DOI 10.1002/pmic.200700917 *Proteomics* 2008, *8*, 8–20

Frontiers in glycomics: Bioinformatics and biomarkers in disease

An NIH White Paper prepared from discussions by the focus groups at a workshop on the NIH campus, Bethesda MD (September 11–13, 2006)

REPORT

Nicolle H. Packer[1], Claus-Wilhelm von der Lieth[2]*†, Kiyoko F. Aoki-Kinoshita[3], Carlito B. Lebrilla[4], James C. Paulson[5], Rahul Raman[6], Pauline Rudd[7], Ram Sasisekharan[6], Naoyuki Taniguchi[8] and William S. York[9]*

[1] Macquarie University, Sydney, Australia
[2] German Cancer Research Center, Heidelberg, Germany
[3] Soka University, Tokyo, Japan
[4] University of California Davis, CA, USA
[5] Scripps Research Institute, La Jolla, CA, USA
[6] MIT, Cambridge, MA, USA
[7] University College Dublin, Dublin, Ireland
[8] Osaka University, Osaka, Japan
[9] Complex Carbohydrate Research Center, Athens, GA, USA

Key issues relating to glycomics research were discussed after the workshop entitled "Frontiers in Glycomics: Bioinformatics and Biomarkers in Disease" by two focus groups nominated by the organizers. The groups focused on two themes: (i) glycomics as the new frontier for the discovery of biomarkers of disease and (ii) requirements for the development of informatics for glycomics and glycobiology. The mandate of the focus groups was to build consensus on these issues and develop a summary of findings and recommendations for presentation to the NIH and the greater scientific community. A list of scientific priorities was developed, presented, and discussed at the workshops. Additional suggestions were solicited from workshop participants and collected using the workshop mailing list. The results are summarized in this White Paper, authored by the co-chairs of the focus groups.

Received: September 26, 2007
Accepted: November 1, 2007

Keywords:
Bioinformatics / Biomarkers / Glycobiology / Glycomics

Correspondence: Professor Nicolle H. Packer, Department of Chemistry and Biomolecular Sciences, CORE of Biomolecular Frontiers, Rm 307, Building E8C, Macquarie University, Sydney, New South Wales, NSW 2109, Australia
E-mail: nicki.packer@mq.edu.au
Fax: +61-2-9850-8313

Abbreviations: ADCC, antibody-dependent cell-mediated cytotoxicity; **AFP,** alpha-fetoprotein; **FSH,** follicle-stimulating hormone; **GAGs,** glycosaminoglycans; **N-glycans,** N-linked oligosaccharides; **O-glycans,** O-linked oligosaccharides

1 Mandate

To describe and identify the state and the potential of glycans as biomarkers for diseases and to recommend the tools that need to be supported so as to develop and enhance biomarker discoveries that employ the glycome.

* Both authors contributed equally to this work.
† Deceased; see note added in proof.

2 The need for glycomics in biomarker discovery

Glycans include short carbohydrate chains (*i.e.*, oligosaccharides) and larger molecules containing many carbohydrate residues (*e.g.*, glycosaminoglycans) that are bound to proteins or lipids but may also be free. Glycomics—the study of the biological role of carbohydrates—is opening up new research fronts, and pharmaceutical and biotechnology companies are probing the glycome for targets for novel drugs or new therapies for infectious disease, cancer, and metabolic disorders (Fig. 1).

Two years ago, in the Massachusetts Institute of Technology's *Technology Review*, Dr. Terry van der Werff nominated glycomics as one of ten emerging technologies that will change the world.

"The glycome was regarded as a much bigger challenge than the genome or proteome—the language of sugars was just too complex. But all that has changed dramatically in the past decade. New technologies are facilitating exploration and a new understanding of the glycome, and chemists now have the tools to assemble large, complex carbohydrate molecules from simple monosaccharide components. It's now obvious that carbohydrates play significant roles in healthy biology as well as disease. Glycomics should be ranked equally with the genome and proteome, and developed as rapidly as possible" (Prof. Mark Von Itzstein 2006, Institute of Glycomics, Australia, and one of the discoverers of the sugar enzyme inhibitor of influenza infection (Relenza™)).

The surface of all cells are elaborated by the addition of sugars to the membrane macromolecules—the ability of the structure of these sugars to be fine-tuned affects the communication between cells, serves as docking pads for bacteria or viruses, and provides clinically useful markers for diseases.

2.1 Glycans are involved in a host of disease-related functions

The importance of carbohydrates in general metabolic processes and the malfunctions resulting in disease are evidenced by the extremely diverse physical and mental deficiencies evident in the 16 or so congenital disorders of glycosylation that have been classified (CDG Syndrome). There are also several congenital muscle dystrophies designated as alpha-dysglycanopathy due to the mutation of glycosyltransferases involved in O-mannosylation of alpha-dystroglycan. These genetic defects are often a mutation in a single glycosyltransferase gene causing widespread phenotypic effects usually devastating to the patient.

The primarily extracellular location of the glycans means that they are intrinsic to cell–cell interactions such as pathogenic infection, fertility, immunity, and cancer. Glycans thus provide an alternative, and to date an under-exploited, molecular class with which to look for disease biomarkers, drug targets, and therapeutics.

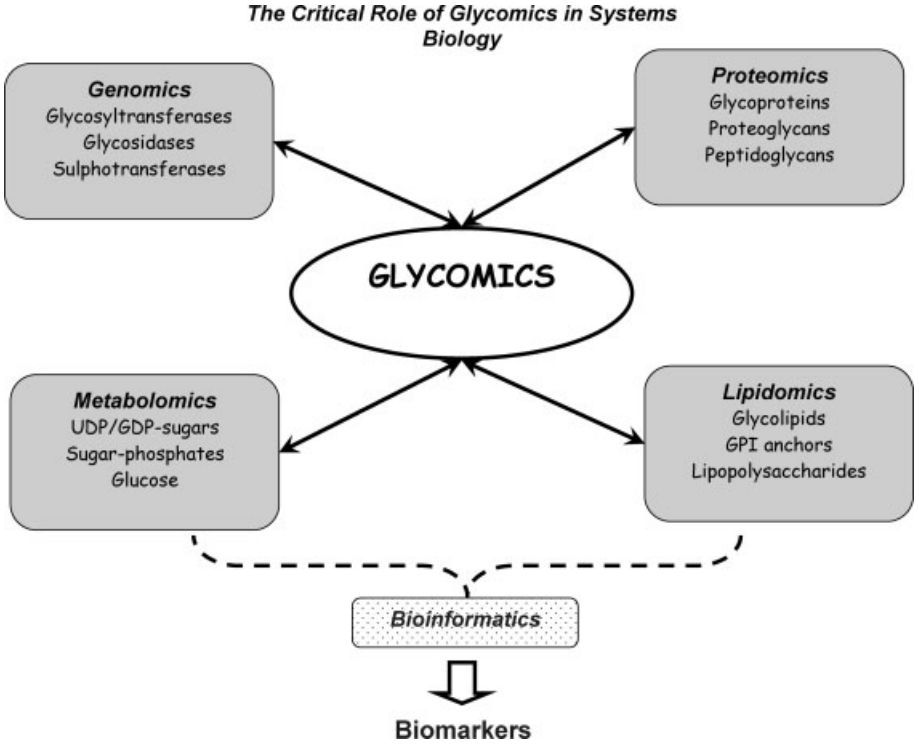

Figure 1. The critical role of glycomics in systems biology.

2.1.1 Glycans are potential biomarkers for cancer

Most tumor antigens are glycoproteins or glycolipids and their monoclonal antibodies were generated against peptide portions of the glycoprotein, or sugar portions of glycolipid, one or two decades ago. Although N-glycans are not known to be immunogenic, a combined protein–sugar epitope on a glycoprotein can form a very specific immunogen. Well-known clinical cancer diagnostic tests use such existing glycoprotein cancer markers (CA125, CA19-9, CEA, CA15-3, MUC1), which are on the whole not specific to a particular cancer and for which the exact epitope has largely not been defined. Alpha-fetoprotein (AFP) and prostate specific antigen (PSA) are glycoproteins which are tissue specific and thus are being used for monitoring primary hepatoma and prostate cancer, respectively. Unfortunately only the polypeptide component of most of these cancer biomarkers is being exploited in clinical tests and needs to be re-evaluated, as these proteins may be elevated in patients with benign or inflammatory diseases. It is highly likely that glycoform variants of these cancer-specific markers will provide greater diagnostic performance in terms of sensitivity and specificity. The biomarker for early cancer diagnosis will then need to be evaluated by longitudinal studies with statistically relevant numbers of human samples.

2.1.1.1 The alpha-fetoprotein story

Quite recently the FDA has approved AFP-L3, the core fucosylated form of AFP as a tumor marker for primary hepatocellular carcinoma. The performance of AFP-L3, with a sensitivity of about 50% in detection of early-stage hepatocellular carcinoma, is significantly better than that of total AFP. Furthermore, the fraction of AFP contributed by AFP-L3 has shown considerable promise as a predictor of progression of cirrhosis to cancer within the next year. These findings highlight the improvement in diagnostic efficacy of a serum glycoprotein when a change in glycosylation is examined, as opposed to looking only at the protein levels of the biomarker.

2.1.1.2 The haptoglobin story

Fucosylated haptoglobin has been reported to be a marker for pancreatic cancer. Increased fucosylation is a promising cancer marker even though the real mechanism still remains unknown. Haptoglobin is a so-called "acute phase protein," and it is well known that cancer is usually associated with inflammation. The data suggest that glycosylation of inflammatory markers such as haptoglobin is a promising target for the early detection of malignancy especially in "silent cancers" such as ovarian and pancreatic cancer which is very difficult to diagnose at an early stage. To differentiate this alteration in glycans caused by cancer it will be important to combine glycomics with proteomics to characterize the organ-specific haptoglobin glycosylation.

At the workshop there was much discussion of this most important question of carbohydrates as cancer biomarkers, with the obvious potential to collaborate and learn from the established Early Detection Research Network (EDRN; http://edrn.nci.nih.gov/) whose mission is to discover, develop, and validate biomarkers for early detection of cancer. The principles that EDRN operates by in validating diagnostic biomarkers can serve as a good model system for carbohydrate biomarkers of cancer and other diseases.

2.1.2 Glycans are potential targets for drugs

Since glycans are located at the cell surfaces of both the microorganism and the mammalian host cell, they present the perfect targets for the development of new antibiotic drugs to prevent or treat the infective process of bacteria, viruses, and fungi.

2.1.2.1 The influenza story

One of the most topical stories today is the potential threat of an influenza pandemic spread of infection in humans by a highly pathogenic avian virus. Influenza is a highly contagious, acute, viral infection of the respiratory tract. The causative agents of the disease are immunologically diverse, single-strand RNA viruses. Type A viruses are the most prevalent and are associated with most serious health risks and epidemics. Glycosylation of both the host cell receptor and the two main viral membrane proteins is intrinsically involved in many aspects of the pathogenicity of this organism.

The action of a major viral envelope protein, neuraminidase, is to cleave the sialic acid from the membrane glycolipid so the new virus particles can be released from host cells. An additional glycosylation site within the neuraminidase (NA) protein globular head has been reported to contribute to the high virulence of the H5N1 virus. In addition, the specificity of this neuraminidase has been the target for the successful development of two successful influenza drug therapies (Relenza™ GlaxoSmithKline, and Tamiflu™, Roche), which are synthetic structural analogs of neuraminic acid that bind specifically to the neuraminidase active site, and thus inhibit the transmission of the virus.

2.1.3 Glycans are potential drugs

Some of the most successful drugs currently on the market are glycan macromolecules.

Heparin (a polysulfated glycosaminoglycan) is one of the most useful (and commercially successful) drugs in medicine, being used as an injectable blood anticoagulant and as an anticoagulant coating for medical devices. Originally extracted from natural sources, its active pentasaccharide moiety has now been defined and is being produced synthetically. Also in the glycosaminoglycan family, hya-

luronic acid is the only currently acceptable treatment for the symptoms of osteoarthritis.

The recombinant protein glycopharmaceuticals are the fastest growing application of glycoproteins as it becomes clear that the therapeutic proteins (such as erythropoietin, interleukin, antibodies, CSF, factor VIII) are all glycoproteins, where biological activity is often critically dependent on appropriate glycosylation.

Importantly, the antibody-dependent cellular cytotoxicity (ADCC) therapy, in which a lytic attack on cells to which recombinant antibodies are bound is triggered by binding of lymphocyte receptors to the antibody constant Fc region, is significantly enhanced by removal of fucose or addition of bisecting GlcNAc to the IgG1 oligosaccharides. The dramatic effect of this specific glycosylation alteration demonstrates the potential of glycan-engineered recombinant antibodies as novel therapeutic candidates.

2.1.4 Glycans are biological imaging molecules

Microscopic staining has used the labeling of cellular carbohydrates for decades to explore the structure of the cell. Acidic stains such as Alcian Blue, fluorescently labeled lectins, and oxidative periodic acid Schiff stains show carbohydrates are located extensively both intra- and extracellularly.

New technology (Bertozzi, UCB) involves metabolic labeling of azidoglycans in various glycoconjugates which can then be covalently tagged, either *ex vivo* or *in vivo*, to tag glycans with imaging probes or epitope tags, thus enabling the noninvasive visualization of the location and function of glycoconjugates.

2.1.5 Specific glycans are implicated in cell signaling

The covalent modification of intracellular proteins by *O*-linked β-*N*-acetylglucosamine (*O*-GlcNAc) is emerging as a crucial regulatory signaling mechanism similar to phosphorylation. Numerous studies point to the significance of *O*-GlcNAc in cellular processes such as nutrient sensing, protein degradation, and gene expression. The involvement of this nucleocytoplasmic PTM in cellular responses to stress is key to survival following injury or disease. The 'ying-yang' phenomenon between this glycosylation and phosphorylation on the same amino acid has the potential to be an important biomarker of changes caused by disease. *O*-GlcNAc has also been implicated in type 2 diabetes mellitus.

The deletion of a fucosyltransferase that causes a lack of core fucosylation of TGF-1 receptors produces changes consistent with a deficiency in TGF-1 signaling and suggests that fucosylation defects may underlie human emphysema. The extended sugar modifications of *O*-fucose on EGF domains is another example of how an important signaling pathway can be modulated by differential receptor glycosylation, and deficiencies in fucose metabolism are known to underlie leukocyte-adhesion deficiency type II.

2.1.6 Glycans are potential targets for vaccines

Carbohydrate epitopes, or glycotopes, are present on the surfaces of cells in the body and on the surfaces of pathogens. Carbohydrate vaccines have been developed, or are being developed, for protection against many microbial pathogens in which the surface sugar antigen is variably recognized by the immune response (*e.g.*, *Leishmania*, HIV, type b Haemophilus influenzae, Neisseria meningitides, Meningococcus). However, in cancer, many of the defined carbohydrate antigens are really altered 'self' antigens and are poor immunologically. The development of successful vaccines for cancer treatment and improved immunization, however, is rapidly progressing as synthetic carbohydrate chemistry, in which the sugar epitope is being attached to immunogenic peptides and lipids, is overcoming these difficulties.

2.1.7 Glycan structures are essential for recombinant protein biotechnology products – Glycobiopharmaceuticals

It is not accidental that many of the recombinant protein drugs produced today by the biotechnology industry are glycobiopharmaceuticals, because the majority of the current products on the market are versions of hormones, cytokines, proteases, and other important pharmacologically active proteins that are excreted into the bloodstream by the organs that produce them. Carbohydrate side chains are usually found on these classes of proteins and contribute to their stability and signal recognition at their sites of action. As generic recombinant glycobiopharmaceuticals appear on the market, the FDA is increasingly requiring detailed characterization of their glycosylation to ensure their quality.

2.1.7.1 The follicle-stimulating hormone story

The natural form of follicle-stimulating hormone (FSH) is composed of two polypeptide subunits each of which has two carbohydrate side chains, such that FSH is glycosylated to the extent of about a quarter of its molecular weight. Natural variation in the structure and composition of these oligosaccharide side chains generates a large number of isoforms of FSH in humans. This heterogeneity has important biological significance in that acidic carbohydrate side chains are associated with isoforms of the hormone with longer serum half-lives. For example, removal of one acidic carbohydrate residue from FSH reduces its *in vivo* half-life from 90 min to 2–3 min.

2.1.7.2 The erythropoietin story

Erythropoietin (EPO) has become the most produced recombinant protein drug in the world because of the large number of patients benefiting from its ability to increase red

blood cell count. As such it is used extensively to counteract anemia in cancer, kidney disease and, illegally, in the sports drug abuse arena.

The protein exists as numerous isoforms as a result of the extensive heterogeneity of glycosylation on three N-linked sites and two O-linked sites. The recombinant drug form has been produced in CHO cell culture, which adds less sialylated sugars and causes the formation of different isoforms. Whereas this difference forms the basis of urinary drug testing in sports it presents a problem for the generic production of the drug as more companies start production in the wake of the expiration of the original patent. The glycosylation profile of these products can vary and is dependent upon the cell expression system as well as the culture conditions. Adding two more sites of glycosylation on erythropoietin by genetically manipulating the expressed amino acid sequence of the CHO cells results in three times the half-life in the body and increased *in vivo* activity of the drug (Aranesp®). Longer retention of the drug has the beneficial clinical effect of decreasing the required injectable administration for patients and consequently confers a significant benefit to the new product.

2.1.7.3 The ADCC and antibody therapy story

Some natural killer T cells have receptors for the Fc domain of antibodies (IgGs) and bind to the Fc portion of the IgG antibody on the surface of cells to release cytolytic components that kill the target cell. This mechanism of killing is referred to as antibody-dependent cell-mediated cytotoxicity (ADCC). Antibody therapy against tumors, such as trastuzumab (Herceptin®) and rituximab (Rituzan®), requires both activation *via* Fc gamma RIIII and inhibition *via* Fc gamma RIIb antibody receptors. The addition of bisecting GlcNAc to the recombinant antibody glycans leads to an increase in ADCC through a 10–20-fold higher affinity for Fc gamma RIII. Similarly, deletion of core fucose from IgG1 oligosaccharides has been seen to enhance ADCC activity by up to 50–100-fold.

2.2 Glycomics-focused approach

2.2.1 A focused effort on glycomics is needed for the exploitation of the glycome as a source of disease markers and drug targets

Glycomics is typically neglected in the framework of proteomics and lipidomics. Researchers often cleave glycans from proteins or other conjugates and ignore the glycan component, as it is perceived that oligosaccharides are difficult to analyze. Despite the fact that over 50% of all proteins are glycosylated, this elimination of critical information is considered necessary for the rapid analytical throughput that characterizes the "-omics" revolution. In addition, large glycoproteins that are highly glycosylated, including mucins, are similarly neglected, since current methods are still incapable of characterizing these large, highly heterogeneous but important compounds which have been widely implicated in epithelial diseases.

2.2.2 A glycan-centered approach is necessary to monitor specific changes in glycosylation

These glycan changes may occur globally or on specific glycoconjugates. The analytical protocols and the bioinformatics methods are sufficiently from proteomics dissimilar that efforts specifically for glycans are necessary. Nevertheless, we can apply lessons learned from proteomics science to develop robust glycan-specific standardization, data analysis, and informatics tools.

New glycan-specific technology and informatics can advance the discovery of glycomics-associated disease and drug processes and speed the development of new diagnostics and drugs

There is now new technology to address the analysis of the micro- and macroheterogeneity inherent in glycoproteins. Advances in 1-D and 2-D NMR spectroscopy, chromatography, CE, MS (MALDI and ESI MS/MS), and microarrays of both glycans and lectins now yield comprehensive and accurate insight into not only the monomeric sugar compositions of the glycan side chains of glycoconjugates, but also insight into the most probable structures of the side chains, including branching patterns, the location of charged and neutral moieties, and of their location on the protein and/or other conjugate.

2.2.3 Glycomics is complementary to proteomics and other -omics

Glycomics has several distinct advantages that make it well suited for disease biomarker discovery:

(i) Changes in glycosylation can be more distinct than changes in protein expression. Specific glycan structures that are not present, or are in low amounts, in normal state proliferate in disease states.

(ii) Changes in glycosylation involve many proteins including those that are highly abundant. For example, most secreted glycoproteins are produced in two areas of the cell, the ER and the Golgi body. Therefore, a single change in a cell's glycosylation machinery can affect the many different glycoconjugates.

(iii) The location of the glycans on the cell surface makes them the first point of contact of cellular interactions and thus crucial in the control of normal metabolic processes. Disruption to these cell–cell interactions is intrinsic to many diseases and provides specific diagnostic and target biomarkers. Cell surface molecules are also strategically exposed for surveillance by the immune system allowing for the potential of immune recognition of abnormal cells.

3 Defining glycomics

3.1 What is glycomics?

The glycome is the entire oligosaccharide constituent of all glycoconjugates from a single or defined biological source. In mammals it includes all oligosaccharides from glycoconjugates such as glycoproteins, glycolipids, those attached to glycosylphosphatidylinositol anchors, and proteoglycans. It may also be defined to include free glycans found in bodily fluids such as serum and milk.

Glycomic analysis ("glycomics") is the characterization of the glycome. This effort often requires the release of all glycans from the corresponding glycoconjugates. However, it is often difficult to profile the complete glycome because of the large diversity in the structures and functions of oligosaccharides. The glycomics approach will therefore always be targeted to specific groups of oligosaccharides and groups of glycoproteins and glycoconjugates.

Functional glycomics requires both the characterization of the glycans and their corresponding attachments to proteins or lipids in order to determine the role of glycan heterogeneity in disease.

Glycomes can vary among animals, plants, and microorganisms that impact human biology, with different monosaccharide constituents and structures occurring that can differentiate these species from the human glycome and provide new diagnostic and drug targets.

3.2 Glycans are characterized by macro- and microheterogeneity

The major characteristic of glycosylation is its large diversity. The classification of oligosaccharides is difficult to define explicitly, as differentiation can include function and structures or both. There are a number of traditional ways to partition oligosaccharides into specific groups such as by biological function, by the attachment to the peptide backbone, by the types of glycoconjugates, and by the molecular polarity (neutral *vs.* anionic). For example, glycans attached to serine or threonine of the peptide backbones (O-linked oligosaccharides or O-glycans) can be released together. However, glycomic analysis of O-glycans requires separation of the highly anionic glycans, such as glycosaminoglycans (GAGs), from neutral and less anionic species, because the distinct physical properties of these groups of glycans make it necessary to use different analytical methods for their characterization. Fortunately, these physical properties often facilitate the required separation.

3.2.1 N-linked oligosaccharides are those that are attached to the peptide backbone through an asparagine

N-linked oligosaccharides (N-glycans) are found predominantly with the consensus sequence N-X-S/T, where X is any amino acid but proline. While a consensus sequence is necessary for N-glycosylation, not all consensus sequences are occupied. N-glycans have a single common core composed of two *N*-acetylglucose amine and three mannoses. Their analysis is facilitated by the existence of an enzyme which cleaves the linkage between the N-glycan and the protein (PNGase F).

Changes in N-glycans, such as degree of sialylation, addition of bisecting *N*-acetylglucosamine, and fucosylation of N-glycans, have all been reported to occur as a consequence of disease. For example, the differential sialylation on transferrin is used as a diagnostic of alcoholism.

3.2.2 O-glycans are those that are connected to the peptide backbone through serine or threonine

This group generally includes protein O-glycans as well as the GAGs released from proteoglycans, but because the biological function and the chemical characteristics of the latter are so different, they are separated into their own group. O-glycans are often released by alkaline sodium borohydride and include many of the mucin oligosaccharides.

There is a significant body of work that shows changes in disease states produce aberrant glycosylation in O-glycans. Attachment of an O-glycan does not depend on a specific consensus amino acid sequence on the peptide backbone. There is no single core, although most are connected to the peptide *via* an *N*-acetylgalactosamine. There are currently eight known core structures for O-glycans and the structures can vary significantly. In humans, O-glycans have common residues that include *N*-acetylglucosamine, *N*-acetylgalactosamine, galactose, fucose, sialic acid, and sulfate.

The potential for O-glycans, or the O-glycan/peptide epitope, as biomarkers is high. There are a number of studies that show aberrant glycosylation in disease is common among mucins, which make up an important constituent of this group. Mucins are large (1–3 MDa) highly O-glycosylated proteins that are frequently ignored in current proteomic approaches because of their size and large heterogeneity. In addition, the monosaccharides which attach the O-linked sugar to the protein are being found to be more diverse than the mucin-type *N*-acetylgalactosamine linkage, with mannose, fucose, glucose, and xylose identified as the linker in different tissues.

3.2.3 Glycosaminoglycans and other highly anionic glycans

GAGs are the covalently attached oligosaccharide chains of proteoglycans (PGs), and confer many of the biological functions of PGs. Some of their important roles include cell signaling, tissue development, inflammation, and cartilage integrity. Proteoglycans are vital components of articular cartilage where they provide resilience against compressive forces in the joints. The most characteristic feature of these proteoglycans is the presence of GAGs, which consist of

tandemly repeated, often sulfated, oligosaccharide chains. In normal cartilage pathology, the most predominant PG, aggrecan, interacts with hyaluronic acid and collagen fibers to provide a stable supportive scaffold. The negatively charged sulfate groups of chondroitin sulfate (CS) on aggrecan create a strong electrostatic repulsion leading to a hydrophilic environment that contributes to cartilage resistance to compression. However, in disease states such as osteoarthritis, there is a breakdown in proteoglycans, causing a disruption in the balanced framework in cartilage. One of the first hallmarks of disease is the release of proteoglycan protein and GAG fragments into the synovial fluid. Early diagnosis of osteoarthritis requires a better understanding of the structural complexity of these GAGs in order to develop more sensitive and specific methods to detect their fragments in the patient.

3.2.4 Glycolipids

Glycolipids are glycosyl derivatives of lipids such as acylglycerols, ceramides, and prenols. Their structure and function have been implicated in a wide range of immune-related human diseases.

For example, inherited deficiency of glucocerebrosidase, a lysosomal hydrolase, results in Gaucher's disease. Patients with Gaucher's disease have altered humoral and cellular immune profiles and increased peripheral blood natural killer T lymphocytes.

Fabry disease is caused by a deficiency of α-galactosidase A which leads to the progressive intralysosomal accumulation of ceramide trihexoside, also known as globotriaosylceramide (Gb3), in different cell types and body fluids. The clinical manifestations are multisystemic and predominantly affect the heart, kidney, and central nervous system. The recent introduction of a specific treatment for Fabry disease in the form of enzyme replacement therapy has led to the need for a biological marker, such as measuring the concentration of the accumulated sugar, ceramide trihexoside, for evaluating the efficacy of treatment and also as a tool for following the long-term effects of treatment.

There are many such examples in which the determination of the glycan structures associated with the whole lipid class of molecules, and their changes in disease, requires specific technologies and informatics to be developed to be able to exploit them as biomarkers or treatment targets.

3.3 Bacterial, viral, and fungal glycosylation

The study of the glycome of pathogenic bacteria, viruses, and fungi can also provide an avenue of biomarker discovery. There remains a general misconception that bacteria do not glycosylate proteins whereas, in fact, most bacteria have extensive glycosylation machinery used in the synthesis of their cell walls (lipopolysaccharides, peptidoglycans) and capsular glycoproteins (as evidenced by the Bacterial Carbohydrate Structure DataBase developed at the Carbohydrate Chemistry Lab of the N.D. Zelinsky Institute of Organic Chemistry (Moscow, Russia); http://www.glyco.ac.ru/bcsdb/start.shtml). An example of the usefulness of these sugars as biomarkers is provided by the diagnostic potential of the specific cell wall lipoarabinomannan excreted in the urine of patients infected with TB (*Mycobacterium tuberculosis*).

As described in Section 2.1.2.1, the mechanism and inhibition of viral influenza infection is wholly dependent on the glycosylation interactions between the virus and the host mammalian cell. In addition, the use of the eukaryotic yeast and fungal expression systems (*Pichia* and *Trichoderma*) as economical producers of human recombinant proteins is dependent on the genetic modification of their glycosyltransferases.

4 Immediate needs for glycan biomarker discovery

The discovery of glycan disease markers will be aided by methods capable of global analysis and methods that can identify specific glycans. An annotated structure bank with compiled physical properties that make it easy to recognize previously annotated structures is key. Because oligosaccharide mixtures are characterized by high structural micro- and macroheterogeneity, separation methods that can deal with the large structural diversity are necessary. The separation methods must also be specifically sensitive, as oligosaccharides often do not contain highly chromophoric substituents for spectroscopic detection or highly ionizable substituents for MS. For these reasons, many of the methods developed for protein and peptide analysis often fail for oligosaccharides. Structural elucidation methods must also be able to deal with limited amounts of material while providing as much information as possible. Application of promising methods such as MS/MS to glycans requires further development, as the fragmentation chemistry of these complex, branched molecules is distinct from that of peptides, complicating the interpretation of their mass spectra.

A multi-institutional study conducted by the Human Proteome Organisation (HUPO) HGPI (human disease glycomics/proteome initiative) in 2006 indicated that MS-based analysis appears to be the most efficient method for identification and quantitation of oligosaccharides in glycomic studies and endorsed the power of MS for glycopeptide characterization with high sensitivity in proteomic programs.

4.1 New and novel methods for rapid structural elucidation

The large diversity in structures presents problems unique to the glycome that will not be solved by a single set of tools and will involve a number of different analytical tools developed specifically for glycans. The diversity of glycans and glycosylation enzymes distinguishes glycomics from genomics and

proteomics, where a small number of rapid, highly sensitive analytical methods have come to dominate. Glycans include distinct structures that are branched, neutral, and anionic with different anomeric linkages. They may also include highly anionic polymeric material that requires different sets of tools and new methods for analysis.

4.1.1 Mass mapping strategies are necessary for the rapid assessment of aberration in glycosylation

Changes in glycosylation can sometimes be sufficiently determined based solely on the composition. For this reason, mass mapping strategies will be useful for observing changes in glycosylation such as the number of fucose and sialic acids. High mass accuracy techniques with high sensitivity are necessary for rapid compositional analysis, albeit this approach will not differentiate between such residues as galactose, mannose, and glucose or between *N*-acetylglucosamine or *N*-acetylgalactosamine.

4.1.2 Methods for profiling glycans with separation techniques will also be key for identifying specific glycan markers

Complete separation of oligosaccharides is rarely accomplished using current methods, many of which were developed for peptides. Oligosaccharides require their own separating media such as graphitized carbon and ion exchange. Many of these media need to be paired with emerging separation methods such as nanoflow LC and CE to separate complicated glycan mixtures into individual components while employing minute amounts of material.

4.1.3 Structural elucidation methods that provide structural information with high sensitivity such as MS are necessary for the further development of structural libraries

Mass spectrometric fragmentation including collision-induced dissociation and laser-induced dissociation provide structural and compositional information. Other forms such as electron capture or electron transfer can provide some information regarding the position of glycosylation in peptides and small proteins. In this effort, bioinformatics methods that can annotate MS spectra and predict structures will significantly improve the analysis and could provide high-throughput methods for structural elucidation.

Glycosidases coupled with arrays can provide sensitive and precise methods of analysis that complement the MS approaches. Glycosidases are enzymes that cleave specific glycan linkages providing key information including the identity of the residue, the linkage, and the anomeric character. However, the number of commercially available glycosidases is currently small and limits the application of these methods.

4.1.4 The continued development of large glycan arrays and the corresponding glycan-specific antibodies

Glycan arrays provide the potential for bedside diagnostics once specific markers are identified for the disease. While major efforts are underway in this area, these arrays need to be further expanded and need to include other types of oligosaccharides such as GAGs.

Synthetic efforts making new oligosaccharides are necessary for the arrays as well as for the development of more glycan-specific antibodies. Sugars are found to be fairly non-immunogenic without an attached peptide and they often raise an unstable IgM response rather than the more robust IgGs. Methods for the improved immunogenicity of glycans to improve antibody production and stability for their use in diagnostic biomarker testing are required if they are to be commercialized.

4.1.5 New and novel methods for the determination of site-specific glycosylation in glycoproteins

New approaches for the determination of site-specific glycosylation in glycoproteins are necessary for tracking disease-specific changes at the level of glycosylation sites. While glycan profiles are currently obtainable, the determination of site-specific glycosylation remains a difficult task. Improved methods are needed to determine the identity of the glycoproteins, the sites of glycosylation, and the glycan heterogeneity at each site. The methods must also be able to determine when the site is fully or partially unoccupied. Methods that examine only the polypeptide aglycon while discarding the glycan moiety provide a very limited subset of the information required for a proper "glycoproteomic analysis." Site-specific glycomics analysis may be the most difficult aspect of glycomic analysis from the technological point of view, but is necessary for finding markers with greater disease specificity. A single site and its associated glycan may provide the most precise marker for specific diseases.

4.2 Bioinformatics

New bioinformatics approaches are required to describe the structures, biosynthesis, and functions of glycans, because methods developed for genomics and proteomics are not directly applicable to these complex, branched molecules. Due to the importance of bioinformatics in glycomics, 50% of the discussion at the workshop was devoted to this topic. Two general aspects were emphasized from the perspective of biomarker research.

4.2.1 An annotated and curated library of fully and partially characterized glycan structures

(i) An annotated and curated structure database of both fully and partially characterized glycans is necessary to

allow the identification of new and partially known structures. This database will then form the core component of a glycan information library (glycoKnowledgebase), which must contain other physical characteristics of structural analysis such as MS fragmentation spectra, NMR spectra, and LC retention times on different systems as well as glycan biosynthetic pathways and glycosidase and glycosyltransferase expression and activity data. In this way, the structural and functional elucidation of specific oligosaccharides can be performed by several groups working independently but employing the information that has already been obtained by other groups.

(ii) It is essential that this library attracts funding to ensure that quality and consistency are built on and guaranteed into the future. Much of the progress accomplished in previous efforts in this area has been lost or compromised due to lack of continued funding.

4.2.2 Informatics methods for interpretation of glycan MS and MS fragmentation spectra

Key to the efforts in structural analysis are bioinformatics tools that can rapidly interpret MS data. These tools will be able to examine mass profiles for composition and MS fragmentation spectra to provide rudimentary or complete structures. Software that can perform these tasks will greatly advance the efforts in glycomics as it has done in proteomics.

5 The urgent requirement for glycan-specific databases and informatics

We are just beginning to understand the importance of carbohydrates in biological information transfer and storage. New knowledge regarding the structural, functional, and physiological aspects of glycans that is gained from high-throughput glycomics experiments will influence future research in ways that are as far-reaching as the advances in our knowledge of genes and proteins have been during the last decades. Similar to genomics and proteomics, the availability of well-structured and curated databases, along with efficient and user-friendly retrieval and analysis software, is of paramount significance for the rapid development of glycobiology. It is likely that the availability of appropriate informatics tools that enable efficient correlation of glycomics data with the other biomedical data will engender a synergism that leads to numerous discoveries that directly impact the diagnosis and treatment of human disease.

5.1 Current status of informatics for glycosciences

The development and use of informatics tools and databases for glycobiology and glycomics research have increased considerably in recent years. However, this field must still be considered as being in its infancy when compared to genomics and proteomics. For example, no *comprehensive* carbohydrate data collections similar to those currently available for genomic and proteomic data have been compiled so far. There is currently no location where information about all carbohydrates reported in refereed scientific papers is systematically stored. Procedures (similar to those for protein sequences) have not yet been established for scientists to report the observation of specific glycan structures in specific environments and to store these observations in a generally accepted database.

5.2 A short history of glyco-related databases

The need to establish a centralized database of all carbohydrate structures published in refereed scientific journals was recognized during the mid-1980s. The driving force behind this initiative was to find easily all publications in which specific carbohydrate (sub)structures are reported. This initiative resulted in the 'Complex Carbohydrate Structure Database' (CCSD)—often named CarbBank according to the retrieval software to access the data—which was developed and maintained by the Complex Carbohydrate Research Center of the University of Georgia (USA). The project was funded by the NIH. The need to install CarbBank as an international effort was clearly recognized and resulted in worldwide curation teams responsible for specific classes of glycans. During the 1990s, a Dutch group assigned NMR-spectra to CCSD entries (SugaBase). Due to a variety of reasons, the funding for CCSD stopped during the second half of the 1990s. Consequently, CarbBank was not further developed and the CCSD was no longer updated. Nevertheless, with 49 897 entries, which correspond to 23 118 distinct glycan structure graphs, the CCSD is still the largest publicly available repository of glycan-related data. All subsequent open access projects initiated at the beginning of the new century made use of the CCSD data.

5.3 Past CarbBank

Although the collapse of CarbBank was a setback, a small informatics-oriented group of scientists at the DKFZ (German Cancer Research Centre), initially interested in elucidating the conformational space of complex carbohydrates, first postulated the imperative to develop informatics for glycobiology as an independent sub-branch of bioinformatics. This group also realized the need to make the CCSD entries publicly available using modern Internet-based tools and to cross-reference the glyco-related data with proteomics and glycomics information. These ideas have led to the development of the GLYCOSCIENCES.de portal and the EUROCarbDB project.

At the beginning of the new century when the gap between encoded and published glycan structures became obvious, several companies started to provide commercial access to glyco-related data, which they extracted from the literature. None developed a successful business model, and only the Australian *GlycoSuiteDB* survives today.

5.4 Glycomics given a new stimulus

An important stimulation was the establishment of the international Consortium for Functional Glycomics (CFG) funded by the US National Institute of General Medical Sciences. It was the first large-scale project that clearly emphasized the need for informatics to manage and automatically annotate the vast amount of experimental data generated by glycomics research. The development of algorithms for the automatic interpretation of MS spectra—a severe bottleneck that hampers the rapid and reliable interpretation of MS data in high-throughput glycomics projects—is critical for all glycomics projects. This is still the most active area of software development, where various experimentally oriented groups have been developing software solutions and algorithms to solve their specific scientific questions.

Another important step was the integration of glyco-related biological pathways into the schemata of the first 'classical' bioinformatics initiative—the Kyoto Encyclopedia of Genes and Genomes (KEGG). Subsequent development of associated databases for glycan structures led to the KEGG GLYCAN approach, which elegantly established the connection between glycan structures and the knowledge of enzymatic reactions to build the glycan structures. Additionally, the KEGG group made significant progress to apply bioinformatics algorithms to the tree-like structures of glycans for comparison and alignment, to develop similarity scores, and to establish a global view of all glycans belonging to related pathways.

As a consequence of the increasing interest in glycomics research, various new databases were started in recent years (see, *e.g.*, the link list at www.eurocarbdb.org/links/). Among these the *EUROCarbDB* project (distributed bottom to top initiative for primary experimental data), the Russian *Bacterial Carbohydrate Structure Database* (aiming to cover all known structures), and the '*Bioinformatics for Glycan Expression*' initiative (development of glyco-related ontologies) of the *Complex Carbohydrate Research Center* are the larger ones. In general the development of glyco-related related tools and databases can be described as a small but quite active field of research.

5.5 Current situation

The current situation in glycoinformatics is characterized by the existence of multiple disconnected and incompatible islands of experimental data, data resources, and specific applications, managed by various consortia, institutions, or local groups. These resources rarely provide communication mechanisms that would leverage these data by allowing their combination and comparison. However, approaches to link the distributed data have been conceptually worked out and examples are already implemented. The collaborative spirit recently exhibited by all of the major glycomics initiatives will significantly help to overcome this unfavorable situation. This positive spirit has recently led to an important milestone, the agreement of an XML standard for the exchange of glycan structures (GLYDE-II).

None of the existing initiatives had the capacity to completely fulfill the mandate of *CarbBank* at the beginning of the 1990s, *i.e.*, to provide comprehensive access to all published carbohydrate structures. In particular, the existing initiatives did not have the worldwide resources to fill the gap of published glycan structures that were not included in *CarbBank* after its termination in the mid-1990s.

It is likely that the tendency to set up local databases designed to support specific areas of research in glycobiology will continue in the near future. The existence of a centralized glycan structure database would substantially increase the ability to annotate and cross-reference local data with other bioinformatics resources. Offering clear guidelines describing the minimal requirements of data exchange formats, which are required for databases to communicate with each other, will hopefully lead to strong interconnections and compatibility among glycobiology and glycomics databases.

5.6 What are the urgent next steps and what can be done immediately to assist in glycan biomarker discovery?

5.6.1 Centralized database for carbohydrate structures

There was a general agreement expressed by many speakers at the NIH workshop that there is an urgent need for a unified, thoroughly curated and sustainable database for carbohydrate structures in biological samples. This lack of appropriate databases is regarded as the biggest defect in glycomics and glycobiology research. "*We need to be able to search databases for what is out there. Imagine genomics and proteomics without GenBank*" (Ajit Varki).

To pave the way for a central carbohydrate structure database, the existing larger initiatives agreed to immediately start with the necessary preparatory steps for the conversion of CarbBank data into the GLYDE-II format. This will be a multi-institutional, international effort, which will be coordinated by the EUROCarbDB/GLYCOSCIENCES.de initiative. The result of the new conversion of CarbBank will provide a clean dataset of fully determined glycan structures in GLYDE-II format. This dataset will constitute the state-of-the-art repertoire of available digital glycan structures. These structures will also constitute the foundation for the future centralized database.

5.6.1.1 The benefits of converting CarbBank data into the GLYDE-II format

The objectives of the conversion to GLYDE-II format will be:

(i) an ideal test set of glycan structures to assess the robustness of the GLYDE-II exchange format as well as its implementation(s).

(ii) the unique definition of the monosaccharide namespace as well as the description of the topology of glycans so that all other database projects can refer to a unique encoding.

(iii) the generation of a new glycan structural dataset, which fulfills clearly defined quality criteria (Gold Standard) based on the experiences gained during the previous conversions of CarbBank entries and the recognized imperative to cross-reference glycomics data with genomics, proteomics, and other established classifications systems.

To achieve the above outlined objectives, the glycosciences community must endorse supplementary standards for data exchange. Obvious requirements are to agree to XML descriptions for the exchange of bibliographic references and to controlled vocabularies for biological species, tissues, cells, and diseases. The general philosophy will be that well-established XML formats (*e.g.*, the MEDLINE®/ PubMed® XML) and vocabularies (*e.g.*, the NCBI Entrez Taxonomy to uniquely describe species) are used for these purposes. To describe the noncarbohydrate attachments at the reducing end of glycans, references/cross-links to well-established databases specialized for such molecules (proteins, lipid, and small organic molecules) will be given. This will allow the structures of complex glycoconjugates to be stored by specifying pointers to specific records in these established databases along with the sites linking the glycan to the noncarbohydrate moiety.

5.6.1.2 A central database for carbohydrate structures will require long-term maintenance and operation

Discussions at the NIH workshop clearly pointed out that the long-term maintenance of a highly curated database that summarizes all results that have been reported in the literature is clearly beyond the scope of the existing larger projects in the USA, Japan, and Europe. Such a central database requires robust manual annotation and curation tools in the hands of qualified experts in glycan structural analysis. The publicly available data in glyco-related databases reflect a gap in knowledge of more than 10 years accumulated by the community after termination of the CarbBank in the mid-1990s. It will be a major effort to close this gap.

In a recently published paper the European Strategy Forum for Research Infrastructures (cordis.europa.eu/esfri/) presented a roadmap emphasizing that "*modern science is inconceivable without recourse to well structured, continuously upgraded (...) and freely accessible databases (...) The infrastructures required are often multi-sited, they are mainly data collection, storage and access systems which not only require long term maintenance and operation, but also continuous upgrades.*".

An obvious demand of such a database is the assurance that the included data will be maintained and made available over the long term. Therefore, the new repository should be located at or closely associated with a well-recognized international nonprofit academic organization that provides open access to biological and experimental data. A central database will assure data consistency, systematic annotation, and cross-referencing with other bioinformatics resources.

5.6.1.3 Urgent needs to organize the input of glyco-related data during the process of publication

As newly published data are generated, it will be necessary to establish procedures similar to those routinely used in genomics and proteomics research, allowing scientists to directly enter structural data and biological annotations during the publication process. Rapid progress in glycobiology will depend on the acceptance of such procedures by glycoscientists. It is likely that many of the required guidelines and standards can be adopted on the basis of protocols that are worked out during the above-mentioned conversion of *CarbBank*. This should entail a specification of the minimal amount of information required for the publication of glycan-related data. Similarly to the MIAME standard for the publication of microarray expression data, such a specification will be important for the replication of the published data. In addition, it will be necessary to develop user-friendly software tools and user interfaces, especially for the input of glycan structures. Scientific editors and publishers as well as the thematically related societies should be involved in the discussion at an early stage.

5.6.2 Support for distributed, federated databases for primary experimental data

There was a general consensus at the NIH meeting that the task of creating a centralized glycan structure database should be separated from the mission to collect the associated primary experimental data. The volume and diversity of glycomics data make it necessary to distribute it in different locations throughout the world. This approach allows those having the technical expertise required for data generation to maintain close ties with the data and its curation. However, such a system can work only if robust standards for data transmission are developed and accepted by the scientific community. The NIH workshop constitutes a significant step forward in this respect. (See Supporting Information for a more detailed description.)

5.6.3 Support for the development of open source software projects for glycomics

To achieve the above-mentioned goals it is obvious that the endeavors pulling together databases have to be combined with efforts to develop robust and user-friendly software to access and analyze the data deposited in the emerging databases. Currently, only a limited amount of

software related to glycomics is freely available to be shared by various projects. The largest bottleneck until now was the lack of a common language to exchange glycan structures and related data. The currently existing software is consequently based on proprietary formats and no mechanism enabling an easy exchange of data is foreseen. With the agreement to accept GLYDE-II as the central format to exchange structural data, a central prerequisite for the development of glyco-related software engineering resource glycomics has been achieved. It will be of paramount significance that software implementations using the GLYDE-II format will be publicly available in the near future.

Experience in other fields of software development and engineering has demonstrated that the *Open Source* philosophy favors the rapid creation of robust solutions within an open, collaborative environment. Its underlying philosophy is that source code is available for anyone to use, modify, and redistribute freely. *Open Source* projects have shown to promote a higher standard of quality, and help to ensure the long-term viability of both data and applications. (See Supporting Information for a more detailed description.)

6 Summary of statements and recommendations of the workshop

6.1 Recommendation 1: Develop a robust, centralized database of curated glycan structures

The focus group assigns the highest priority to the implementation of a thoroughly curated repository of carbohydrate structures. Each record in the database will contain (i) a single glycan structure that meets well-defined confidence criteria and (ii) provenance information for the structure, including literature references, a description of its biological source, and/or the primary analytical data used in its assignment. (See Recommendation 2.) The new database should be closely associated with a well-recognized international nonprofit organization that provides open access to biological and experimental data and has funding to be able to maintain and curate the entries into the future.

In contrast to the genomic and proteomic areas, no comprehensive, up-to-date data collections for carbohydrates containing carefully curated data that summarize results reported in the literature have been compiled so far. This lack of appropriate databases is regarded as the biggest defect in glycomics and glycobiology research. There is an urgent need for a unified, thoroughly curated and sustainable database for carbohydrate structures in biological samples. This "core" of structural information will provide the foundation for a wide range of glycomics and glycobiology research and act as an anchor that unites the experimental data that support the identification of specific structures. A central database requires robust manual annotation and curation tools in the hands of qualified experts to maintain database consistency.

6.2 Recommendation 2: Develop an infrastructure to implement a worldwide network of databases containing experimental and analytical data relevant to the structures and functions of glycans

The focus group assigns the highest priority to the creation of a bioinformatics infrastructure supporting worldwide database networks for primary experimental data. This builds on a major achievement of the workshop, the agreement to use GLYDE-II as a common structural data exchange format. Additional data representation standards must be adopted along with guidelines for good laboratory practice and quality control procedures. Robust database models must be developed and supported.

Databases that include experimental results at various stages of processing fulfill a fundamentally different purpose than the glycan structures database. They include raw data and algorithmically and/or manually interpreted/annotated data, for example, to associate specific structures with specific spectral features or biological functions. For these data to be useful, these annotations must include detailed information regarding the biological source, the complete history of sample preparation, instrumentation methods, and post-acquisition processing techniques. The focus group believes that, due to the potential enormity and diversity of this type of data, it is best maintained in a distributed form at the site of its generation or in designated "satellite" sites that specialize in certain analytical techniques (MS, NMR, HPLC, CE *etc.*).

6.3 Recommendation 3: Support the development of analytical tools specifically for glycans and glycoconjugates

The focus group believes that the analytical technology available for the specific analysis of glycoconjugates is lagging behind that of the technologies available to the scientific community for the study of genomics and proteomics and their function in disease and assigns the highest priority to the support of the development of glycan-specific analytical tools.

The primary method used for the discovery of biomarkers of disease currently is MS and this rapidly developing technology needs to be exploited for its capacity in glycomics analysis. In addition, the development of glycan arrays needs to be expanded to include other glycoconjugates such as glycolipids and GAGs. The availability of purified glycosidases isolated from different sources for use in structural elucidation along with the discovery and characterization of new glycan-specific lectins and antibodies needs to be supported to facilitate the conversion of the discovered glycan biomarkers into functional diagnostic assays.

6.4 Recommendation 4: Support the development of open source software for automated analysis of analytical data and data mining in the glycomics domain

The focus group assigns a high priority to supporting open source software projects that will provide robust solutions for often-required functions in glycomics research. These include software for the automated interpretation of high-throughput analytical data (such as mass spectral data) and data mining tools that facilitate, for example, the discovery of correlations between glycan structure and function.

There is the lack of freely available robust software for glycomics applications and data analysis. The glycosciences community will significantly benefit from free software and Web services that are robust and applicable to a wide range of glycomics-related questions. There is broad potential for software development including tools for (semi)automatic assignment of experimental data, automatically generated semantic annotations, and graphical user interfaces that allow manual annotation of the data. Additionally, approaches for retrieval and knowledge discovery such as tools for similarity searching as well as glycomics-specific data mining are urgently needed. The demand to have powerful software at hand will increase significantly with the availability of a centralized glycan structure database as well as with the available experimental data deposited in the associated federated networks of databases.

6.5 Recommendation 5: Facilitate the transition from glycan discovery to validated diagnostic biomarkers

Once the glycan biomarkers are identified and characterized in various disease applications, their usefulness can only be realized if there is support for validating them in statistically relevant clinical samples. A high priority to progress the translation of glycomic biomarker discovery to the bedside will involve access to statistically significant numbers of human patient samples and sample tissue banks as well as access to model disease systems.

This recommendation can be facilitated by integration with the existing NIH protocols (*e.g.*, EDRN) for validating genomic and proteomics diagnostic biomarkers. Collaboration with clinical groups that can organize sufficient numbers of well-documented disease-specific samples of tissues and biological fluids is essential.

6.6 Recommendation 6: Invest in the education and training of young scientists as future leaders of glycomics research

The focus group strongly recommends investing in interdisciplinary educational programs aimed at training scientists in all aspects of glycoscience and glycomics.

The opportunities of glycomics research can only be achieved if a sufficient number of well-trained scientists can be attracted to work in this field. The need for trained scientists is especially great in the area of informatics for glycomics. There is a clear need for additional investments in education and interdisciplinary programs for training young scientists in this area. Informatics education for glycobiologists must encompass a broad variety of disciplines that extend beyond fundamental information technology to include computational methods for determining molecular structure and molecular interactions. Training in the development and use of functional assays and spectroscopy/spectrometry is also necessary in the context of (systems) glycobiology and glycomedicine. Such training can be accomplished by programs ranging from summer schools to specialized PhD and MD/PhD courses.

We would like to thank Dr. Pamela Marino, Program Director at the US National Institute of General Medical Sciences, for her vision and organizational skills.

We would also like to acknowledge the input from all the participants at the Glycomics workshop and thank those who actively contributed to the discussions, both during and after the meeting.

A short summary report of the workshop presentations has been published earlier in Proteomics: Taniguchi N., Paulson J. C., Proteomics 2007, 7, 1360–1363.

The authors have declared no conflict of interest.

Note added in Proof
Vale Claus-Wilhelm (Willi) von der Lieth

The glycobiology community is saddened to have to report the unexpected death of Dr. Glaus-Wilhelm (Willi) von der Lieth on November 16, 2007. Dr. von der Lieth, of the German Cancer Research Center, Molecular Structure Analysis Group in Heidelberg, was a global leader in the development of informatics systems for glycobiology. His effectiveness in this area stemmed from his deep understanding of the structural complexity of glycans and the need for novel computational approaches for analysing and understanding the biological relevance of these structures. He had a great enthusiasm for sharing ideas and worked tirelessly toward the goal of a robust worldwide infrastructure for glyco-informatics based on the philosophy of open access within the area.

To this end, Willi was the inspiration behind the EUROCarbDB project, and his vision, coupled with his professional and personal skills, were seminal in encouraging many bioinformatics groups to integrate their databases. The continuation of these efforts by the members of the consortium, as he would have wished, will provide a lasting resource for the field in his memory.

Willi will be greatly missed both as a friend and a colleague. His willingness to do more than his share of work to bring together different approaches for computational glycobiology has really made a difference to the field and we owe him a great deal of gratitude for his efforts. Perhaps the most fitting tribute to Dr. von der Lieth will be the realization of his dream, which will nevertheless be more difficult without his leadership.

