Linking Glycomics Repositories with Data Capture

Matthew P. Campbell[1], Frédérique Lisacek[2], Marc R. Wilkins[3], Pauline M. Rudd[4], Daniel Kolarich[6], Catherine A. Hayes[5], Niclas G. Karlsson[5] and Nicolle H. Packer[1]*

[1]Biomolecular Frontiers Research Centre, Macquarie University, Sydney, Australia

[2]Proteome Informatics Group, Swiss Institute of Bioinformatics, Switzerland

[3]Systems Biology Initiative, University of New South Wales, Sydney, Australia

[4]NIBRT Dublin-Oxford Glycobiology Lab, University College Dublin, Ireland

[5]Medical Biochemistry, University of Gothenburg, Sweden

[6]Max Planck Institute of Colloids and Interfaces, Department of Biomolecular Systems, Berlin, Germany

E-Mail: *nicki.packer@mq.edu.au

Abstract

As in proteomics, the speed of the advances in glycomic discovery is dependent upon the development of a specific bioinformatic knowledgebase that links various forms of glycoanalytical data to repositories of known glycan structures. The UniCarbKB project (www.unicarbkb.org) is a partnership of leading international research groups come together in an effort to develop and provide an informatic framework for the storage of high-quality structural glycan collections including informative meta-data and annotated experimental datasets. UniCarbKB seeks to advance the integration of data capture and management within the glycomics discipline.

INTRODUCTION

About 25 years ago glycobiology began to emerge as a field focused on the functional aspects of glycans. The term Glycobiology, coined by Professor Raymond Dwek FRS (Director of the Oxford Glycobiology Institute), entered the Oxford English Dictionary in 1988, "to emphasize the importance of oligosaccharides (sugars, carbohydrates) in their biological context in the hope that this would reveal their functions". Glycans, both in form of polysaccharides or glyco-conjugates (bound to proteins and lipids), are the most abundant class of biomolecules and are increasingly being implicated in human health [1 – 5].

Glycosylation is by far the most important post-translational modification in terms of the number of proteins modified and the diversity generated. This elaborate modification enhances the functional diversity of the relatively limited number of proteins translated from the genome. Since glycoproteins, glycolipids and glycan-binding proteins (also referred to as lectins, which recognise particular glycan epitopes) are mainly located on the cell's surface, they present a primary interface for a myriad of biologically significant events that can be attributed to glycan recognition. These include interactions between mammalian eggs and sperm [6], between pathogens/parasites and hosts to establish the infection process [7, 8], and as part of the immune system [9, 10].

Glycobiology is intertwined with fields such as immunology, virology, reproductive biology and drug discovery, making significant discoveries in the fields of molecular biology and in the fight against disease. Glycobiologists are exploring the many roles of sugars, discovering that different structures have essential functions in all major organ systems and are involved in the progression of many diseases [11]. These findings have led to the development of carbohydrate-based drugs and vaccines, some of which show great promise for treating or preventing various diseases, including influenza, malaria, cancer and AIDS [12 – 15]. Not surprisingly, considering the central role they play in molecular encounters, carbohydrate-based drugs and therapeutics represent a greater than US $ 20 billion market, including approvals of the first (preventive and therapeutic) cancer vaccines. Other examples include the glycoprotein, erythropoietin, that is used in treating kidney disease (as well as in sports drug abuse) and which is the highest biosimilar drug revenue earner in the pharmaceutical industry; and Relenza and Tamiflu that are inhibitors of the infective viral-sugar interaction and which have been used in avian and swine flu treatment.

The rapidly expanding glycoscience field is being increasingly recognised as an important component of life science research, as highlighted in a report commissioned by the European Science Foundation, which stated in its policy briefing that "to further develop diagnostic tools, preventive medicines (vaccines) and therapeutic drugs, a better understanding of glycosylated molecules is required" (Structural *Medicine – The Importance of Glycomics for Health and Disease,* ESF Science Policy Briefing 27; 2006). The growing importance for glycobiology in biomedical and disease research has also recently been recognised by a

major 7-year (US $ 18 million) grant awarded by the US National Heart, Lung and Blood Institute (NHLBI) under its new "Program of Excellence in Glycosciences" and efforts in Japan to bring together glyco-databases through the Government sponsored MEXT Integrated Database Project. These programmes aim to translate emerging discoveries in the glycosciences to new discoveries and therapies.

An increasing number of researchers accept this importance of sugars in basic biology, however, many are concerned that the barrier for entry into this research field remains too high, potentially hampering new high-impact discovery and innovative research efforts. New technologies and systematic glycomics (the system-wide study of the entire complement of sugars of an organism) investigations churn out lists of complex data on a routine basis and the amount of information available has expanded beyond comprehension [16, 17]. Analogous to the essential function and growth of informatics in genomics and proteomics, the need to develop appropriate databases and bioinformatics tools for glycomics is obvious. There is a demand for new methods to index and retrieve information rich datasets, and software to annotate and integrate the data. Infrastructure capabilities are now required to disseminate research collections, to allow researchers to know what information is out there, as well as analytical tools to handle and interpret the volume and complexity of experimental data being produced by a myriad of different technologies.

GLYCOMICS AND BIOINFORMATICS: DISCONNECTED ISLANDS

Public availability of quality curated data collections and informatics tools with which to interpret the flood of associated data has been of paramount importance in most life sciences and has become one of the foundations of modern biology. The advances in genomics (NCBI) and proteomics (UniProt) have clearly demonstrated how the existence of well-curated extensive collections of biomolecular data, together with the development of software to facilitate the mining and correlation of this data, is crucial to research advancements. In contrast to these other areas, glycomics lacks accessible, curated and comprehensive data collections and informatics tools that summarise and integrate the experimentally determined structure, interactions, characteristics and biological origin and link them with the potential function of glycans.

This lack of glycan knowledge bases and the sparseness of resources hampers the realisation of bioinformatics tools for the interpretation of experimental data and the automated determination of glycan structure; thereby limiting the possibility of large-scale glycomics studies. The complexity of the glycan structures, and the variety of techniques that are used for their study, requires a more sophisticated approach to the development of computerised tools that will have the same impact on glycomics as, for example, the Mascot protein identification tool has had on proteomics discovery.

A small international community of researchers has emerged in the last decade with the aim of building the foundations for informatics solutions in the realm of glycobiology and glycomics. A variety of databases and tools have developed in recent years to acquire, annotate and organise data collections, managed by various consortia, institutions or local groups [18 – 25]. The consortia, based in US (Consortium for Functional Glycomics), Europe (EuroCarbDB) and Japan (KEGG Glycan) are currently undergoing funding difficulties. The existence of these disconnected and incompatible islands of experimental data, resources and proprietary applications have been identified by glycobiologists as the single biggest hindrance to their research [26].

There is a general consensus within the community that the availability of comprehensive and up-to-date eResearch resources comprising knowledge bases supported by efficient software is a prerequisite for successfully conducting large-scale glycomics projects and interfacing them with related genomics and proteomics activities.

The Establishment of UniCarbKB to Build a Universal Knowledgebase for Glycoscience Research

Glyco-bioinformatics is slowly expanding, and benefits from the developments made in the genomics and proteomics fields, but it is in its infancy and the available informatics resources retain focus on separate methodologies and/or functionalities. UniCarbKB [27] aims to promote the translation of research by building an effective infrastructure that enables glyco-related research (Figure 1).
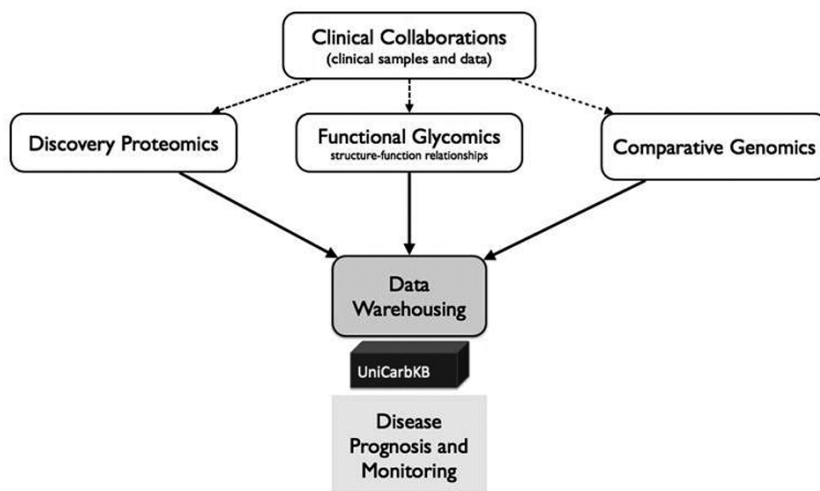

Figure 1. The complexity of glycomics and its relationship to genomics and proteomics, all of which require the integration of robust databases and tools to interpret the implications of glycosylation.

Research in the life sciences can only flourish if data is preserved and made accessible to all facets of the community to promote scientific progress. This is exemplified by the biological strides and technology advancements over the last decade that has provided the 'parts lists' for many living organisms.

The absence of a single comprehensive resource that collects research findings reported in peer-reviewed publications, along with supporting experimental data in an accessible and searchable form, acquired from a wide variety of analytical technologies, is a limiting factor in efforts to progress the field [28]. It has been readily acknowledged that this type of resource has been essential to the progress of genomics and proteomics.

The challenge for glycobiology and glycomics is to address the problem of data distribution and the fragmentation of knowledge. This can be achieved by integrating the many sources of information into a single and systematic entity, to be called UniCarbKB, and to continue efforts to advance the interpretation of captured data through the development of novel data analysis methods and algorithms for the efficient representation and mining of large experimental data sets.

The proposed international initiative will build on our previous efforts and those of our international collaborators to construct UniCarbKB [27] (Figure 2), utilising multidisciplinary/inter-sectorial integration of data to support glycoscience research and its translation to medicine and the biotechnology industry.
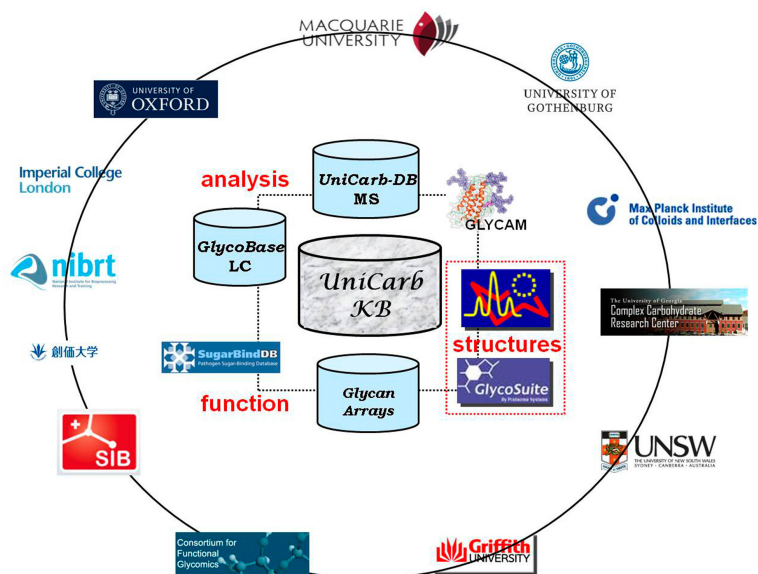
Figure 2. Overview of UniCarbKB: the optimistic endpoint.

The establishment of UniCarbKB will support:

- Provision of an open access infrastructure for glycobiological data from common analytical strategies,

- Working with international partners to enable the standardised collection, integration, and storage to promote the sharing of user-generated data,

- Providing comprehensive integrated data sets, with a web-based publically accessible portal, enabling user entered data collection,

- Curating and collecting published glycoanalytical data,

- Negotiating with editors of discipline journals to ensure glycobiological data is placed in the publically accessible database,

- Enhancing data quality through annotation features and cross-referencing with internationally recognised databases used for genomics and proteomics,

- Designing software tools to enable sharing between institutions and distributed databases for glycosciences,

- Supporting the development of algorithms to enable rapid and reliable automatic interpretation and mining of mass spectrometry spectra, chromatographic and electrophoretic behaviour, molecular modelling and array data.

UniCarbKB introduces three new resources that leverage existing infrastructure. Firstly, the introduction of MIRAGE, a standard focused on the analytical technologies and data interpretation issues associated with glycomics analysis. This standard will describe the information required to reproduce and process data that was used to characterize a group of structures or an individual structure. Secondly, the development of a data-exchange or remote communication protocol to enable databases, tools and informatics developers to access data collections using REST-style protocols, this will support the integration of 'omics' databases similar to resources distributed by UniProt [29]. Finally, the development of new algorithms and tools to facilitate the rapid and reliable processing, interpretation and annotation of analytical MS data for glycan structure determination with a high degree of automation by i) library based matching and scoring based on experimental MS spectra data sets (UniCarb-DB [30]) and ii) improved *in-silico* fragmentation based on features observed from library generation.

APPROACH

To help achieve this ambitious goal we have focused the work into a series of milestone packages and deliverables that cover software development to address technical challenges such as data storage and representation, and the interpretation of complex and large datasets.

Implementation of a federated and unified web portal to access disparate sources

A publically accessible portal will be deployed to support the integration of existing global repositories, GlycoSuiteDB [31], EUROCarbDB [19], GlycoBase [32] and UniCarb-DB [30],(Table 1) which will provide access to a large manually curated collection of glycan structures extracted from the literature together with associated experimental data (UPLC/HPLC, Mass Spectrometry).

Table 1: Currently available glycan structural databases

Carbohydrate Database	Web address	Number of unique structures	Content	Status in 2011
Bacterial Carbohydrate Structure Database (BCSDB)	http://www.glyco.ac.ru/bcsdb3/	9,071	bacterial glycans	Last V.3 update Sep 22, 2011
EurocarbDB	http://www.ebi.ac.uk/eurocarb/home.action	13,471	glycoconjugate glycans	Last user entry Sep 7 2010
GlycoSuiteDB	http://glycosuitedb.expasy.org/glycosuite/glycodb	3,238	glycoprotein glycans	manually curated entry from literature V.8 up to 2005
Consortium of Functional Glycomics (CFG GlycanDB)	http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp	7,500	synthetic and glycoconjugate glycans	Last update 2010
Kyoto Encyclopedia of Genes and Genomes (KEGG Glycan database)	http://www.genome.jp/kegg/glycan/	10,979	glycoconjugate glycans	Last update Dec 11, 2007

Based on the specifications and the descriptions of each platform used to characterise glycan structures and their occurrence in nature and biological functions, UniCarbKB will modify the database architecture, as well as user interfaces and retrieval options, to ensure that the quality of data available and functions fit the needs of research. The modality by which the users will access the services will be through a common portal that will be developed and deployed as a component on a public web site of UniCarbKB. The portal concept hides the complexity of the underlying data architecture, and the makeup of multiple data resources. Likely expansion is the addition of 3D-structures determined by molecular modelling, leveraging the Complex Carbohydrate Research Centre expertise and high-performance computing nodes accessible through GLYCAM (http://glycam.ccrc.uga.edu) and glycan interaction array data from the US Consortium of Functional Glycomics (http://www.functionalglycomics.org/glycomics/publicdata/primaryscreen.jsp) and the Australian Glycomics Institute, Griffith University.

Building and deploying structure curation and data management workflows

One of the major goals for UniCarbKB is to develop and provide a robust and easy-to-use application which would allow researchers to feed their structural and experimental data directly into the database. The technical design of the database and the evaluation of minimum reporting requirements for storing structure and meta-data/literature descriptions will lead to enhanced workflows, interfaces and an improved tool environment to manage, assign and deposit data, thus moving towards a collaborative research resource for glycomics.

The availability of such tools will lead to the creation of one single, highly curated and centralised knowledgebase to manage published carbohydrate structure information, thus removing the key stumbling block in glycobiology where information is dispersed across multiple sites.

Semi-automated workflows to capture mass spectrometry data

Mass Spectrometry (MS) is a highly sensitive method, which is frequently applied to provide information on the carbohydrate repertoire of cells, tissue and organs. MS/MS-spectra coupled with databases of glycan structures will enable the automated assignment of the carbohydrate moieties of glycoproteins. However, no comprehensive depository of experimentally determined MS spectra is available. Moreover, no standardised procedures and agreements exist on quality measures that have to be fulfilled before entering into a database.
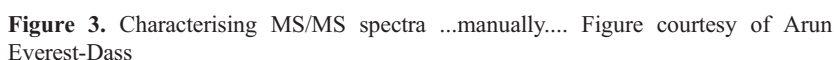
UniCarbKB is orientated to demonstrate how the data flow of experimental data from spectrometry to the database can be done using automated procedures. It is also our aim to extend UniCarb-DB's collection [30] of MS/MS spectra by providing workflows and connecting tools used to capture and process MS data. The existence of an accepted repository for glyco-related spectroscopic data will guarantee that the loss of primary data will be considerably reduced.

Automating assignment and interpretation of mass spectrometry data

We aim to support the improvement of existing algorithms, which will enable a rapid and reliable automatic interpretation and detection of carbohydrates from the MS/MS data instead of the highly labour intensive manual annotation currently being used (Figure 3). The existence of sufficiently large high quality collections of reference spectra in the Uni-Carb-DB component of UniCarbKB will allow us to extend algorithms and improve capacity by incorporating library and *in silico* approaches – an urgent requirement in high throughput investigations.

Figure 3. Characterising MS/MS spectra ...manually.... Figure courtesy of Arun Everest-Dass

Establishing and communicating exchange standards by deploying web service solutions

A RESTful web service application will provide a novel informatics remote interface to enable communication between the databases, building on the success of the international Working Group on Glycomics Data(base) Standards and the MIRAGE standards, to implement a common data exchange standard between remote structural and experimental databases created from a myriad of different technologies. It will provide enhanced search functionality, through the provision of a unique architecture that connects several international databases, whilst providing a structure for connecting future remote Research Cloud and data storage nodes.

Establish the interlink between the UniCarbKB glycan reference database and the protein databases

There have been few intensive programmes to cross-reference sugar-containing databases with those available in the genomics and protein sciences in a systems biology context. UniCarbKB will develop cross-linking strategies to establish relationships with the home of the proteomics databases, Swiss Institute of Bioinformatics (SIB) by making use of existing remote API packages, whilst opening the web services (above) to establish a new data communication pipeline.

OUTCOMES

Contributing, retrieving and data-mining a curated resource

These approaches will provide efficient functionalities for i) contributing structural and meta-data ii) submitting high-quality analytical data and iii) browsing and searching within the various component databases described above. The user will be able to search for all glycans associated with: (a) a particular taxonomic branch or individual species, (b) a particular organ, tissue or cell type, (c) a particular disease or pathology, (d) a chemical perturbation or drug treatment.

Delivering new bioinformatics to promote discovery

The informatics technologies and libraries will be able to be applied to screen data collections in a high-throughput manner. The bioinformatics technologies will address limitations and inaccuracies and provide a forum for comparing in silico approaches with traditional labour intensive interpretations. Access to a screening tool will assist routine data analysis, provide mechanisms to enhance the translation of discovery research, and improve capacity.

The construction and availability of the UniCarbKB resource will thus provide freely available data and bioinformatics tools to promote glycoscience and discovery research and contribute to the advancement of biology through basic investigator-driven research in bioinformatics.

CONCLUSIONS

Fifty years after the double helix, the reference DNA sequence of *Homo sapiens* is now available for downloading with many tools from one site centred in the USA. The National Centre for Biotechnology Information (NCBI) thus advances science and health by providing central access to genomic information. About twenty years after the start of proteomics a curated protein database and informatics tools are available at one site centred in Europe (UniProt). The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information. It may be somewhat ambitious, but we would like the same to be said one day for glycobiology, viz. that a gold standard curated database and informatics tools for glycomics are available.

As the president of Hewlett-Packard said in 1999 "The goal is to transform data into information, and information into insight". To do this in glycobiology requires innovative approaches, a well-organised comprehensive data repository and unique software.

UniCarbKB will offer the glycoscience community new enhanced tools and capabilities to search this growing mass of information, to share user-generated content, and to provide robust solutions to assist data annotation.

UniCarbKB will be an innovative glycomics knowledge platform dedicated to bridging the gap between data-capture and data interpretation to help glycobiological researchers understand and direct investigations. The creation of the knowledgebase will set the foundations for continued development to enhance the global recognition and importance of glycobiology. The key stumbling blocks impeding progress will be removed by 1) providing a freely accessible, highly curated, amalgamated database of known glycan structures and mass spectrometric experimental evidence; 2) merging curated data collections into a centralised database, that comprise manually annotated structural and experimental verified data; 3) extending and developing new databases containing information on different properties of the carbohydrates attached to proteins and lipids; 4) providing a technical framework to deploy tools and workflows to assist the interpretation of experimental data 5) hosting of the site on the established international proteomics portal, UniProt/SwissProt, to try to ensure sustainability of access.

The successful construction of UniCarbKB will also provide a leading edge and innovative approach to connect glycomics with Systems Biology, an emerging and central discipline essential to all areas of biology and medicine. A key component will be integration of glycomics data with genomic/proteomic findings in biomedical studies where glycan structures can serve as biomarkers for specific disease or malfunctions. Links will be fostered between glyco-bioinformatics and systems informatics groups that employ similar methods for genomics, proteomics, metabolomics and lipidomics research.

The UniCarbKB data acquisition consortium, at present, comprises glycoscientists at Macquarie University and University of New South Wales, Australia; University of Gothenburg, Sweden; the Swiss Institute for Bioinformatics, Switzerland; the Max Planck Institute, Germany; and the National Institute for Bioprocessing Research and Training (NIBRT), Ireland. In the future it is hoped that the partnership will expand to include quality data on all other aspects of glycomics information. Several groups (Figure 2) have already indicated their interest in being an active part of this initiative.

UniCarbKB thus is an international initiative that aims to collect and distribute resources and practices from glycobiology practitioners to the whole biological research community. The mission is to provide a comprehensive, high quality catalogue of information on carbohydrates, and to continue efforts to advance the interpretation of captured data through the development of novel data analysis methods and algorithms for the efficient representation and mining of large experimental glycomics data sets.

REFERENCES

[1] Hart, G.W. & Copeland, R.J. (2010) Glycomics hits the big time. *Cell* **143:**672 – 676.
doi: http://dx.doi.org/10.1016/j.cell.2010.11.008.

[2] Dwek, R.A. (1996) Glycobiology: Toward Understanding the Function of Sugars.
Chem. Rev. **96:**683 – 720.
doi: http://dx.doi.org/10.1021/cr940283b.

[3] Bonomelli, C. *et al.* (2011) The Glycan Shield of HIV Is Predominantly Oligo-
mannose Independently of Production System or Viral Clade. *PLoS One* **6:**e23521.
doi: http://dx.doi.org/10.1371/journal.pone.0023521.

[4] Ohtsubo, K. & Marth, J.D. (2006) Glycosylation in cellular mechanisms of health
and disease. *Cell* **126:**855 – 867.
doi: http://dx.doi.org/10.1016/j.cell.2006.08.019.

[5] Schachter, H. & Freeze, H.H. (2009) Glycosylation diseases: quo vadis? *Biochim.
Biophys. Acta* **1792:**925 – 930.
doi: http://dx.doi.org/10.1016/j.bbadis.2008.11.002.

[6] Pang, P.C. *et al.* (2011) Human sperm binding is mediated by the sialyl-Lewis(x)
oligosaccharide on the zona pellucida. *Science* **333:**1761 – 1764.
doi: http://dx.doi.org/10.1126/science.1207438.

[7] Issa, S. *et al.* (2010) O-linked oligosaccharides from salivary agglutinin: *Helicobacter
pylori* binding sialyl-Lewis x and Lewis b are terminating moieties on hyperfuco-
sylated oligo-*N*-acetyllactosamine. *Glycobiology* **20:**1046 – 1057.
doi: http://dx.doi.org/10.1093/glycob/cwq066.

[8] Moran, A.P., Gupta, A. & Joshi, L. (2011) Sweet-talk: role of host glycosylation in
bacterial pathogenesis of the gastrointestinal tract. *Gut* **60:**1412 – 1425.
doi: http://dx.doi.org/10.1136/gut.2010.212704.

[9] Rudd, P.M., Elliott, T., Cresswell, P., Wilson, I.A. & Dwek, R.A. (2001) Glyco-
sylation and the immune system. *Science* **291:**2370 – 2376 (2001).
doi: http://dx.doi.org/10.1126/science.291.5512.2370.

[10] Arnold, J.N., Wormald, M.R., Sim, R.B., Rudd, P.M. & Dwek, R.A. (2007) The
impact of glycosylation on the biological function and structure of human immuno-
globulins. *Annu. Rev. Immunol.* **25:**21 – 50 (2007).
doi: http://dx.doi.org/10.1146/annurev.immunol.25.022106.141702

[11] Gabius, H.J. (2011) Glycobiomarkers by glycoproteomics and glycan profiling
(glycomics): emergence of functionality. *Biochem. Soc. Trans.* **39:**399 – 405.
doi: http://dx.doi.org/10.1042/BST0390399.

[12]   Astronomo, R.D. & Burton, D.R. (2010) Carbohydrate vaccines: developing sweet solutions to sticky situations? *Nat. Rev. Drug Discov.* **9:**308 – 324.
       doi: http://dx.doi.org/10.1038/nrd3012.

[13]   von Itzstein, M. (2007) The war against influenza: discovery and development of sialidase inhibitors. *Nat. Rev. Drug Discov.* **6:**967 – 974.
       doi: http://dx.doi.org/10.1038/nrd2400

[14]   Fuster, M.M. & Esko, J.D. (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat. Rev. Cancer* **5:**526 – 542.
       doi: http://dx.doi.org/10.1038/nrc1649.

[15]   Calarese, D.A. *et al.* (2003) Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300:**2065 – 2071.
       doi: http://dx.doi.org/10.1126/science.1083182.

[16]   Pabst, M. & Altmann, F. (2011) Glycan analysis by modern instrumental methods. *Proteomics* **11:**631 – 643.
       doi: http://dx.doi.org/10.1002/pmic.201000517.

[17]   Geyer, H. & Geyer, R. (2006) Strategies for analysis of glycoprotein glycosylation. *Biochim. Biophys. Acta* **1764:**1853 – 1869.
       doi: http://dx.doi.org/10.1016/j.bbapap.2006.10.007.

[18]   Frank, M. & Schloissnig, S. (2010) Bioinformatics and molecular modelling in glycobiology. *Cell Mol. Life Sci.* **67:**2749 – 2772.
       doi: http://dx.doi.org/10.1007/s00018-010-0352-4.

[19]   von der Lieth, C.W. *et al.* (2011) EUROCarbDB: An open-access platform for glycoinformatics. *Glycobiology* **21:**493 – 502.
       doi: http://dx.doi.org/10.1093/glycob/cwq188.

[20]   Lutteke, T. *et al.* (2006) GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiology research. *Glycobiology* **16:**71R-81R.
       doi: http://dx.doi.org/10.1093/glycob/cwj049.

[21]   Aoki-Kinoshita, K.F. (2008) An introduction to bioinformatics for glycomics research. *PLoS Comput. Biol.* **4:**e1000075.
       doi: http://dx.doi.org/10.1371/journal.pcbi.1000075.

[22]   Raman, R. *et al.* (2006) Advancing glycomics: implementation strategies at the consortium for functional glycomics. *Glycobiology* **16:**82R-90R.
       doi: http://dx.doi.org/10.1093/glycob/cwj080.

[23]   Hashimoto, K. *et al.* (2006) KEGG as a glycome informatics resource. *Glycobiology* **16:**63R-70R.
       doi: http://dx.doi.org/10.1093/glycob/cwj010.

[24] Ranzinger, R., Herget, S., Wetter, T. & von der Lieth, C.W. (2008) GlycomeDB –
      integration of open-access carbohydrate structure databases. *BMC Bioinformatics*
      9:384.
      doi: http://dx.doi.org/10.1186/1471-2105-9-384.

[25] Toukach, P., Joshi, H.J., Ranzinger, R., Knirel, Y. & von der Lieth, C.W. (2007) Sharing
      of worldwide distributed carbohydrate-related digital resources: online connection of
      the Bacterial Carbohydrate Structure DataBase and GLYCOSCIENCES.de. *Nucleic
      Acids Res.* **35:**D 280 – 286.
      doi: http://dx.doi.org/10.1093/nar/gkl883.

[26] Lutteke, T. (2008) Web resources for the glycoscientist. *Chembiochem* **9:**2155 –
      2160.
      doi: http://dx.doi.org/10.1002/cbic.200800338.

[27] Campbell, M.P. *et al.* (2011) UniCarbKB: Putting the pieces together for glycomics
      research. *Proteomics* **11:**4117 – 4121.
      doi: http://dx.doi.org/10.1002/pmic.201100302.

[28] Packer, N.H. *et al.* (2008) Frontiers in glycomics: bioinformatics and biomarkers in
      disease. An NIH white paper prepared from discussions by the focus groups at a
      workshop on the NIH campus, Bethesda MD (September 11 – 13, 2006). *Proteomics*
      8:8 – 20.
      doi: http://dx.doi.org/10.1002/pmic.200700917.

[29] Magrane, M. & Consortium, U. (2011) UniProt Knowledgebase: a hub of integrated
      protein data. *Database: the journal of biological databases and curation* 2011,
      bar009.

[30] Hayes, C.A. *et al.* (2011) UniCarb-DB: a database resource for glycomic discovery.
      Bioinformatics **27:**1343 – 1344.
      doi: http://dx.doi.org/10.1093/bioinformatics/btr137.

[31] Cooper, C.A., Harrison, M.J., Wilkins, M.R. & Packer, N.H. (2001) GlycoSuiteDB: a
      new curated relational database of glycoprotein glycan structures and their biological
      sources. *Nucleic Acids Res.* **29:**332 – 335.
      doi: http://dx.doi.org/10.1093/nar/29.1.332.

[32] Campbell, M.P., Royle, L., Radcliffe, C.M., Dwek, R.A. & Rudd, P.M. (2008)
      GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics*
      24:1214 – 1216.
      doi: http://dx.doi.org/10.1093/bioinformatics/btn090.