

Emerging glycomics technologies

Jeremy E Turnbull & Robert A Field

The nascent field of glycomics is currently undergoing rapid development, largely as a result of advances in technologies for analyzing glycan structure, unraveling glycan-protein interactions and establishing the functional significance of glycans. A meeting was held in November 2006 to explore the challenges and opportunities ahead for this emerging 'omics' domain.

As a field, glycobiology encompasses the determination of the structure and function of complex sugars (glycans). This has become a critical facet of postgenome science—in particular, proteomics, as many proteins are post-translationally modified by glycosylation, and these modifications alter and regulate biological activities. Thus, glycans represent a major class of post-translational modifications that dramatically enhance the functional diversity of proteins (Fig. 1). In recent years the vision of glycobiologists has increasingly turned to the concept of the 'glycome'—the complete set of glycan structures expressed by specific cells, tissues or organisms. This has in turn led to a need for analysis of larger numbers of glycan structures, which has provided the impetus for development of technologies with high-throughput potential.

The emerging omics domain of glycomics has lagged behind that of genomics and proteomics, mainly because of the inherent difficulties in analysis of glycan structure and function¹. However, the field is now rapidly advancing on a number of fronts. A one-day conference entitled "Glycomics: Challenges and Technologies" was organized by Euroscicon, chaired by Professor Jeremy Turnbull (University of Liverpool) and held at Imperial College, University of London on November 28, 2006 to explore the future challenges and emerging technologies of glycomics. The plenary speakers were experts

*Jeremy E. Turnbull is in the School of Biological Sciences at the University of Liverpool, Liverpool L69 7ZB, UK, and Robert A. Field is in the Department of Biological Chemistry, John Innes Centre, Norwich NR4 7UH, UK.
e-mail: j.turnbull@liverpool.ac.uk or rob.field@bbsrc.ac.uk*

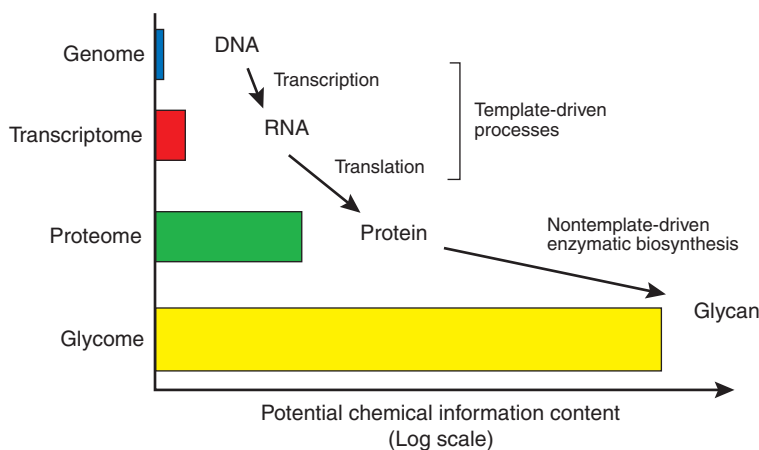


Figure 1 Glycome enhancement of the molecular and functional diversity of the proteome. Protein expression is based on a genetically encoded template, but post-translational modifications of proteins dramatically enhance their functional diversity. The glycome represents the main class of post-translational modifications, providing biological access to vast information space at minimum genetic cost. Note that in the information flow from the genome to the glycome, the biosynthesis of glycans is not encoded via a template-driven system. Note also the log scale of potential chemical information.

from a range of fields that converge in the multidisciplinary topic of glycomics, with several additional short talks invited from submitted abstracts. Here we highlight the key themes that emerged from the meeting.

What is glycomics?

The meeting was opened with an introductory presentation and an open discussion led by Turnbull, who introduced the themes of the meeting and posed the question "What is glycomics?" It was recognized that glycomics is much more than just a catchy new title for the general field of glycobiology. The key feature of glycomics is the move toward large-scale analysis of the structure and function of diverse glycans, and ultimately of whole gly-

comes. In this respect glycomics clearly has the potential to earn its place alongside other established omics fields such as genomics, proteomics and metabolomics. A wide variety of technologies are now being brought to bear on the technically difficult problems of glycan structural analysis and investigation of functional roles. It was apparent from the discussion that glycomics also encompasses opportunities to exploit genomics, proteomics and transcriptomics strategies applied specifically to glycan-related molecules (for example, glycan biosynthetic enzymes and glycan-binding proteins). Overall, the field of functional glycomics is now emerging as a vehicle to move carbohydrates into the mainstream of biology and biomedicine.

Analytical high-throughput technologies

One of the key bottlenecks in the development of glycobiology has been analysis of glycan structures, which are often only available in small quantities from natural sources. In the past, analytical tools have lagged far behind those used on DNA and proteins, but this landscape has changed considerably over the last decade. In particular, as in proteomics and metabolomics, mass spectrometry methods have come to the fore as a powerful tool for sensitive and definitive glycan analysis. Stuart Haslam (Imperial College, London) described the state of the art in MS applications for glycomics, using both electrospray and matrix-assisted laser desorption ionization instruments as examples. He demonstrated how MALDI-TOF can be exploited in screening strategies, thereby permitting very complex cell or tissue glycomes to be profiled and important minor structures to be identified on a timescale of a few weeks. Examples were given of MALDI mapping of mouse and human tissue-specific glycan profiles. These methods also provide a window on the dynamics of glycome expression². In contrast, ESI-MS/MS and MALDI-TOF/TOF methods can be used for detailed analysis of selected structures. Haslam also described how glycoproteomics approaches can be used to define the variant glycans located at specific attachment sites within glycoproteins³; analyzing these different 'glycoforms' is an important issue for proteomics because they can significantly alter protein function.

The significance of MS techniques for glycomics was reinforced by Rachel Martin (Shimadzu Biotech Ltd), who described the power of ion-trap MALDI-TOF for high-accuracy glycan analysis on femtomole quantities of sample. Sequential fragmentation by collision-induced dissociation permits MS to the power n to be performed on selected glycans and subsequent fragments, thereby yielding highly accurate analysis of glycan sequences. Johannes Stanta, from the Lowe group (University of Cambridge), outlined an MS approach using a new surface-enhanced laser desorption ionization (SELDI)-TOF chip for analyzing serum and tissue glycoproteins. By generically capturing glycans on the chip surface with boronic acid derivatives, this group is exploring the potential of this approach to identify glycoprotein disease markers (see below).

Emergence of glycoarrays

A major area of technology development has been the recent emergence of glycoarrays. Microarrays of glycans displayed on surfaces for high-throughput interrogation of interactions with cognate partners (generally proteins) are expected to be an important tool for advancing

the glycomics field. Ten Feizi (Imperial College, London) described "designer" arrays aimed at deciphering the glycode that are based on an established platform of glycolipids and neoglycolipids (glycans linked to a lipid tail to permit efficient clustered immobilization) arrayed onto nitrocellulose-coated glass slides⁴. Using a piezo-type spotting robot, researchers have arrayed a library of ~300 diverse sequence-defined glycans and investigated the binding of various glycan-binding proteins at femtomole levels of sensitivity using a slide scanner for fluorescence detection. Feizi demonstrated several examples of applications of the arrays to the discovery of ligands for glycan-binding proteins and to the determination of their fine specificities for glycan structures. Examples included dectin-1 (a receptor on leukocytes involved in innate immunity to fungal pathogens) and siglec family members (receptors on cells of the immune, hematopoietic and nervous systems)^{5,6}. The use of the arrays for exploring glycomes such as the O-linked glycans found in brain tissue was also demonstrated.

A diversified glycoarray platform based on oriented attachment of glycans via their reducing ends on self-assembled monolayers on gold surfaces was presented by Zheng-liang Zhi (University of Liverpool). Working with partners in the UK GlycoArray Consortium, which is funded by Research Councils UK (<http://www.glycochips.org.uk>), the Liverpool group has developed a fluorescence readout method that can be used to detect binding of protein partners for both heparin and N- and O-linked glycans⁷. The wider application of the platform for biosensor measurement of glycan-protein interactions was also demonstrated; this application is important because it will permit accurate measurements of the kinetics and affinities for specific interactions. This will add significant value to the information that can be gleaned from arrays, as would complementary information from functional assay screens. Zhi also reported that the potential of the platform for MALDI-MS investigation of glycan-protein interactions is being explored by the Consortium. Bastien Castagner (ETH Zurich) described another glycoarray approach in which fully synthetic heparin saccharides were immobilized via amine-terminated linkers placed at the reducing end of the synthetic structures to allow for immobilization onto *N*-hydroxysuccinimide-activated glass slides⁸. Kurt Drickamer (Imperial College London) described the effective exploitation of arrays in ELISA format for establishing the structural basis for distinct ligand-binding and targeting properties of protein receptors⁹ (see also <http://www.functionalglycomics.org>).

There is considerable interest in the continued development and application of glycoarray platforms such as these, because they will permit (for the first time) large-scale analysis of highly diverse libraries of glycan structures with multiple protein ligands. It is anticipated that they will make an important contribution to the glycomics field, along with complementary lectin array technologies¹⁰. Andrew Sutcliffe (Procognia Ltd) described a glycoanalysis platform based on a range of glycan-binding proteins (lectins) immobilized on chip surfaces and used to capture glycoproteins from biological samples, with detection using antibodies against specific proteins. This approach permits fingerprinting of the differential expression of glycoforms of selected proteins of interest, and it is of particular value in the optimization of bioprocessing, manufacturing and quality control of recombinant glycoproteins.

Chemical glycomics

The ability of synthetic and biosynthetic chemistry to produce fully defined glycan structures is proving to be a critical aspect in the development of glycomics, spinning off the related field of chemical glycomics¹¹. Synthetic chemistry provides routes not only to proof of specificity regarding the biological functions of glycans¹², but also to opportunities for 'glycoengineering', the generation of non-natural glycan-based structures with novel bioactivities.

Castagner outlined the power of automated solid-phase synthesis strategies for oligosaccharide preparation¹³ and highlighted the potential for microfluidics devices to revolutionize both research and production-scale synthesis. The impact of these and related technologies on biomedicine is already evident in the total synthesis of a *Bacillus anthracis* tetrasaccharide antigen for the creation of an anthrax vaccine candidate¹⁴ and in the use of synthetic glycosylphosphatidylinositols (GPIs) as candidate vaccines for malaria¹⁵.

In the context of carbohydrate chemistry and biochemistry, Rob Field (John Innes Centre) highlighted the issue of predicting gene function from sequence and gave examples that show that even highly homologous enzyme structures can give rise to different activities¹⁶. In contrast, the relatively relaxed substrate specificity of several sugar nucleotide-processing enzymes (for example, nucleotidyltransferases and glycosyltransferases) can be exploited for generating natural and unnatural glycans and new glycosylated antibiotics, without recourse to complex synthetic chemistry¹⁷.

The development of biocompatible methodologies for glycoprotein synthesis was addressed by Ben Davis (University of Oxford), with a view to site-specific engineering of defined

glycosylation variants at more than one position. In connection with the development of new P-selectin glycoprotein ligand-1 (PSGL-1) analogs, defined glycosylation was achieved using site-directed mutagenesis in combination with orthogonal disulfide exchange¹⁸ and Huisgen 1,3-dipolar cycloaddition chemistries. The power of synthetic chemistry to deliver multigram quantities of homogeneous N-linked glycan structures of commercial interest was also discussed.

Glycoinformatics

The growing scale of data generation from glycomics technologies underlines the critical importance of development of requisite bioinformatics platforms. These are clearly required to integrate the diverse datasets produced by the different technologies, with a goal of building a systems biology approach to glycan structure-function relationships¹. Haslam noted the difficulty of annotating complex MS spectra, which emphasizes the need for bioinformatics solutions¹⁹. Serge Perez (Centre de Recherches sur les Macromolécules Végétales (Cermav), Grenoble) outlined the important information content in the conformational space inhabited by glycans and described various glycoinformatics and modeling approaches for studying their molecular dynamics. Databases of the three-dimensional structures of ~100 bioactive glycans have been set up by this group (<http://www.cermav.cnrs.fr/glyco3d>). Perez also described how glycan arrays can be interrogated to discover the natural ligands, which can then be modeled based on predominant low-energy conformations in order to investigate potential mechanisms of interaction with protein partners.

Claus-Wilhelm von der Leith (German Cancer Research Center) focused on the need for relational glycoinformatic databases, emphasizing that information storage on glycans, for which topology of residues is critical, is more akin to that of small chemical molecules than to the sequence information of DNA and proteins. Advances in software are clearly required, along with common platforms to allow data sharing. Von der Leith described the range of existing databases (see www.eurocarbdb.org and links therein) and the recent moves toward acceptance of a Glycan Data Exchange (GLYDE) standardized data format. He also described the international plans to create a federated network of distributed databases for the glycosciences, with a hub at the European Bioinformatics Institute and other sites as nodes, and with potential for future expansion.

Bioinformatics were also exploited by Drickamer, who described the use of genomics

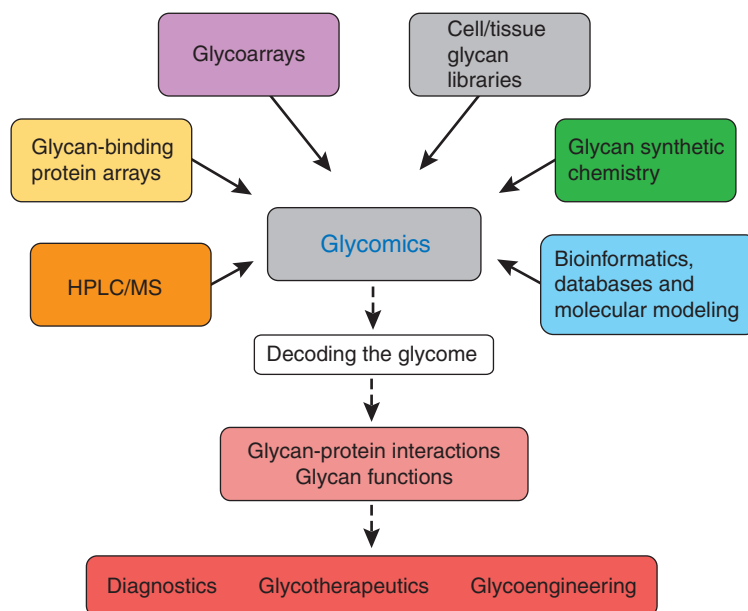


Figure 2 Cracking the glycode: emerging technologies underpinning glycomics. A range of technologies are now being exploited to undertake large-scale analyses of the structure-function relationships of the glycome. These approaches hold the potential to decode the glycome, leading to new insights and biomedical applications.

approaches to identify common lectin binding sites for glycan ligands by searching for common protein folds. Though this approach has been used successfully, he sounded a cautionary note regarding the difficulties in making predictions from genomics data concerning glycan-binding protein receptors. It is clear that despite the fact that multiple glycan ligands may bind to any one receptor, there are far too few receptors identified so far for the large numbers of glycans known to be expressed. This could mean that many different families of receptors, or other functional roles for glycans, have yet to be identified.

Biomedical applications

Throughout the meeting there was ample evidence of the potential impact of glycomics in the commercial arena, perhaps the best example being diagnostics. Several speakers highlighted the exciting opportunities for mining the glycome for disease biomarkers. Approaches based on a combination of HPLC and MS analysis for the identification of signature glycan structures and glycosylation patterns in a variety of cancer types²⁰ were demonstrated by Pauline Rudd (National Institute for Bioprocessing Research and Training Dublin-Oxford Glycobiology Laboratory). The HPLC technology has been automated in a 96-well plate format and, combined with computer-assisted data interpretation using extensive databases, it has the

potential for high-throughput discovery of new glycodiagnostics and monitoring of recombinant therapeutics. In addition, Stanta described the application of a glycan-capturing SELDI-TOF chip that allows identification of serum glycoprotein biomarkers of diseases such as schizophrenia. Haslam also highlighted the potential for MALDI-MS mapping to provide glycan profiles from tissues, which could potentially be used to diagnose diseases such as liver sclerosis and cancer. It was evident from these examples that glycan biomarkers are an exciting field with huge potential, and significant developments in this area can be anticipated in the next few years.

Analytical and array technologies also have clear potential for application in the production of recombinant protein therapeutics, though hurdles still exist. Daryl Fernandes (Ludger Ltd) described the regulatory landscape for biopharmaceutical glycosylation, reviewing the issues that variant glycoforms of proteins impose regarding stability, safety and efficacy. The pressing need for improved international guidelines on appropriate characterization of glycoforms and glycan profiles of glycoprotein therapeutics was explained. Future moves in this direction are likely to be driven by the use of technologies proven to be sufficiently accurate and robust to be chosen for standardization. Indeed, evaluation of methods with this in mind was described by Paula Vickers (LGC Ltd),

who is involved in the Measurements for Biotechnology program (funded by the UK Department of Trade and Industry as part of the National Measurement System), in which different methods of interest to pharmaceutical and biotechnology companies have been studied. As an example, she described validation of methods coupling capillary electrophoresis or LC with MS for improved discrimination of glycoprotein microheterogeneity. The ultimate goal of the program is to develop a best practice guide and a panel of standards linked to the regulatory guidelines (www.mfbprog.org.uk).

Clearly there are strong drivers for the development of improved approaches for production of recombinant glycoproteins. As Sutcliffe pointed out, there are a number of biogenics coming off patent in the next five years (a market worth US \$15 billion), so glycomics approaches could underpin opportunities for technology development for monitoring and control of glycosylation of recombinant protein therapeutics. In this regard, the production of highly pure single glycoforms of recombinant proteins through glycoengineering, as reported by Davis, also has major commercial applications. Fully defined recombinant therapeutics could herald a step-change in the usage potential of these molecules, providing routes to

better-targeted, more bioactive drugs with reduced off-target side effects.

Summary

In the emerging field of glycomics, a diverse range of technologies and strategies are coming into play to crack the glyco-code (Fig. 2). Arrays, HPLC, MS, natural glycan libraries and synthetic chemistry are at the forefront of technology developments and are supported by progress in the development of essential tools for data handling, bioinformatics and molecular modeling of glycans. The application of existing genomics and transcriptomics platforms is also proving useful in studies on the biosynthesis of the glycome and in identification of glycan-binding proteins. Many exciting applications of glycomics approaches have become evident, including diagnostics, new routes to glycotherapeutics and defined recombinant protein drugs. We envisage that the collective enterprise of glycomics over the next decade will begin the process of decoding the glycome, thereby yielding many new insights into its myriad functions and producing diverse advances in the biomedical arena.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

1. Raman, R., Raguram, S., Venkataraman, G., Paulson, J.C. & Sasisekharan, R. *Nat. Methods* **2**, 817–824 (2005).
2. Comelli, E.M. *et al. J. Immunol.* **177**, 2431–2440 (2006).
3. Chalabi, S. *et al. Biochemistry* **45**, 637–647 (2006).
4. Feizi, T. & Chai, W.G. *Nat. Rev. Mol. Cell Biol.* **5**, 582–588 (2004).
5. Palma, A.S. *et al. J. Biol. Chem.* **281**, 5771–5779 (2006).
6. Campanero-Rhodes, M.A. *et al. Biochem. Biophys. Res. Commun.* **344**, 1141–1146 (2006).
7. Zhi, Z.L., Powell, A.K. & Turnbull, J.E. *Anal. Chem.* **78**, 4786–4793 (2006).
8. de Paz, J.L., Noti, C. & Seeberger, P.H. *J. Am. Chem. Soc.* **128**, 2766–2767 (2006).
9. Guo, Y. *et al. Nat. Struct. Mol. Biol.* **11**, 591–598 (2004).
10. Hsu, K.L., Pilobello, K.T. & Mahal, L.K. *Nat. Chem. Biol.* **2**, 153–157 (2006).
11. Turnbull, J.E. & Linhardt, R.J. *Nat. Chem. Biol.* **2**, 449–450 (2006).
12. Gama, C.I. *et al. Nat. Chem. Biol.* **2**, 467–473 (2006).
13. Seeberger, P.H. & Werz, D.B. *Nat. Rev. Drug Discov.* **4**, 751–763 (2005).
14. Werz, D.B. & Seeberger, P.H. *Angew. Chem. Int. Edn.* **44**, 6315–6318 (2005).
15. Schofield, L., Hewitt, M.C., Evans, K., Siomos, M.A. & Seeberger, P.H. *Nature* **418**, 785–789 (2002).
16. Tello, M. *et al. Chem. Commun. (Camb)*. **2006**, 1079–1081 (2006).
17. Yang, M. *et al. J. Am. Chem. Soc.* **127**, 9336–9337 (2005).
18. Bernardes, G.J.L., Gamblin, D.P. & Davis, B.G. *Angew. Chem. Int. Edn.* **45**, 4007–4011 (2006).
19. Goldberg, D., Sutton-Smith, M., Paulson, J.C. & Dell, A. *Proteomics* **5**, 865–875 (2005).
20. Tabares, G. *et al. Glycobiology* **16**, 132–145 (2006).