

Glycomics: an integrated systems approach to structure-function relationships of glycans

Rahul Raman¹, S Raguram¹, Ganesh Venkataraman^{1,3}, James C Paulson² & Ram Sasisekharan¹

In comparison with genomics and proteomics, the advancement of glycomics has faced unique challenges in the pursuit of developing analytical and biochemical tools and biological readouts to investigate glycan structure-function relationships. Glycans are more diverse in terms of chemical structure and information density than are DNA and proteins. This diversity arises from glycans' complex nontemplate-based biosynthesis, which involves several enzymes and isoforms of these enzymes. Consequently, glycans are expressed as an 'ensemble' of structures that mediate function. Moreover, unlike protein-protein interactions, which can be generally viewed as 'digital' in regulating function, glycan-protein interactions impinge on biological functions in a more 'analog' fashion that can in turn 'fine-tune' a biological response. This fine-tuning by glycans is achieved through the graded affinity, avidity and multivalency of their interactions. Given the importance of glycomics, this review focuses on areas of technologies and the importance of developing a bioinformatics platform to integrate the diverse datasets generated using the different technologies to allow a systems approach to glycan structure-function relationships.

Given the limited number of genes in the entire genome of mammals including humans, protein post-translational modifications (PTM) regulating protein function have a more important role in cell phenotype than was previously suspected. The central dogma of molecular biology has been revised to include protein PTM such as glycosylation. Glycosylation, or the attachment of glycans or carbohydrates to proteins, is perhaps the most extensive and complex form of protein PTM, and it provides for the needed functional diversity to generate extensive phenotypes from a limited genotype.

Based on their backbone chemical structure, glycans can be classified broadly as linear and branched sugars. Branched glycans are present as *N*-linked and *O*-linked glycosylation on glycoproteins or on glyco-

lipids¹⁻³. The majority of the linear sugars are glycosaminoglycans, which contain long polymers of sulfated disaccharide repeat units that are *O*-linked to a core protein, forming a proteoglycan aggregate⁴ (Fig. 1). There is accumulating evidence for the role of glycans in cell growth and development^{1,5-9}, tumor growth and metastasis¹⁰⁻¹⁵, anticoagulation¹⁶⁻¹⁸, immune recognition/response¹⁹⁻²³, cell-cell communication^{24,25} and microbial pathogenesis^{20,26-30}. Owing to their ubiquitous presence at the cell-extracellular interface, glycans are located in an environment of many proteins such as growth factors, cytokines, immune receptors, enzymes and others. The numerous biological roles of glycans are attributed to their interactions with these proteins, and thus, glycans modulate protein activity at the cell-extracellular interface.

¹Biological Engineering Division, Center for Biomedical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. ²Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ³Present address: Momenta Pharmaceuticals, 675 West Kendall Street, Cambridge, Massachusetts 02142, USA. Correspondence should be addressed to R.S. (rams@mit.edu).

The study of glycans presents unique challenges that necessitate a systems approach involving multiple components as well as integration of information at molecular, cellular, tissue and higher levels³¹. Three fundamental and interrelated aspects of glycans make this field both intriguing and challenging. First, the biosynthesis of glycans is a non-template driven process

involving coordinated expression of several glycosyltransferases, some of which have additional tissue-specific isoforms^{1–3,32,33}. The complex biosynthesis and lack of proofreading machinery leads to inherent heterogeneity and large diversity of glycan structures. Furthermore, this complicates investigation by a functional genetics approach to knock in or knock out specific structures and directly evaluate their effect on the whole-organism phenotype. Second, the chemical heterogeneity and diversity of glycans has challenged the development of analytical techniques to accurately define their chemical structures. Also, owing to their mode of biosynthesis, ubiquitous subcellular distribution and glycoprotein diversity arising from one or more glycosylation sites, glycans always need to be considered as a heterogeneous mixture of different chemical structures when isolated from cells and tissues. Finally, understanding the biochemical basis of glycan-protein interaction is complicated by the multivalency and graded affinity involving an ensemble of glycans making multiple contacts with multivalent protein binding sites^{4,24}. Thus, glycomics—defined as a systems or integrated approach to glycan investigation—is necessary to truly delineate glycan structure-function relationships.

Recognizing the need to take an integrated approach to advance glycan structure-function relationships, several international collaborative efforts, namely the Consortium for Functional Glycomics

Table 1 | Large-scale glycomics initiatives

Web resource (URL)	Glycomics initiatives
CFG (http://www.functionalglycomics.org)	Consortium for Functional Glycomics (CFG; USA)
EuroCarbDB (http://www.eurocarbdb.org)	Collaborative Glycomics Initiative (Europe)
HGPI (http://www.hgpi.jp)	Human Disease Glycomics/Proteomics Initiative (Japan)
CCRC (http://www.ccrcc.uga.edu)	Complex Carbohydrate Research (Georgia, USA)

(CFG; a multimillion dollar initiative funded by US National Institute of General Medical Sciences), EuroCarb, the Japanese Consortium for Glycomics and many other resources have been established (Table 1). Motivated by the need to address the challenges outlined above, these collaborative efforts are resulting in the development of novel resources and technologies for glycomics.

Using CFG as a model system, this review aims to provide a perspective on the different technologies ranging from a functional genetics approach to structural characterization of glycans and biochemical aspects of glycan-protein interactions. The focus of this review is on the datasets provided by these technologies, how they are interrelated and how they need to be integrated to enable glycomics. The review also discusses the development of a bioinformatics platform that bridges multiple datasets collected using the different technologies to provide a systems framework for glycomics. Although emphasis is given to branched sugars in this review, the overall concepts behind the integrated systems approach are applicable to both linear and branched sugars.

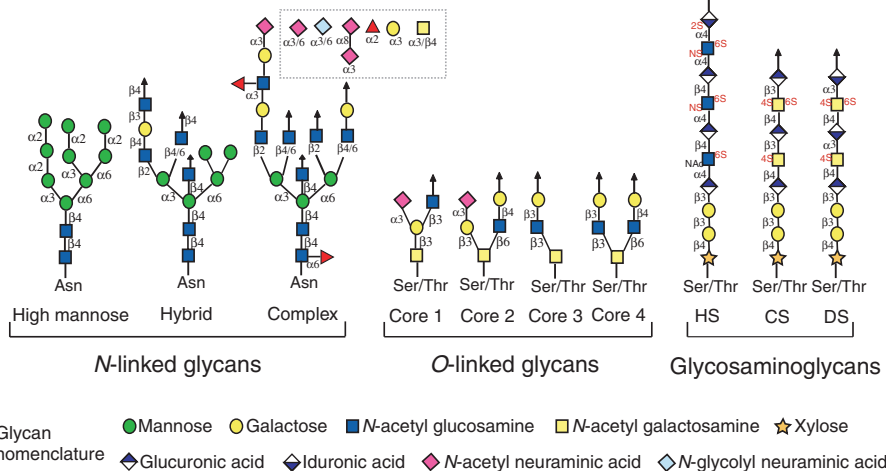
Functional genetics approach to glycomics

Toward obtaining functional readouts on the various biological roles of glycans, there have been advances in transgenic technologies to evaluate the effect of knockouts of glycan biosynthesis enzymes.

The number of known glycan biosynthetic enzymes has increased dramatically since the identification of the first set of glycosyltransferase genes in 1980s. Human and mouse glycosyltransferases have been primarily used in engineering glycosylation of proteins and antibodies that are used as therapeutics to improve their therapeutic parameters³¹. The glycosyltransferases have also been used in *de novo* synthesis of specific glycan structures^{34,35}.

There are 98 genes in humans corresponding to glycosyltransferases that have been annotated in the KEGG database (Table 2). As a part of the collaboration with the Japanese glycomics initiative, the CFG has expanded this list to around 200 genes and is now in the process of annotating them in terms of the reaction specificity of the enzymes. Characterization of the glycosyltransferases has permitted studies in which these genes are knocked out in somatic cell cultures to investigate effects of alteration or complete inhibition of glycosylation on cellular phenotype¹. Given that glycans are at the interface of the cell-extracellular region, it was necessary to develop whole-organism

© 2005 Nature Publishing Group <http://www.nature.com/naturemethods>



genetics to understand how genotype influences the phenotype of the entire organism. Advances in mammalian transgenic technologies have led to engineering knockouts of these genes *in vivo*.

Whole-organism functional genetic studies have provided valuable insights by directly linking the role of glycosylation of proteins and glycan diversification to the phenotype at the cellular and the whole-organism level. For example, knockouts of the GlcNAc and GalNAc transferases involved in the early stages of *N*-linked and *O*-linked glycan biosynthesis, and diversification often result in severe phenotypes—specifically, developmental defects, immune dysfunction, inflammation deficits and even embryonic lethality¹. This is perhaps not surprising as such mutations affect the majority of cells and their glycoproteins. But transgenic mice containing a knockout of later-stage enzymes such as fucosyl and sialyl transferases, which are involved in the cell type diversification and capping of *N*-linked and *O*-linked glycans, provide more discrete and consequently more difficult-to-detect phenotypes. Recent phenotype analysis of knockout strains of sialyl and fucosyl transferases has revealed interesting

phenotypes that provide evidence for specific glycan sequences in mediating aspects of cell-surface biology^{36–38}.

Large initiatives such as CFG now are generating transgenic mouse lines representing knockouts of later-stage fucosyl and sialyl transferases. These transgenic mice are subject to a battery of phenotype analysis studies, namely (i) hematology and coagulation chemistry, (ii) histological staining of tissues, (iii) immunology assays such as FACS, Ig level analysis, measurement of B- and T-cell proliferation upon induction with various agents and cytokines and (iv) various metabolism and behavioral tests. These studies have generated volumes of new data that provide many parameters to quantify distinct phenotypic abnormalities in these mice.

Owing to the chemical diversity of glycans, the above highlight the complexity and challenges in unraveling how glycans modulate whole-organism phenotype. Thus it is necessary to couple the functional genetics and whole-organism phenotyping studies with measuring gene expression of glycan biosynthesis enzymes, their binding proteins and to correlate these measurements with the repertoire of

Table 2 | Web-based resources for glycomics

Web resource (URL)	Key datasets or information
Consortium for Functional Glycomics (CFG; USA)	
GBP Molecule Page (http://www.functionalglycomics.org/glycomics/molecule/jsp/gbpMolecule-home.jsp)	Information portal with access to CFG and public databases
Glycan Database (http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp)	Database of glycan structures with search interfaces and links to CFG glycan array and MALDI-MS data
Glycan Profiling Data (http://www.functionalglycomics.org/glycomics/publicdata/glycoprofiling.jsp)	Raw and annotated MALDI-MS profiles of glycans from mouse and human cells and tissues
Glycan Array Screen Data (http://www.functionalglycomics.org/glycomics/publicdata/primaryscreen.jsp)	Raw data, bar graph of mean binding signal of GBP to each glycan in the array with links to their structures in glycan database
Gene Microarray Data (http://www.functionalglycomics.org/glycomics/publicdata/microarray.jsp)	Gene expression profiles of glycan biosynthesis enzymes and GBPs in various cells and tissues supplied by investigators
Transgenic Mice Phenotyping Data (http://www.functionalglycomics.org/lycomics/publicdata/phenotyping.jsp)	Experimental protocols, data files corresponding to various phenotyping analysis of transgenic mice
Kyoto Encyclopedia of Genes and Genomes (KEGG; Japan)	
KEGG Glycan Database (http://glycan.genome.jp)	Database of glycan structures obtained from CarbBank and updated with structures from other labs
KEGG Pathways dDatabase (http://www.genome.jp/kegg/pathway.html)	Collection of 15 glycan biosynthesis pathways with links to around 100 glycan biosynthesis enzymes
Glycomics Initiative of the German Cancer Research Institute (Glycosciences.de; Germany)	
Glycan Database (http://www.glycosciences.de/sweetdb/structure/)	Database of glycan structures
Glycans in Protein Data Bank (PDB) (http://www.glycosciences.de/sweetdb/start.php?action=form_pdb_data)	Glycan structures extracted from PDB entries using computational tools
Glycan NMR Profiles (http://www.glycosciences.de/sweetdb/nmr/)	Characteristic chemical shifts monosaccharides in different glycans
Computational Tools for Glycans (http://www.glycosciences.de/tools/index.php)	Collection of tools to analyze and query glycan structures and predict glycosylation sites on glycoproteins
Three-dimensional Modeling of Glycans (http://www.glycosciences.de/modeling/index.php)	Collection of tools to investigate conformational aspects of glycans and model their 3D structures
Other glycomics resources	
Glycosuite Database (http://www.glycosuitedb.org)	Commercial database and tools for glycans
Sugabase (http://www.boc.chem.uu.nl/sugabase/sugabase.html)	Glycan NMR database: chemical shifts of glycan structures
Lectin Database (http://www.imperial.ac.uk/research/animalllectins/)	Collection of information on animal lectins
Three-dimensional Lectin Database (http://www.cermav.cnrs.fr/lectines/)	Three-dimensional structures of lectins in the PDB
Bacterial Glycan Database (http://www.glyco.ac.ru/bcsdb/)	Database of bacterial glycan structures
CAZy (http://afmb.cnrs-mrs.fr/CAZY/)	Carbohydrate active enzymes database

glycan structures present on specific tissue or cell types. The development of technologies for taking such an approach is discussed in the following sections.

Development of glyco-gene microarray for glycomics

Measurement of simultaneous expression of several thousand genes in different cells to construct genetic networks and pathways has been an important component of a systems approach to molecular and cell biology. There have been considerable advances in the development of commercially available genome-wide microarrays (such as Affymetrix chips) to improve gene expression measurements, such as enhancing signal-to-noise ratio, among others. Investigating gene expression of enzymes involved in glycan biosynthesis and that of glycan binding proteins provides a new dimension to study processes at the cell-extracellular interface. There are challenges in using genome-wide arrays to investigate the dynamic nature of glycans-protein interactions. Some of these challenges arise from a limited representation of glycan biosynthesis enzymes on human and mouse genome microarrays and limited sensitivity in measuring expression of these genes relative to other downstream events³⁹.

Glyco-gene-based DNA microarrays, which focus on glycan biosynthesis and binding protein genes, were designed to overcome the above challenges. After careful consideration of choice of DNA printing technology, DNA probe format and appropriate sequence for the probes, an Affymetrix array-based glyco-gene microarrays were designed³⁹. These customized microarrays have been valuable resources for advancing glycomics. For example, over the past few years, around 400 samples representing various tissues and cell types have been analyzed on the CFG glyco-gene microarrays as a part of focused experiments to study differences in expression of glycan biosynthesis genes. The focus of the experiments include analysis of different tumor cell lines, cell types or tissues from glycosyltransferase or glycan binding protein (GBP) knockout mice strains, cells under mechanical stress such as chondrocytes and others. Thus these glyco-gene microarrays provide information on simultaneous expression of glycan biosynthetic enzymes that can be then correlated with the actual glycan structures, which had been characterized in a given sample.

Glycan analysis—from high-throughput to fine structure characterization

An important aspect of functional glycomics is the characterization of the primary chemical structure of glycans. Owing to heterogeneity of glycans arising from their complex nontemplate-driven biosynthesis, glycans isolated from cells and tissues comprise a heterogeneous repertoire of structures. Independent of the challenges in isolating glycans, there is a practical need to characterize the entire repertoire of glycan structures from the cell surface or on proteins as glycan-GBP interactions involve multivalent binding with several glycan structures

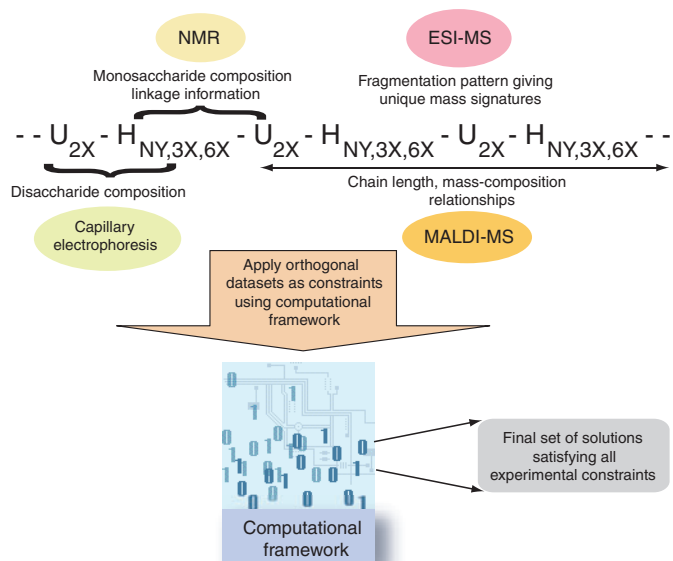
Figure 2 | Informatics approach to characterize glycans. Schematic of the technology developed to sequence HS-glycosaminoglycans in a rapid and unbiased fashion. The best set of attributes obtained using each methodology, namely disaccharide composition from capillary electrophoresis, monosaccharide composition and linkage information from NMR and chain length from MALDI-MS are incorporated as constraints into a computational framework that iteratively reduces the large solution space to the final set of solutions that satisfy all the experimental constraints. U, uronic acid (α -L-iduronic or β -D-glucuronic acid); H, α -D-glucosamine; linkage between U-H and H-U is 1–4. X represents sulfation sites (SO_3^-) and Y represents sites of acetylation (COCH_3).

on the cell surface. Several biochemical and analytical methodologies have been developed to address the above challenges.

There has always been trade-off between the sensitivity of fine structure characterization of glycan mixture (in terms of the exact sequence of each glycan) and the ability to perform a high-throughput analysis of a large number of glycans in the mixture. Mass spectrometric (MS) methods have been valuable in obtaining high-throughput mass profile of glycans from entire cells and tissues^{40–45}. Matrix-assisted laser desorption/ionization (MALDI)-MS-based analysis of human and mouse cells and tissues (as done by CFG) has provided a snapshot of the mass profiles along with the most likely set of glycan structures annotated by an expert. More recently, automated algorithms to annotate MALDI-MS spectra based on incorporating the domain knowledge have been developed⁴⁶. Whereas such a high-throughput analysis provides a good snapshot of the most likely structures to be present in a given tissue, the exact structures of the glycans in terms of the explicit monosaccharides and linkages are difficult to assign, particularly at higher molecular weights. A more-detailed linkage analysis of each glycan in the mixture is necessary for rigorous assignment of structure and also to quantify its relative abundance in the mixture.

High-performance liquid chromatography (HPLC) is a well-developed technique to obtain a profile of glycans in a mixture based on their elution profile^{47,48}. The glycans are labeled using a radioactive or fluorescent tag and are quantified based on the intensity values. HPLC-based analyses also provide quantitative information on relative abundance of different glycans. One such approach involves the use of a library of exo-glycosidases that specifically cleave individual monosaccharides from the nonreducing end of the glycan. The shifts in the chromatographic elution profile of glycans upon treatment with each glycosidase are used to assign the structure based on the specificity of cleavage⁴⁸.

Fine structure characterization of glycans involves selection of a mass or chromatographic peak followed by fragmentation or depolymerization and characterization of the fragments formed. Techniques involving different types of MALDI-MS and electrospray ionization (ESI)-MS analysis capture specific mass ions and fragment them to do a MS-MS or MS-MS-MS fragmentation pattern analysis. There have been considerable advances in the development of computational



glycan arrays. For example, the CFG has developed two types of glycan arrays—a well-based plate array on which a specific glycan ligand is prepared as a solution of fixed concentration and a solid-phase, printed array comprising NHS-activated glass slides on which the glycans are printed⁶⁸. Compared to the well-based array, the printed array better mimics the physiological distribution of glycans on a cell surface that will be presented to the multivalent GBPs. The GBP is introduced into the array upon treatment with a primary antibody, and the signal is obtained by using a secondary antibody attached to a fluorophore, similar to the enzyme-linked immunosorbent assay (ELISA). Typically, to promote multivalent high-affinity binding, conditions such as protein concentration are optimized to ensure that the GBP would be present in the dimeric or other multimeric forms. The glycan arrays have been extensively used to screen for novel ligand specificities for GBPs and for development of antibodies to target specific glycan motifs (Fig. 3). For example, using the CFG glycan arrays, detailed insights into the distinct ligand specificities of dendritic cell-specific ICAM grabbing nonintegrin (DC-SIGN) and DC-SIGN-related protein (DC-SIGNR) were identified. Both of these GBPs bound to high mannose-containing glycan structures, whereas only DC-SIGN bound to Lewis^x (motif in which Gal and Fuc are linked b4 and a3, respectively, to GlcNAc) and other fucosylated motifs²⁰. This analysis was important in understanding the structural basis for differences in host cell and pathogen recognition by these proteins. Thus the glycan array data are rapidly expanding the knowledge on the ligand specificity of different GBPs, thus providing a biochemical context to understanding how cellular phenotype is modulated by glyco-related gene expression.

Bioinformatics platform for glycomics

It is clear that there is a need to cut across multiple datasets to truly understand the structure-function relationships of glycans. A critical component that enables this process is a bioinformatics platform to store, integrate and process the information generated by the above methods and disseminate it in a meaningful fashion via the internet to the scientific community worldwide. The evolution of information generation for glycomics is different from that for genomics and proteomics. Representation of glycan chemical structure analogous to the primary sequence of proteins or DNA was challenging owing to the chemical complexity and branching patterns⁶⁹. Issues with representation of glycans were augmented by challenges in characterizing glycan structures using analytical techniques in the past. Owing to these challenges, earlier efforts to develop databases for storing glycan structures such as Complex Carbohydrate Structures Database (CCSD) had to be discontinued. In recent years, with the recognition of the importance of glycobiology and with advances in technology for characterizing glycan structures, academic and commercial organizations including the CFG are making considerable efforts to build databases such as Glycosuite database⁷⁰, KEGG Glycan database⁷¹ and tools^{46,72–75} for representation and analysis of glycan structures (Table 2).

An important aspect of the systems approach to glycan investigation is to define relationships between different entities that would facilitate the integration of information. As an analogy, gene ontologies developed by the Gene Ontology Consortium go beyond cataloging gene information. They capture relationships between molecular function of the gene product in context of a biological process. To capture complex relationships between diverse data, it is necessary to develop an object-based relational database. Datasets from the differ-

ent glycomics technologies include Microsoft Excel spread sheets to ASCII-delimited data of field and field-value pairs in which each individual parameter is important for data analysis and integration. For example, there are three primary objects in glycomics datasets—GBPs, glycan biosynthetic enzymes and the glycan structures. The different methodologies that provide datasets are further organized into secondary and other levels of objects with defined inter-relationships and relationships to the primary objects. The **Supplementary Note** online comprises an animated presentation showing the important components and features of large-scale glycomics databases using CFG as a model system.

The blueprint of the object-based relational database is the data model or ontology diagram that captures data definitions and inter-relationships, which is quite complex for glycomics databases (**Supplementary Note**). It is therefore important to develop a software architecture that keeps this complexity hidden from the user during data acquisition and dissemination. The three-tier software architecture comprising the back-end relational database to store the data and annotate their relationships, a middleware application layer that communicates between the database and the user interface, and the top layer comprising the user interfaces to the database is best-suited for this purpose. This software architecture facilitates the scientist to easily deposit the data into the database, which is automatically organized into the relational tables by the middleware application layer.

Central to the data integration is the ability to link orthogonal data sets derived from identical or similar samples. For example, the gene expression profile of a specific tissue or cell line isolated from a given strain of transgenic mice needs to be automatically associated in the database with orthogonal information such as glycan profile, histological staining and immunological profile from a similar or identical sample. Such an integration would allow researchers to cut across multiple datasets and start asking the questions such as “does the expression of glycosyltransferase correlate with glycan profile of that tissue?” or “can the pathological analysis of the tissue be explained on the basis of gene expression profile?”

Another emerging concept in data integration is the molecule page interface, which provides a portal to information and data ranging from molecule to mouse. The molecule pages are evolving into standardized interfaces not only for glycomics but also for genomics and proteomics initiatives⁷⁶. In the case of glycomics, the molecule page interface was introduced by CFG to capture information pertaining to different families of glycan binding proteins (**Supplementary Note**). The CFG molecule pages contain three main components: (i) automatic acquisition of information from other public databases on that molecule, (ii) automatic interface with CFG data pertaining to that molecule, (iii) contribution from a group of experts on that particular molecule.

Finally, the bioinformatics platform needs to support computational tools to perform data mining analysis on the large scale glycomics data sets. For example the prediction of glycan structures based on gene expression profiles of glycan biosynthetic enzymes and the identification of patterns in an ensemble of glycans that govern the multivalent high-affinity interactions with specific GBPs are now enabled because of the user-friendly access to diverse datasets via relational databases.

Conclusions and future directions

Technological advances in DNA microarrays and mass spectrometric methods coupled with availability of genome-wide sequence informa-

tion has steered postgenomics research towards a 'systems' approach to understanding cellular phenotype as a function of its gene and protein components. Glycomics has emerged as a fundamental field in providing an important dimension to this approach.

The challenges that lie ahead for advancing glycomics include: (i) explaining how glycan diversity is regulated as a function of its biosynthesis, (ii) understanding the basis for specificity of glycan-protein interactions, and (iii) elucidating how an ensemble of glycans displayed on the cell surface govern extracellular signal transduction and cell-cell communication via multivalent interactions with proteins. The availability of diverse data sets and their integration via object-oriented relational databases has motivated the development of computational tools to perform data mining and pattern analysis to begin addressing these questions. It is important to address these kinds of questions before modeling biochemical pathways and network interactions.

There has also been an increasing awareness for the need to develop data exchange formats such as XML for consistent description of glycan structures⁷⁴ and glycomics data sets across different large scale glycomics initiatives. There is a practical need to set standards for incorporating glycan structures into a database to develop the glycan database into an international resource similar to GenBank and SwissProt. In summary, it is envisioned that large scale glycomics initiatives would continue their focus on developing and applying technologies to advance this important field.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported by the US National Institute of General Medical Sciences Glue Grant U54 GM62116. We thank the CFG members, the members of Bioinformatics Core (B) of CFG for the work described in this review and V. Sasisekharan for critical reading of the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

Published online at <http://www.nature.com/naturemethods/>
Reprints and permissions information is available online at <http://ngp.nature.com/reprintsandpermissions/>

- Lowe, J.B. & Marth, J.D. A genetic approach to mammalian glycan function. *Annu. Rev. Biochem.* **72**, 643–691 (2003).
- Varki, A. *et al.* *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, New York, 1999).
- Taylor, M.E. & Drickamer, K. *Introduction to glycobiology* (Oxford University Press, Oxford and New York, 2003).
- Raman, R., Sasisekharan, V. & Sasisekharan, R. Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chem. Biol.* **12**, 267–277 (2005).
- Hwang, H.Y., Olson, S.K., Esko, J.D. & Horvitz, H.R. *Caenorhabditis elegans* early embryogenesis and vulval morphogenesis require chondroitin biosynthesis. *Nature* **423**, 439–443 (2003).
- Inatani, M., Irie, F., Plump, A.S., Tessier-Lavigne, M. & Yamaguchi, Y. Mammalian brain morphogenesis and midline axon guidance require heparan sulfate. *Science* **302**, 1044–1046 (2003).
- Lin, X. Functions of heparan sulfate proteoglycans in cell signaling during development. *Development* **131**, 6009–6021 (2004).
- Haltiwanger, R.S. & Lowe, J.B. Role of glycosylation in development. *Annu. Rev. Biochem.* **73**, 491–537 (2004).
- Cipollo, J.F., Awad, A.M., Costello, C.E. & Hirschberg, C.B. *N-glycans of Caenorhabditis elegans* are specific to developmental stages. *J. Biol. Chem.* **280**, 26063–26072 (2005).
- Sasisekharan, R., Shriver, Z., Venkataraman, G. & Narayanasami, U. Roles of heparan-sulphate glycosaminoglycans in cancer. *Nat. Rev. Cancer* **2**, 521–528 (2002).
- Liu, D., Shriver, Z., Venkataraman, G., El Shabrawi, Y. & Sasisekharan, R. Tumor cell surface heparan sulfate as cryptic promoters or inhibitors of tumor growth and metastasis. *Proc. Natl. Acad. Sci. USA* **99**, 568–573 (2002).
- Fuster, M.M., Brown, J.R., Wang, L. & Esko, J.D. A disaccharide precursor of sialyl Lewis X inhibits metastatic potential of tumor cells. *Cancer Res.* **63**, 2775–2781 (2003).
- Ishida, H. *et al.* A novel β 1,3-*N*-acetylglucosaminyltransferase (β 3Gn-T8), which synthesizes poly(*N*-acetylglucosamine), is dramatically upregulated in colon cancer. *FEBS Lett.* **579**, 71–78 (2005).
- Iwai, T. *et al.* Core 3 synthase is downregulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc. Natl. Acad. Sci. USA* **102**, 4572–4577 (2005).
- Dube, D.H. & Bertozzi, C.R. Glycans in cancer and inflammation—potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **4**, 477–488 (2005).
- Casu, B., Guerrini, M. & Torri, G. Structural and conformational aspects of the anticoagulant and anti-thrombotic activity of heparin and dermatan sulfate. *Curr. Pharm. Des.* **10**, 939–949 (2004).
- Petitou, M. & van Boeckel, C.A. A synthetic antithrombin III binding pentasaccharide is now a drug! What comes next? *Angew. Chem. Int. Edn Engl.* **43**, 3118–3133 (2004).
- Shriver, Z., Liu, D. & Sasisekharan, R. Emerging views of heparan sulfate glycosaminoglycan structure/activity relationships modulating dynamic biological functions. *Trends Cardiovasc. Med.* **12**, 71–77 (2002).
- Kinjo, Y. *et al.* Recognition of bacterial glycosphingolipids by natural killer T cells. *Nature* **434**, 520–525 (2005).
- Guo, Y. *et al.* Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.* **11**, 591–598 (2004).
- Crocker, P.R. Siglecs in innate immunity. *Curr. Opin. Pharmacol.* **5**, 431–437 (2005).
- Rudd, P.M., Wormald, M.R. & Dwek, R.A. Sugar-mediated ligand-receptor interactions in the immune system. *Trends Biotechnol.* **22**, 524–530 (2004).
- Rudd, P.M., Elliott, T., Cresswell, P., Wilson, I.A. & Dwek, R.A. Glycosylation and the immune system. *Science* **291**, 2370–2376 (2001).
- Collins, B.E. & Paulson, J.C. Cell surface biology mediated by low affinity multivalent protein-glycan interactions. *Curr. Opin. Chem. Biol.* **8**, 617–625 (2004).
- Crocker, P.R. Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling. *Curr. Opin. Struct. Biol.* **12**, 609–615 (2002).
- Fry, E.E. *et al.* The structure and function of a foot-and-mouth disease virus-oligosaccharide receptor complex. *EMBO J.* **18**, 543–554 (1999).
- Ganesh, V.K., Smith, S.A., Kotwal, G.J. & Murthy, K.H. Structure of vaccinia complement protein in complex with heparin and potential implications for complement regulation. *Proc. Natl. Acad. Sci. USA* **101**, 8924–8929 (2004).
- Liu, J. *et al.* Characterization of a heparan sulfate octasaccharide that binds to herpes simplex virus type 1 glycoprotein d. *J. Biol. Chem.* **277**, 33456–33467 (2002).
- Mahdavi, J. *et al.* *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* **297**, 573–578 (2002).
- Miller, S.I., Ernst, R.K. & Bader, M.W. LPS, TLR4 and infectious disease diversity. *Nat. Rev. Microbiol.* **3**, 36–46 (2005).
- Shriver, Z., Raguram, S. & Sasisekharan, R. Glycomics: a pathway to a class of new and improved therapeutics. *Nat. Rev. Drug Discov.* **3**, 863–873 (2004).
- Sasisekharan, R. & Venkataraman, G. Heparin and heparan sulfate: biosynthesis, structure and function. *Curr. Opin. Chem. Biol.* **4**, 626–631 (2000).
- Sugahara, K. *et al.* Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. *Curr. Opin. Struct. Biol.* **13**, 612–620 (2003).
- Blixt, O. *et al.* Chemoenzymatic synthesis of 2-azidoethyl-ganglio-oligosaccharides GD3, GT3, GM2, GD2, GT2, GM1, and GD1a. *Carbohydr. Res.* **340**, 1963–1972 (2005).
- Hanson, S., Best, M., Bryan, M.C. & Wong, C.H. Chemoenzymatic synthesis of oligosaccharides and glycoproteins. *Trends Biochem. Sci.* **29**, 656–663 (2004).
- Homeister, J.W., Daugherty, A. & Lowe, J.B. α (1,3) fucosyltransferases FucT-IV and FucT-VII control susceptibility to atherosclerosis in apolipoprotein E^{-/-} mice. *Arterioscler. Thromb. Vasc. Biol.* **24**, 1897–1903 (2004).
- Smithson, G. *et al.* Fuc-TVII is required for T helper 1 and T cytotoxic 1 lymphocyte selectin ligand expression and recruitment in inflammation, and together with Fuc-TIV regulates naive T cell trafficking to lymph nodes. *J. Exp. Med.* **194**, 601–614 (2001).
- Martin, L.T., Marth, J.D., Varki, A. & Varki, N.M. Genetically altered mice with different sialyltransferase deficiencies show tissue-specific alterations in sialylation and sialic acid 9-*O*-acetylation. *J. Biol. Chem.* **277**, 32930–32938 (2002).
- Comelli, E.M., Amado, M., Head, S.R. & Paulson, J.C. Custom microarray for

- glycobiologists: considerations for glycosyltransferase gene expression profiling. *Biochem. Soc. Symp.* **69**, 135–142 (2002).
40. An, H.J., Peavy, T.R., Hedrick, J.L. & Lebrilla, C.B. Determination of *N*-glycosylation sites and site heterogeneity in glycoproteins. *Anal. Chem.* **75**, 5628–5637 (2003).
 41. Cipollo, J.F., Costello, C.E. & Hirschberg, C.B. The fine structure of *Caenorhabditis elegans* *N*-glycans. *J. Biol. Chem.* **277**, 49143–49157 (2002).
 42. Dell, A. & Morris, H.R. Glycoprotein structure determination by mass spectrometry. *Science* **291**, 2351–2356 (2001).
 43. Morelle, W., Page, A. & Michalski, J.C. Electrospray ionization ion trap mass spectrometry for structural characterization of oligosaccharides derivatized with 2-aminobenzamide. *Rapid Commun. Mass Spectrom.* **19**, 1145–1158 (2005).
 44. Morelle, W. *et al.* Fragmentation characteristics of permethylated oligosaccharides using a matrix-assisted laser desorption/ionization two-stage time-of-flight (TOF/TOF) tandem mass spectrometer. *Rapid Commun. Mass Spectrom.* **18**, 2637–2649 (2004).
 45. Kameyama, A. *et al.* Detection of oligosaccharides labeled with cyanine dyes using matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **76**, 4537–4542 (2004).
 46. Goldberg, D., Sutton-Smith, M., Paulson, J. & Dell, A. Automatic annotation of matrix-assisted laser desorption/ionization *N*-glycan spectra. *Proteomics* **5**, 865–875 (2005).
 47. Rudd, P.M. *et al.* A high-performance liquid chromatography based strategy for rapid, sensitive sequencing of *N*-linked oligosaccharide modifications to proteins in sodium dodecyl sulphate polyacrylamide electrophoresis gel bands. *Proteomics* **1**, 285–294 (2001).
 48. Royle, L. *et al.* An analytical and structural database provides a strategy for sequencing *O*-glycans from microgram quantities of glycoproteins. *Anal. Biochem.* **304**, 70–90 (2002).
 49. Ethier, M., Saba, J.A., Ens, W., Standing, K.G. & Perreault, H. Automated structural assignment of derivatized complex *N*-linked oligosaccharides from tandem mass spectra. *Rapid Commun. Mass Spectrom.* **16**, 1743–1754 (2002).
 50. Joshi, H.J. *et al.* Development of a mass fingerprinting tool for automated interpretation of oligosaccharide fragmentation data. *Proteomics* **4**, 1650–1664 (2004).
 51. Lohmann, K.K. & von der Lieth, C.W. GlycoFragment and GlycoSearchMS: web tools to support the interpretation of mass spectra of complex carbohydrates. *Nucleic Acids Res.* **32**, W261–W266 (2004).
 52. Tang, H., Mechref, Y. & Novotny, M.V. Automated interpretation of MS/MS spectra of oligosaccharides. *Bioinformatics* **21** (Suppl. 1), i431–i439 (2005).
 53. Park, Y. & Lebrilla, C.B. Application of Fourier transform ion cyclotron resonance mass spectrometry to oligosaccharides. *Mass Spectrom. Rev.* **24**, 232–264 (2005).
 54. Hakansson, K. *et al.* Electron capture dissociation and infrared multiphoton dissociation MS/MS of an *N*-glycosylated tryptic peptic to yield complementary sequence information. *Anal. Chem.* **73**, 4530–4536 (2001).
 55. McFarland, M.A. *et al.* Structural characterization of the GM1 ganglioside by infrared multiphoton dissociation, electron capture dissociation, and electron detachment dissociation electrospray ionization FT-ICR MS/MS. *J. Am. Soc. Mass Spectrom.* **16**, 752–762 (2005).
 56. Froesch, M. *et al.* Coupling of fully automated chip electrospray to Fourier transform ion cyclotron resonance mass spectrometry for high-performance glycoscreening and sequencing. *Rapid Commun. Mass Spectrom.* **18**, 3084–3092 (2004).
 57. Zamfir, A. *et al.* Fully automated chip-based mass spectrometry for complex carbohydrate system analysis. *Anal. Chem.* **76**, 2046–2054 (2004).
 58. Guerrini, M., Bisio, A. & Torri, G. Combined quantitative ¹H and ¹³C nuclear magnetic resonance spectroscopy for characterization of heparin preparations. *Semin. Thromb. Hemost.* **27**, 473–482 (2001).
 59. Manzi, A.E. *et al.* Exploring the glycan repertoire of genetically modified mice by isolation and profiling of the major glycan classes and nano-NMR analysis of glycan mixtures. *Glycobiology* **10**, 669–689 (2000).
 60. Lopez, M. *et al.* Microheterogeneity of the oligosaccharides carried by the recombinant bovine lactoferrin expressed in *Mamestra brassicae* cells. *Glycobiology* **7**, 635–651 (1997).
 61. Venkataraman, G., Shriver, Z., Raman, R. & Sasisekharan, R. Sequencing complex polysaccharides. *Science* **286**, 537–542 (1999).
 62. Guerrini, M. *et al.* A novel computational approach to integrate NMR spectroscopy and capillary electrophoresis for structure assignment of heparin and heparan sulfate oligosaccharides. *Glycobiology* **12**, 713–719 (2002).
 63. Sharon, N. & Lis, H. *Lectins* (Kluwer Academic Publishers, Boston, 2004).
 64. Plante, O.J., Palmacci, E.R. & Seeberger, P.H. Automated solid-phase synthesis of oligosaccharides. *Science* **291**, 1523–1527 (2001).
 65. Hirabayashi, J. Lectin-based structural glycomics: glycoproteomics and glycan profiling. *Glycoconj. J.* **21**, 35–40 (2004).
 66. Feizi, T., Fazio, F., Chai, W. & Wong, C.H. Carbohydrate microarrays—a new set of technologies at the frontiers of glycomics. *Curr. Opin. Struct. Biol.* **13**, 637–645 (2003).
 67. Galanina, O.E., Mecklenburg, M., Nifantiev, N.E., Pazynina, G.V. & Bovin, N.V. GlycoChip: multiarray for the study of carbohydrate-binding proteins. *Lab Chip* **3**, 260–265 (2003).
 68. Blixt, O. *et al.* Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. USA* **101**, 17033–17038 (2004).
 69. von der Lieth, C.W., Bohne-Lang, A., Lohmann, K.K. & Frank, M. Bioinformatics for glycomics: status, methods, requirements and perspectives. *Brief. Bioinform.* **5**, 164–178 (2004).
 70. Cooper, C.A., Harrison, M.J., Wilkins, M.R. & Packer, N.H. GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res.* **29**, 332–335 (2001).
 71. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280 (2004).
 72. Aoki, K.F. *et al.* Application of a new probabilistic model for recognizing complex patterns in glycans. *Bioinformatics* **20** (Suppl. 1), 16–114 (2004).
 73. Bohne-Lang, A., Lang, E., Forster, T. & von der Lieth, C.W. LINUCS: linear notation for unique description of carbohydrate sequences. *Carbohydr. Res.* **336**, 1–11 (2001).
 74. Kikuchi, N. *et al.* The carbohydrate sequence markup language (CabosML): an XML description of carbohydrate structures. *Bioinformatics* **21**, 1717–1718 (2005).
 75. Lutteke, T., Frank, M. & von der Lieth, C.W. Data mining the protein data bank: automatic detection and assignment of carbohydrate structures. *Carbohydr. Res.* **339**, 1015–1020 (2004).
 76. Li, J. *et al.* The Molecule Pages database. *Nature* **420**, 716–717 (2002).