC, Adema GJ, Van Kooyk Y, et al. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* (2000) 100:575–85. doi:10.1016/S0092-8674(00)80693-5
- Geijtenbeek TB, Kwon DS, Torensma R, Van Vliet SJ, Van Duijnhoven GC, Middel J, et al. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* (2000) 100:587–97. doi:10.1016/S0092-8674(00)80694-7
- Appelmelk BJ, Van Die I, Van Vliet SJ, Vandenbroucke-Grauls CM, Geijtenbeek TB, Van Kooyk Y. Cutting edge: carbohydrate profiling identifies new pathogens that interact with dendritic cell-specific ICAM-3-grabbing nonintegrin on dendritic cells. *J Immunol* (2003) 170:1635–9. doi:10.4049/jimmunol.170.4.1635
- Lozach PY, Lortat-Jacob H, de Lacroix de Lavalette A, Staropoli I, Foug S, Amara A, et al. DC-SIGN and L-SIGN are high affinity binding receptors for hepatitis C virus glycoprotein E2. *J Biol Chem* (2003) 278:20358–66. doi:10.1074/jbc.M301284200
- Simmons G, Reeves JD, Grogan CC, Vandenbergh LH, Baribaud F, Whitbeck JC, et al. DC-SIGN and DC-SIGNR bind ebola glycoproteins and enhance infection of macrophages and endothelial cells. *Virology* (2003) 305:115–23. doi:10.1006/viro.2002.1730
- Tailleux L, Schwartz O, Herrmann JL, Pivert E, Jackson M, Amara A, et al. DC-SIGN is the major *Mycobacterium tuberculosis* receptor on human dendritic cells. *J Exp Med* (2003) 197:121–7. doi:10.1084/jem.20021468
- Sancho D, Reis e Sousa C. Signaling by myeloid C-type lectin receptors in immunity and homeostasis. *Annu Rev Immunol* (2012) 30:491–529. doi:10.1146/annurev-immunol-031210-101352
- van Kooyk Y, Geijtenbeek TB. DC-SIGN: escape mechanism for pathogens. *Nat Rev Immunol* (2003) 3:697–709. doi:10.1038/nri1182
- Ernst B, Magnani JL. From carbohydrate leads to glycomimetic drugs. *Nat Rev Drug Discov* (2009) 8:661–77. doi:10.1038/nrd2852
- Kansas GS. Selectins and their ligands: current concepts and controversies. *Blood* (1996) 88:3259–87.
- Chang J, Patton JT, Sarkar A, Ernst B, Magnani JL, Frenette PS. GMI-1070, a novel pan-selectin antagonist, reverses acute vascular occlusions in sickle cell mice. *Blood* (2010) 116:1779–86. doi:10.1182/blood-2009-12-260513
- Lang R, Schoenen H, Desel C. Targeting Syk-Card9-activating C-type lectin receptors by vaccine adjuvants: findings, implications and open questions. *Immunobiology* (2011) 216:1184–91. doi:10.1016/j.imbio.2011.06.005
- Zelensky AN, Gready JE. The C-type lectin-like domain superfamily. *FEBS J* (2005) 272:6179–217. doi:10.1111/j.1742-4658.2005.05031.x
- Blomhoff R, Tolleshaug H, Berg T. Binding of calcium ions to the isolated asialoglycoprotein receptor. Implications for receptor function in suspended hepatocytes. *J Biol Chem* (1982) 257:7456–9.
- Onizuka T, Shimizu H, Moriwaki Y, Nakano T, Kanai S, Shimada I, et al. NMR study of ligand release from asialoglycoprotein receptor under solution conditions in early endosomes. *FEBS J* (2012) 279:2645–56. doi:10.1111/j.1742-4658.2012.08643.x
- Tabarani G, Thepaut M, Stroebel D, Ebel C, Vives C, Vachette P, et al. DC-SIGN neck domain is a pH-sensor controlling oligomerization: SAXS and hydrodynamic studies of extracellular domain. *J Biol Chem* (2009) 284:21229–40. doi:10.1074/jbc.M109.021204
- Brown J, O'Callaghan CA, Marshall AS, Gilbert RJ, Siebold C, Gordon S, et al. Structure of the fungal beta-glucan-binding immune receptor dectin-1: implications for function. *Protein Sci* (2007) 16:1042–52. doi:10.1110/ps.072791207
- Weis WI, Drickamer K. Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem* (1996) 65:441–73. doi:10.1146/annurev.bi.65.070196.002301
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* (2001) 46:3–26. doi:10.1016/S0169-409X(00)00129-0
- Egner U, Hillig RC. A structural biology view of target drugability. *Expert Opin Drug Discov* (2008) 3:391–401. doi:10.1517/17460441.3.4.391
- Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov* (2002) 1:727–30. doi:10.1038/nrd892
- Cheng AC, Coleman RG, Smyth KT, Cao Q, Souillard P, Caffrey DR, et al. Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* (2007) 25:71–5. doi:10.1038/nbt1273
- Chen IJ, Hubbard RE. Lessons for fragment library design: analysis of output from multiple screening campaigns. *J Comput Aided Mol Des* (2009) 23:603–20. doi:10.1007/s10822-009-9280-5
- Sheridan RP, Maiorov VN, Holloway MK, Cornell WD, Gao YD. Drug-like density: a method of quantifying the “bindability” of a protein target based on a very large set of pockets and drug-like ligands from the Protein Data Bank. *J Chem Inf Model* (2010) 50:2029–40. doi:10.1021/ci100312t



33. Edfeldt FN, Folmer RH, Breeze AL. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov Today* (2011) **16**:284–7. doi:10.1016/j.drudis.2011.02.002
34. Henrich S, Salo-Ahen OM, Huang B, Rippmann FF, Cruciani G, Wade RC. Computational approaches to identifying and characterizing protein binding sites for ligand design. *J Mol Recognit* (2010) **23**:209–19. doi:10.1002/jmr.984
35. Perot S, Sperandio O, Miteva MA, Camproux AC, Villoutreix BO. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov Today* (2010) **15**:656–67. doi:10.1016/j.drudis.2010.05.015
36. Halgren TA. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* (2009) **49**:377–89. doi:10.1021/ci800324m
37. Hajduk PJ, Huth JR, Fesik SW. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* (2005) **48**:2518–25. doi:10.1021/jm049131r
38. Jordan JB, Poppe L, Xia X, Cheng AC, Sun Y, Michelsen K, et al. Fragment based drug discovery: practical implementation based on 19F NMR spectroscopy. *J Med Chem* (2012) **55**:678–87. doi:10.1021/jm201441k
39. Hajduk PJ, Huth JR, Tse C. Predicting protein druggability. *Drug Discov Today* (2005) **10**:1675–82. doi:10.1016/s1359-6446(05)03624-x
40. Schmidtke P, Barril X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J Med Chem* (2010) **53**:5858–67. doi:10.1021/jm100574m
41. Borrok MJ, Kiessling LL. Non-carbohydrate inhibitors of the lectin DC-SIGN. *J Am Chem Soc* (2007) **129**:12780–5. doi:10.1021/ja072944v
42. Mangold SL, Prost LR, Kiessling LL. Quinoxaline inhibitors of the lectin DC-SIGN. *Chem Sci* (2012) **3**:772–7. doi:10.1039/c2sc00767c
43. Hann MM, Leach AR, Harper G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci* (2001) **41**:856–64. doi:10.1021/ci000403i
44. Barelier S, Pons J, Gehring K, Lancelin JM, Krimm I. Ligand specificity in fragment-based drug design. *J Med Chem* (2010) **53**:5256–66. doi:10.1021/jm100496j
45. Davis BJ, Erlanson DA. Learning from our mistakes: the ‘unknown knowns’ in fragment screening. *Bioorg Med Chem Lett* (2013) **23**:2844–52. doi:10.1016/j.bmcl.2013.03.028
46. Dalvit C, Fagerness PE, Hadden DT, Sarver RW, Stockman BJ. Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. *J Am Chem Soc* (2003) **125**:7696–703. doi:10.1021/ja034646d
47. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic Acids Res* (2012) **40**:D290–301. doi:10.1093/nar/gkr1065
48. Feinberg H, Park-Snyder S, Kolatkar AR, Heise CT, Taylor ME, Weis WI. Structure of a C-type carbohydrate recognition domain from the macrophage mannose receptor. *J Biol Chem* (2000) **275**:21539–48. doi:10.1074/jbc.M002366200
49. Nagae M, Ikeda A, Kitago Y, Matsumoto N, Yamamoto K, Yamaguchi Y. Crystal structures of carbohydrate recognition domain of blood dendritic cell antigen-2 (BDCA2) reveal a common domain-swapped dimer. *Proteins* (2014) **82**:1512–8. doi:10.1002/prot.24504
50. Chemical Computing Group I. *Molecular Operating Environment (MOE)*. Montreal, QC: Chemical Computing Group Inc. (2013).
51. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* (2004) **20**:289–90. doi:10.1093/bioinformatics/btg412
52. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2014).
53. Meier M, Bider MD, Malashkevich VN, Spiess M, Burkhard P. Crystal structure of the carbohydrate recognition domain of the H1 subunit of the asialoglycoprotein receptor. *J Mol Biol* (2000) **300**:857–65. doi:10.1006/jmbi.2000.3853
54. Wurzburg BA, Tarchevskaya SS, Jardetzky TS. Structural changes in the lectin domain of CD23, the low-affinity IgE receptor, upon calcium binding. *Structure* (2006) **14**:1049–58. doi:10.1016/j.str.2006.03.017
55. Watson AA, Brown J, Harlos K, Eble JA, Walter TS, O’Zallaghan CA. The crystal structure and mutational binding analysis of the extracellular domain of the platelet-activating receptor CLEC-2. *J Biol Chem* (2007) **282**:3165–72. doi:10.1074/jbc.M610383200
56. Watson AA, Lebedev AA, Hall BA, Fenton-May AE, Vagin AA, Dejnirattisai W, et al. Structural flexibility of the macrophage dengue virus receptor CLEC5A: implications for ligand binding and signaling. *J Biol Chem* (2011) **286**:24208–18. doi:10.1074/jbc.M111.226142
57. Zhang JG, Czabotar PE, Policheni AN, Caminschi I, Wan SS, Kitsoulis S, et al. The dendritic cell receptor Clec9A binds damaged cells via exposed actin filaments. *Immunity* (2012) **36**:646–57. doi:10.1016/j.immuni.2012.03.009
58. Feinberg H, Mitchell DA, Drickamer K, Weis WI. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* (2001) **294**:2163–6. doi:10.1126/science.1066371
59. Graves BJ, Crowther RL, Chandran C, Rumberger JM, Li S, Huang KS, et al. Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature* (1994) **367**:532–8. doi:10.1038/367532a0
60. Swaminathan GJ, Weaver AJ, Loegering DA, Checkel JL, Leonidas DD, Gleich GJ, et al. Crystal structure of the eosinophil major basic protein at 1.8 Å. An atypical lectin with a paradigm shift in specificity. *J Biol Chem* (2001) **276**:26197–203. doi:10.1074/jbc.M100848200
61. Feinberg H, Taylor ME, Razi N, McBride R, Knirel YA, Graham SA, et al. Structural basis for langerin recognition of diverse pathogen and mammalian glycans through a single binding site. *J Mol Biol* (2011) **405**:1027–39. doi:10.1016/j.jmb.2010.11.039
62. Feinberg H, Powlesland AS, Taylor ME, Weis WI. Trimeric structure of langerin. *J Biol Chem* (2010) **285**:13285–93. doi:10.1074/jbc.M109.086058
63. Park H, Adsit FG, Boyington JC. The 1.4 angstrom crystal structure of the human oxidized low density lipoprotein receptor lox-1. *J Biol Chem* (2005) **280**:13593–9. doi:10.1074/jbc.M500768200
64. Sheriff S, Chang CY, Ezekowitz RA. Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil. *Nat Struct Biol* (1994) **1**:789–94. doi:10.1038/nsb1194-789
65. Furukawa A, Kamishikiryō J, Mori D, Toyonaga K, Okabe Y, Toji A, et al. Structural analysis for glycolipid recognition by the C-type lectins Mincle and MCL. *Proc Natl Acad Sci U S A* (2013) **110**:17438–43. doi:10.1073/pnas.1312649110
66. Somers WS, Tang J, Shaw GD, Camphausen RT. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell* (2000) **103**:467–79. doi:10.1016/S0092-8674(00)00138-0
67. Abergel C, Chenivessé S, Stinnakre MG, Guasco S, Brechot C, Claverie JM, et al. Crystallization and preliminary crystallographic study of HIP/PAP, a human C-lectin overexpressed in primary liver cancers. *Acta Crystallogr D Biol Crystallogr* (1999) **55**:1487–9. doi:10.1107/S0907444999007969
68. Gerbaud V, Pignol D, Loret E, Bertrand JA, Berland Y, Fontecilla-Camps JC, et al. Mechanism of calcite crystal growth inhibition by the N-terminal undecapeptide of lithostathine. *J Biol Chem* (2000) **275**:1057–64. doi:10.1074/jbc.275.2.1057
69. Shrive AK, Martin C, Burns I, Paterson JM, Martin JD, Townsend JP, et al. Structural characterisation of ligand-binding determinants in human lung surfactant protein D: influence of Asp325. *J Mol Biol* (2009) **394**:776–88. doi:10.1016/j.jmb.2009.09.057
70. Nielsen BB, Kastrup JS, Rasmussen H, Holtet TL, Graversen JH, Etzerodt M, et al. Crystal structure of tetranectin, a trimeric plasminogen-binding protein with an alpha-helical coiled coil. *FEBS Lett* (1997) **412**:388–96. doi:10.1016/S0014-5793(97)00664-9
71. Kastrup JS, Nielsen BB, Rasmussen H, Holtet TL, Graversen JH, Etzerodt M, et al. Structure of the C-type lectin carbohydrate recognition domain of human tetranectin. *Acta Crystallogr D Biol Crystallogr* (1998) **54**:757–66. doi:10.1107/S0907444997016806
72. Volkamer A, Griewel A, Grombacher T, Rarey M. Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J Chem Inf Model* (2010) **50**:2041–52. doi:10.1021/ci100241y
73. Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M. Combining global and local measures for structure-based druggability predictions. *J Chem Inf Model* (2012) **52**:360–72. doi:10.1021/ci200454v
74. Stambach NS, Taylor ME. Characterization of carbohydrate recognition by langerin, a C-type lectin of Langerhans cells. *Glycobiology* (2003) **13**:401–10. doi:10.1093/glycob/cwg045
75. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinel T, et al. *KNIME: The Konstanz Information Miner*. Berlin: Springer (2008).
76. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* (2010) **53**:2719–40. doi:10.1021/jm901137j

77. Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. *J Med Chem* (1996) **39**:2887–93. doi:10.1021/jm9602928
78. Sauer WH, Schwarz MK. Molecular shape diversity of combinatorial libraries: a prerequisite for broad bioactivity. *J Chem Inf Comput Sci* (2003) **43**:987–1003. doi:10.1021/ci025599w
79. Carr HY, Purcell EM. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys Rev* (1954) **94**:630–8. doi:10.1103/Physrev.94.630
80. Meiboom S, Gill D. Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum* (1958) **29**:688–91. doi:10.1063/1.1716296
81. Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* (2004) **9**:430–1. doi:10.1016/S1359-6446(04)03069-7
82. Lehotzky RE, Partch CL, Mukherjee S, Cash HL, Goldman WE, Gardner KH, et al. Molecular basis for peptidoglycan recognition by a bactericidal lectin. *Proc Natl Acad Sci U S A* (2010) **107**:7722–7. doi:10.1073/pnas.0909449107
83. Ng KK, Park-Snyder S, Weis WI. Ca<sup>2+</sup>-dependent structural changes in C-type mannose-binding proteins. *Biochemistry* (1998) **37**:17965–76. doi:10.1021/bi981972a
84. Nielbo S, Thomsen JK, Graversen JH, Jensen PH, Etzerodt M, Poulsen FM, et al. Structure of the plasminogen kringle 4 binding calcium-free form of the C-type lectin-like domain of tetranectin. *Biochemistry* (2004) **43**:8636–43. doi:10.1021/bi049570s
85. Wielens J, Headey SJ, Rhodes DI, Mulder RJ, Dolezal O, Deadman JJ, et al. Parallel screening of low molecular weight fragment libraries: do differences in methodology affect hit identification? *J Biomol Screen* (2013) **18**:147–59. doi:10.1177/1087057112465979
86. Mashalidis EH, Sledz P, Lang S, Abell C. A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat Protoc* (2013) **8**:2309–24. doi:10.1038/nprot.2013.130

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 May 2014; accepted: 26 June 2014; published online: 10 July 2014.

Citation: Aretz J, Wamhoff E-C, Hanske J, Heymann D and Rademacher C (2014) Computational and experimental prediction of human C-type lectin receptor druggability. *Front. Immunol.* **5**:323. doi: 10.3389/fimmu.2014.00323

This article was submitted to *Immunotherapies and Vaccines*, a section of the journal *Frontiers in Immunology*.

Copyright © 2014 Aretz, Wamhoff, Hanske, Heymann and Rademacher. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Carbohydrates in cyberspace

Elizabeth Yuriev<sup>1\*</sup> and Paul A. Ramsland<sup>2,3,4,5\*</sup>

<sup>1</sup> Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC, Australia, <sup>2</sup> Centre for Biomedical Research, Burnet Institute, Melbourne, VIC, Australia, <sup>3</sup> Department of Immunology, Alfred Medical Research and Education Precinct, Monash University, Melbourne, VIC, Australia, <sup>4</sup> Department of Surgery Austin Health, University of Melbourne, Heidelberg, VIC, Australia, <sup>5</sup> CHIRI Biosciences, School of Biomedical Sciences, Curtin University, Perth, WA, Australia

**Keywords:** Carbohydrate, curation, glycan, molecular modeling, online database, three-dimensional structure

Many research areas, particularly those focused on structural aspects of biomolecules, have “moved into cyberspace” in the last 20–30 years. This move has become even more prominent in the last 5–10 years, amplified by the increased space accessibility (from mobile devices to the Cloud) and by ever more sophisticated and powerful resources available for high-performance computing. The field of structural glycobiology has greatly benefitted from such progress and has taken advantage of the computational tools and resources developed specifically for the structural analysis of carbohydrates. In this short opinion piece, we do not intend to provide a comprehensive coverage of all such resources. We will briefly survey four websites [GLYCAM, UniCarb KnowledgeBase (UniCarbKB), Glycosciences.de, and Glyco3D], offering tools for structural glycobiology, and will provide some examples of studies where such tools have been successfully used for research relevant to understanding of carbohydrate recognition by proteins involved in immunity and infection. For a more comprehensive listing of computational resources for structural glycobiology, readers should consult Ref. (1–4).

GLYCAM-Web (5) is focused on the prediction of three-dimensional (3D) structures of carbohydrates and macromolecular structures that include carbohydrates. The server is created and operated by the research group of Prof. Robert J. Woods in the Complex Carbohydrate Research Center (CCRC) (6) at the University of Georgia in Athens. The server can perform conformational modeling of oligosaccharides as well as 3D modeling of glycoproteins. The main interfaces for modeling oligosaccharide conformations are the Carbohydrate Builder [including a special Builder for glycosaminoglycans (GAGs)], the Glycoprotein Builder, and the Oligosaccharide Libraries. The server is user-friendly and allows a range of upload and download options. Given a range of file formats – often, program-specific – used in molecular modeling, user-friendliness related to the portability of file formats is a significant “real-estate” feature highly valued by the cyberspace dwellers.

Most of the back-end software used by the server is open source and freely available to the public. This feature is also highly valued by viewers. The reason is not only reduced cost but also the openness of science that could be performed with open-source software. We consider such open access as critical in the field of molecular modeling: it allows researchers to use the program codes and develop them further, thus advancing the development of better modeling methods.

The main engine room of the server is the GLYCAM force field, the outcome of more than 20 years of development. Its latest incarnation, GLYCAM06 (7), has been significantly modified to satisfy the following requirements: transferability of the parameter set to all carbohydrate ring conformations and sizes, carbohydrate derivatives, and other biomolecules; self-containment and transferability to many quadratic force fields; ability to treat both  $\alpha$ - and  $\beta$ -anomers without specific atom types. The GLYCAM-Web server (5) is a very widely and extensively used resource. Its statistics records show that an average of 1700–1800 unique users access the site monthly (data examined: December 2014–March 2015).

The GLYCAM-Web server (5) and the GLYCAM force field (7) were used for the investigation of recognition specificity of ABO blood group antigens by antibodies (8). Such circulating antibodies

## OPEN ACCESS

### Edited by:

Lee Mark Wetzler,  
Boston University School of Medicine,  
USA

### Reviewed by:

Serge Perez,  
Centre national de la recherche  
scientifique, France

### \*Correspondence:

Elizabeth Yuriev  
elizabeth.yuriev@monash.edu;  
Paul A. Ramsland  
pramsland@burnet.edu.au

### Specialty section:

This article was submitted to  
Immunotherapies and Vaccines,  
a section of the journal *Frontiers in  
Immunology*

**Received:** 23 April 2015

**Accepted:** 26 May 2015

**Published:** 10 June 2015

### Citation:

Yuriev E and Ramsland PA (2015)  
Carbohydrates in cyberspace.  
*Front. Immunol.* 6:300.  
doi: 10.3389/fimmu.2015.00300

could cause a hyperacute immune response and sometimes death resulting from mismatched blood transfusion or organ transplantation (9). Another recent example of application of GLYCAM was in the study of structure and immune recognition of HIV-1 envelope (10), where it was used for modeling the *N*-linked glycan shield.

UniCarb KnowledgeBase (11, 12) was conceived to support online data storage and search capabilities for glycomics and glycobiology by integrating structural, experimental, and functional information. The aim of UniCarbKB is to advance the understanding of structures, pathways, and networks involved in glycosylation and processes mediated by carbohydrates. This information-rich resource is freely accessible, supports data annotation, and promotes adoption of common standards, necessary for seamless and meaningful integration of structural and functional data. UniCarbKB is the outcome of the collaborative effort of researchers from Australia (Macquarie University and University of NSW), Germany (Max Planck Institute of Colloids and Interfaces), Ireland (National Institute for Bioprocessing Research and Training), Japan (Soka University), Sweden (University of Gothenburg), Switzerland (Swiss Institute for Bioinformatics and University of Geneva), and USA (CCRC).

UniCarb KnowledgeBase provides querying interfaces for three-structural databases (GlycoSuiteDB, EUROCarbDB, and GlycoBase) and a link to the UniCarb-DB (13). UniCarbKB continues efforts started with GlycoSuiteDB (14) in curating structural glycobiology information from research literature. GlycoBase (15) and UniCarb-DB (16) are databases of experimental glycan structures. Support for EUROCarbDB (17) is planned for the future release.

At the time of writing, over 906 references, 3238 glycan structure entries, and 898 glycoproteins have been curated in UniCarbKB. A total of 598 protein glycosylation sites have been annotated using experimentally confirmed glycan structures. Beyond storing, curating, and providing searchable access to structural and functional data, UniCarbKB offers a range of tools for data pre-process and analysis: GlycanBuilder (18, 19) – for fast and intuitive drawing of glycan structures; GlycoMod tool (20, 21) – for predicting the possible glycan structures on proteins, from their experimentally determined masses. UniCarbKB is connected to the GLYCAM site (described above) and RINGS – a web resource offering algorithms and data mining tools for glycobiology research (22). Further, through its SugarBindDB module, UniCarbKB is linked to the Functional Glycomics Gateway of the Consortium for Functional Glycomics (CFG) (23) and the Glyco3D Portal for Structural Glycosciences (24, 25).

The GlycoMod tool of the UniCarbKB was used in a study of the formation of subcellular-specific *N*-glycosylation glycoprotein determinants (26). Lee et al. used liquid chromatography and mass spectrometry-based quantitative glycomics to investigate eight human breast epithelial cells with diverse genotypes and phenotypes – mostly human breast cancer cell lines – and have shown that the secreted glycoproteins consistently displayed more processed, primarily complex type, *N*-glycans than the high-mannose-rich microsomal glycoproteins. They

have also demonstrated that secreted glycoproteins displayed significantly more  $\alpha$ -sialylation and  $\alpha$ -1,6-fucosylation, but less  $\alpha$ -mannosylation, than both cell-surface and microsomal glycoproteomes.

Of particular interest to those working on the glycobiology of infectious disease is the SugarBindDB module of UniCarbKB – a database of pathogen lectins and corresponding glycan targets (27, 28). The SugarBindDB advances a pathogen-capture technology specifically for the binding of pathogen lectins to carbohydrate epitopes. The contents of the database come from publicly accessible information. After experts (glycobiologists, microbiologists, and medical histologists) locate candidate papers, the binding data is extracted for human pathogens and specific glycans.

The SugarBindDB database can be searched using names of different entities involved in the carbohydrate-protein interactions. For pathogenic agents (e.g., *Pseudomonas aeruginosa*), the outcomes include agent type and taxonomy ID. For lectins themselves (e.g., PA-IIL), the database stores genes, Protein Data Bank (PDB) codes, and associated Glyco3D and CFG references (see below for more information about the Glyco3D resources). A query could also be constructed for specific ligands (e.g., Lewis X). The entries in SugarBindDB are limited to sugar sequences that demonstrate consistent binding to pathogen lectins. Sugar sequences include “glycotope” within a larger carbohydrate structure (the precise lectin-binding site) or entire structures, depending on specific findings within the cited papers. Similar to the GLYCAM-Web server mentioned above, the SugarBindDB is user-friendly in that it allows displaying glycans in a variety of formats; specifically, it supports the modified 2D condensed IUPAC nomenclature, Oxford symbol notation, and the colored CFG cartoons. A sugar query can be input as a structure drawn with GlycanBuilder (18, 19). The SugarBindDB database can also be browsed by disease (e.g., Influenza) or by published references (by author or title descriptions) or searched by the affected area in the pathology (e.g., intestine). Search outputs are very informatively interconnected, i.e., the biological context of each ligand is shown in relation to a disease, to other known ligands, to similar lectins, etc. At the time of writing, over 178 references, 549 pathogenic agents, and 200 ligands have been curated in SugarBindDB.

Glyco3D (25) was developed at the Centre de Recherches sur les Macromolécules Végétales (CERMAV-CNRS), France. It comprises a group of databases containing the 3D data collected from an extensive screening of scientific literature: BiOligo (more than 250 bioactive oligosaccharides along with about 120 of their constituting disaccharide and about 80 monosaccharide segments); PolySac3DB (157 polysaccharides); Lectin3D (more than 1,000 structures); GAG3D (GAG-binding proteins co-crystallized with their ligands); MABs (a limited set of high resolution structures of carbohydrate-antibody complexes); GT3D (glycosyltransferases crystallized with or without their carbohydrate ligands). Lectins are oligomeric proteins, which can specifically recognize carbohydrates and are involved in many processes crucial to infection and immunity (29); ~70% of the Lectin3D structures are present in complex with a carbohydrate ligand, from monosaccharides

to oligosaccharides or glycoproteins, allowing great insight into the molecular basis of carbohydrate recognition by lectins. GAGs are complex anionic polysaccharides (e.g., heparin), specifically recognized by protein receptors, which participate in the regulation of many processes, including cell adhesion. Anti-carbohydrate antibodies, demonstrating specificity to various carbohydrate epitopes, are of a high importance in immunology and vaccine development (30).

The individual databases within Glyco3D are publically accessible via the portal (24), furnished with a user-friendly graphical user interface with several search options. All 3D structures, established using a variety of structure determination methods, can be visualized and some basic measurements are possible on the spot. For a more detailed analysis, the structures can be downloaded in a variety of commonly used molecular file formats.

Why would a user access 3D structural information relevant to glycobiology via databases such as Glyco3D while it can be obtained from the original sources, mainly PDB (31)? The answer is: in databases, such as Glyco3D, the data is manually curated and value-added. Such curation and value-adding is indispensable, given poor quality of structural information for carbohydrates in PDB. Agirre et al. have recently demonstrated that 64% of all *N*-glycan D-pyranosides in PDB reflect poor fit to the electron density (32). They have revealed significant structural errors, such as incorrect conformations, stereochemical configurations, and linkages, resulting from incorrect model building from the start, as well as more subtle errors, resulting from incorrect refinement. Even subtle errors could be detrimental if incorrect structures are interpreted in a biological context. The reasons for errors are various, but the main one seems to be poor chemical understanding on the part of crystallographers and the lack of appropriate torsional restraints, often in disagreement with the low-resolution data. The detriment of such errors goes beyond incorrect interpretations in specific cases. It also leads to skewed statistics based on all published/deposited structures, creating new abnormal “norms.” Thus, structural glycobiology is now ripe for remediation of the type that

the protein crystallography field underwent in not too distant past (33).

Tools for structural validation of carbohydrate 3D structural data are available, have been developed specifically for carbohydrates, and are publically accessible (34). For example, pdb-care (35) can check residue notation (which can be a source of confusion for carbohydrate residues) and atom connectivities and evaluate biological correctness of glycan structures. CARP (carbohydrate Ramachandran plot) (36) analyses carbohydrate data by generating a separate Phi/Psi plot for each linkage type, such as in the study of conformational preferences of *Shigella flexneri* O-antigens and their implications for vaccine design (37). Information necessary for such plots can be obtained from the GlyTorsion database (36), containing glycosidic torsions observed in PDB entries, or from conformational maps based on molecular dynamics simulations provided by GlycoMapsDB (38). Both pdb-care and CARP are available at the Glycosciences.de portal (39).

To summarize, the online resources available for structural glycobiology reviewed here demonstrate good scientific citizenship features: multinational collaborative approach to design and development, open access to data and code, and the drive for common and accepted standards. The collaborative approach is illustrated not only by initiatives such as UniCarbKB, but also by efforts to cross-link and integrate existing resources (e.g., SugarBindDB and Glyco3D) and map functional and structural information (e.g., mapping of SugarBindDB to GlycoSuiteDB to link saccharide sequences with their physical location in the body). Further, somewhat altruistic, aspect of these developments is the extensive effort that goes into manual curation of data. Efforts like these should be recognized, credited, and supported more strongly by relevant scientific societies and funding bodies.

## Acknowledgments

The authors gratefully acknowledge the contribution toward this study from the Victorian Operational Infrastructure Support Program received by the Burnet Institute.

## References

1. von der Lieth C-W, Lütke T, Frank M. *Bioinformatics for Glycobiology and Glycomics: An Introduction*. Chichester: John Wiley & Sons (2009). 494 p.
2. von der Lieth C-W, Lütke T, Frank M, editors. Appendix 1: list of available websites. *Bioinformatics for Glycobiology and Glycomics*. Chichester: John Wiley & Sons (2009). p. 447–52.
3. Aoki-Kinoshita KE. Using databases and web resources for glycomics research. *Mol Cell Proteomics* (2013) 12:1036–45. doi:10.1074/mcp.R112.026252
4. Lütke T, Frank M. *Glycoinformatics*. New York, NY: Springer (2015). 331 p.
5. GLYCAM Web-Server. (2015). Available from: <http://glycam.org/>
6. Complex Carbohydrate Research Center. (2015). Available from: <http://www.ccrcc.uga.edu/research/index.php>
7. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* (2008) 29:622–55. doi:10.1002/jcc.20820
8. Makeneni S, Ji Y, Watson DC, Young NM, Woods RJ. Predicting the origins of anti-blood group antibody specificity: a case study of the ABO A- and B-antigens. *Front Immunol* (2014) 5:397. doi:10.3389/fimmu.2014.00397
9. Ramsland P. Blood brothers: carbohydrates in xenotransplantation and cancer immunotherapy. *Immunol Cell Biol* (2005) 83:315–7. doi:10.1111/j.1440-1711.2005.01367.x
10. Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* (2014) 514:455–61. doi:10.1038/nature13808
11. Campbell MP, Peterson R, Mariethoz J, Gasteiger E, Akune Y, Aoki-Kinoshita KE, et al. UniCarbKB: building a knowledge platform for glycoproteomics. *Nucleic Acids Res* (2014) 42:D215–21. doi:10.1093/nar/gkt1128
12. UniCarbKB. (2015). Available from: <http://www.unicarbkb.org/>
13. UniCarb-DB. (2015). Available from: [http://unicarb-db.biomedicine.gu.se/unicarbdb/show\\_mucin.action](http://unicarb-db.biomedicine.gu.se/unicarbdb/show_mucin.action)
14. Cooper CA, Harrison MJ, Wilkins MR, Packer NH. GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res* (2001) 29:332–5. doi:10.1093/nar/29.1.332
15. GlycoBase. (2015). Available from: [https://glycobase.nibrt.ie/glycobase/show\\_nibrt.action](https://glycobase.nibrt.ie/glycobase/show_nibrt.action)
16. Campbell MP, Nguyen-Khuong T, Hayes CA, Flowers SA, Alagesan K, Kolarich D, et al. Validation of the curation pipeline of UniCarb-DB: building



- a global glycan reference MS/MS repository. *Biochim Biophys Acta* (2014) **1844**:108–16. doi:10.1016/j.bbapap.2013.04.018
17. *EUROCarbDB*. (2015). Available from: <http://www.ebi.ac.uk/pdbe-srv/eurocarbdb/>
  18. Damerell D, Ceroni A, Maass K, Ranzinger R, Dell A, Haslam SM. The GlycanBuilder and GlycoWorkbench glycoinformatics tools: updates and new developments. *Biol Chem* (2012) **393**:1357–62. doi:10.1515/hsz-2012-0135
  19. *GlycanBuilder*. (2015). Available from: <http://www.unicarbk.org/builder>
  20. Cooper CA, Gasteiger E, Packer NH. GlycoMod – a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* (2001) **1**:340–9. doi:10.1002/1615-9861(200102)1:2<340::AID-PROT340>3.0.CO;2-B
  21. *GlycoMod*. (2015). Available from: <http://web.expasy.org/glycomod/>
  22. *RINGS*. (2015). Available from: <http://rings.t.soka.ac.jp/>
  23. *Functional Glycomics Gateway (CFG)*. (2015). Available from: <http://www.functionalglycomics.org/static/index.shtml>
  24. Perez S, Sarkar A, Breton C, Drouillard S, Rivet A, Imberty A. *Glyco3D: A Portal for Structural Glycoscience*. (2013). Available from: <http://glyco3d.cermav.cnrs.fr>
  25. Pérez S, Sarkar A, Rivet A, Breton C, Imberty A. Glyco3D: a portal for structural glycosciences. In: Lütteke T, Frank M, editors. *Glycoinformatics*. New York, NY: Springer (2015). p. 241–58.
  26. Lee LY, Lin CH, Fanayan S, Packer NH, Thaysen-Andersen M. Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants. *Front Immunol* (2014) **5**:404. doi:10.3389/fimmu.2014.00404
  27. Shakhsher B, Anderson M, Khatib K, Tadoori L, Joshi L, Lisacek F, et al. SugarBind database (SugarBindDB): a resource of pathogen lectins and corresponding glycan targets. *J Mol Recognit* (2013) **26**:426–31. doi:10.1002/jmr.2285
  28. *SugarBindDB*. (2015). Available from: <http://sugarbind.expasy.org/>
  29. Sharon N, Lis H. *Lectins*. Netherlands: Springer (2007). 454 p.
  30. Dingjan T, Spendlove I, Durrant LG, Scott AM, Yuriev E, Ramsland PA. Structural biology of antibody recognition of carbohydrate epitopes and potential uses for targeted cancer immunotherapies. *Mol Immunol* (2015). doi:10.1016/j.molimm.2015.02.028
  31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res* (2000) **28**:235–42. doi:10.1093/nar/28.1.235
  32. Agirre J, Davies G, Wilson K, Cowtan K. Carbohydrate anomalies in the PDB. *Nat Chem Biol* (2015) **11**:303. doi:10.1038/nchembio.1798
  33. Read RJ, Adams PD, Arendall WB III, Brunger AT, Emsley P, Joosten RP, et al. A new generation of crystallographic validation tools for the protein data bank. *Structure* (2011) **19**:1395–412. doi:10.1016/j.str.2011.08.006
  34. Emsley P, Brunger AT, Lütteke T. Tools to assist determination and validation of carbohydrate 3D structure data. In: Lütteke T, Frank M, editors. *Glycoinformatics*. New York, NY: Springer (2015). p. 229–40.
  35. Lütteke T, von der Lieth CW. pdb-care (PDB carbohydrate residue check): a program to support annotation of complex carbohydrate structures in PDB files. *BMC Bioinformatics* (2004) **5**:69. doi:10.1186/1471-2105-5-69
  36. Lütteke T, Frank M, von der Lieth CW. Carbohydrate structure suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Res* (2005) **33**:D242–6. doi:10.1093/nar/gki013
  37. Theillet FX, Simenel C, Guerreiro C, Phalipon A, Mulard LA, Delepierre M. Effects of backbone substitutions on the conformational behavior of *Shigella flexneri* O-antigens: implications for vaccine strategy. *Glycobiology* (2011) **21**:109–21. doi:10.1093/glycob/cwq136
  38. Frank M, Lütteke T, von der Lieth CW. GlycoMapsDB: a database of the accessible conformational space of glycosidic linkages. *Nucleic Acids Res* (2007) **35**:287–90. doi:10.1093/nar/gkl907
  39. *Glycosciences.de*. (2015). Available from: <http://www.glycosciences.de/>

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Yuriev and Ramsland. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## ADVANTAGES OF PUBLISHING IN FRONTIERS



### FAST PUBLICATION

Average 90 days  
from submission  
to publication



### COLLABORATIVE PEER-REVIEW

Designed to be rigorous –  
yet also collaborative, fair and  
constructive



### RESEARCH NETWORK

Our network  
increases readership  
for your article



### OPEN ACCESS

Articles are free to read,  
for greatest visibility



### TRANSPARENT

Editors and reviewers  
acknowledged by name  
on published articles



### GLOBAL SPREAD

Six million monthly  
page views worldwide



### COPYRIGHT TO AUTHORS

No limit to  
article distribution  
and re-use



### IMPACT METRICS

Advanced metrics  
track your  
article's impact



### SUPPORT

By our Swiss-based  
editorial team