Beilstein-Institut

Cracking the Sugar Code by Navigating the Glycospace
June 27th – July 1st, 2011, Potsdam, Germany

Expanding Knowledge on the Biological Context of Glycan Structures

Frédérique Lisacek[1]*, Matthew P. Campbell[2] and Nicolle H. Packer[2]

[1]Proteome Informatics Group, Swiss Institute of Bioinformatics, Geneva, Switzerland

[2]Biomolecular Frontiers Research Centre, Faculty of Science, Macquarie University, Sydney, Australia

E-Mail: *frederique.lisacek@isb-sib.ch

Received: 20th December 2011/Published: 11th July 2012

Abstract

Advance in bioinformatics has been substantially influenced by the management of biological data flows. In the most common –omics domains, such as genomics and proteomics, data quality and integration have raised questions and have solutions that can partially address the similar current predicament of glycomics data flows. Some lessons learnt from these can benefit the study of glycans at a time when high throughput data production is well underway.

Introduction

In the last thirty years, knowledge in many fields of the Life Sciences has significantly progressed with the introduction of computer resources. In fact, bioinformatics support has moved forward each time a significant data overflow has hit a domain of biology. The boom in gene sequencing in the early 1980 s prompted the creation of the first sequence databases (e. g. GenBank, [1]), shortly followed by efforts to account for the concurrent increase in gene products (e. g., Swiss-Prot, [2]). Likewise, when the list of complete genomes started to grow, specialised genome databases emerged (e. g., PEDANT, [3]). Software was improved in parallel with meeting the need for aligning, searching and comparing gene sequences, which were the central object of study and hence the major type of initial data.

The evolution of proteome bioinformatics matches the same growth pattern with a slight timeframe shift. The refinement of mass spectrometers in the late 1990s data opened the gates of another data flood complementing the genomics sequence data. The high throughput production of mass spectra led to the development of corresponding databases such as the PeptideAtlas [4], along with a range of analytical informatics tools for processing mass spectra and correlating them with protein sequence information.

The recent improvement in analytical techniques used in glycobiology has now led to a rise for the need for glyco-bioinformatics. Although sugar structures are unique entities and as such require the development of specific bioinformatics resources, the extensive use of mass spectrometry in solving glycome structures bridges with proteome bioinformatics. Previous experience in designing software and databases can be built on to benefit glycobiology applications.

To begin with, lessons in data integration of genomics and proteomics information can help in creating a single virtual space in which knowledge of the structure and function of glycosylation would be organised for the support of integrated data interpretation. The integration of proteomics knowledge into UniProtKB (http://www.uniprot.org/) that followed the trend set by Swiss-Prot [2] is a definite source of inspiration for exploring and elucidating relationships between the structure(s) and function(s) of biological entities including glycans. The steadily increasing intensive use of UniProtKB unambiguously reflects a need for such types of resources.

Secondly, experience in setting standards for data storage and exchange that will facilitate subsequent data comparisons is crucial. Biological data tends to be at first generated somewhat anarchically as a consequence of multiple, and in-the-making, technologies. The stabilisation of techniques, confirmed by their reproducibility, opens the possibility of a unified representation of the data from different experimental techniques. Again, glycomics has recently reached that stage and standards can now be implemented and enforced.

Finally, no major bioinformatics progress was ever made without grounding and testing new tools in real applications. Genomics first focused on model organisms to fine-tune browsers and analytical tools. Protein-protein interactions studies that heavily rely on bioinformatics are benchmarked with yeast data. Such examples are numerous.

This paper highlights these three lessons learnt from genomic and proteomics informatic developments and investigates their benefit to glycomics as an introduction to current efforts described in other articles in this volume; the collection of data on the glycosylation of proteins (M.P. Campbell *et al. Linking Glycomics Repositories with Data Capture*) and the setting of exchange formats (W. York and R. Ranzinger, *MIRAGE: Minimum Information*

Required for A Glycomics Experiment, in the same volume). The third lesson is illustrated with the first practical steps undertaken to study and interrelate the sugars mediating host-pathogen interactions.

Bioinformatics Background

Database quality assessment

The uneven content and quality of data in public databases is a central issue of bioinformatics. Fast and massive data production is often achieved to the detriment of quality. Given the significant amount of information that is automatically inferred from experimental data (e. g., sequence annotation from complete genome data), error propagation can be far-reaching. Frequent updates are intended to address the quality problem by monitoring and correcting reported errors. The **frequency of update** is thus one among many criteria for assessing on-line data quality. Reliability and exhaustiveness are two other criteria that are used to assess the quality of a source.

The **reliability** of a source often comes through usage and steadily increases with time. Internal standards (e. g., evidence tagging, normalised scores) contribute to it. Note that the selection of data sources remains a subjective task often simply geography-dependent. Some methods (e. g., [5]) are defined to rank public biological data according to reliability metadata. This kind of approach aims at guiding and assisting users undertaking bioinformatics analyses. More initiatives are under development in the context of the Semantic Web (e. g., [6]).

Exhaustiveness is a changing criterion. Exhaustiveness long meant ever-increasing size and was applied to counting objects (sequences, species, etc., 10^6 to 10^8 orders of magnitude). An exhaustive set is now accepted as including a relatively small number of entities (10^3 or 10^4 order of magnitude). For instance, the number of entries (genes or proteins) of a species-specific genome or proteome database is not expected to grow but information related to each entry is supposed to expand. Properties of objects are actually expected to be, if not exhaustive, at least comprehensive. Database expansion has thereby shifted from breadth to depth and from objects to properties. Collecting and bringing properties into line involves human expertise and the resulting curated data is definitely knowledge as opposed to data. Such a trend is exemplified with the recent evolution of Amos Bairoch, the founder in the 1980 s of the all-encompassing Swiss-Prot database, (exponentially growing to soon reach the million landmark) to be the initiator in 2008 of the NextProt database dedicated to the 20,000 human gene products.

Data integration

The tasks of integration can be summarised as (i) the selection of sources from which information is extracted and pooled in one location, (ii) the quantification and interconnection of gathered information that includes the assessment of the relative importance of each piece of information, (iii) the visualisation of interconnected information, and (iv) the gradual definition and implementation of one or more underlying principles for structuring information (i.e., ontologies and models). This last task involves questions related to the chronology and dynamics of events that are the most challenging aspects of data integration.

Standards and formats

The Protein Standard initiative (PSI, http://www.psidev.info) was launched 10 years ago in the field of proteomics. It followed a similar earlier move in transcriptomics [7] where the accumulated microarray data started to be uselessly repeated in the absence of a standardised representation of what seemed to be straightforward 2-dimensional datasets. The PSI imposed a laborious exercise in harmonising input/output formats from MS instruments and analysis software, and in normalising information extraction from databases and repositories. These efforts are contemporaneous with the rising of XML in bioinformatics, as this mark-up language facilitates sharing both (semi-structured) formats and data [8]. Not surprisingly, most current proteomics standards are XML-based [9].

Exchange formats: the example of protein-protein interaction annotation

One of the recent proteomics standards governs information extraction from the literature to support curation in databases of protein-protein interactions such as IntAct (http://www.ebi.ac.uk/intact) or MINT (http://mint.bio.uniroma2.it/mint). More specifically, the Minimum Information required for reporting a Molecular Interaction experiment (MIMIx) includes the identity of molecules that participate in an interaction (with accession number), the methods by which both the interaction and the identity of the participants were established and the role of these molecules in the context of the experiment (as distinct from their biological role).

The International Molecular Exchange (IMEx) consortium of protein-protein interaction databases committed to annotating records to an agreed, published standard (http://imex.sourceforge.net/) and to exchange these records on a regular basis. Members of the IMEx consortium act both as hosts who provide the storage facilities for submitted data and as annotators of the information existing in the literature. This guarantees not only the quality but also the visibility of inter-connected resources, and emphasises the advantage of adopting data curation standards. This connectivity and subsequent advantage is made obvious in diagrams of the Pathguide portal (http://www.pathguide.org/interactions.php) that catalogues pathway and protein-protein interaction databases.

Network biology

The shift from objects to properties described in the section *Data integration* is contemporary with the rapid development of the network representation trend. In the last decade, high throughput techniques of all –omics fields have produced results that are, in many cases, systematically mapped as networks of interactions using Cytoscape (www.cytoscape.org), the most popular software for network visualisation and manipulation.

Where Glycomics Fits into the Picture

Data standards and databases

Several databases describing the variety and the complexity of glycan structures on glycoproteins are available online. However, these resources are either focused on describing sugars or on describing proteins in a rather mutually exclusive manner. Published literature is the only common thread available to link two complementary types of information. In fact, the exchange and cross-referencing of data between glyco-related databases is often hampered by the lack of generally accepted exchange formats and structure description standards. Conventional encoding with the IUPAC nomenclature was adopted for linear representation but machine-readable formats for 2D structures are also needed. GlycoCT [10] was proposed by a bioinformatics group headed by Willi von der Lieth at the DKFZ (German Cancer Research Centre) in Heidelberg who re-initiated the development of informatics for glycobiology after the demise of the Complex Carbohydrate Research Centre's structure database, CarbBank, in 1997. This group developed the first glycoscience portal (http://www.glycosciences.de) and was later the leader of the EUROCarbDB project (http://www.eurocarbdb.org). Despite their efforts, the GlycoCT format is not widely spread.

Despite the delay, glycomics is now following in the footsteps of other -omics fields by introducing the initiative on Minimum Information Required for A Glycomics Experiment (MIRAGE) standard this year (see contribution of W. York and R. Ranzinger, *MIRAGE: Minimum Information Required for A Glycomics Experiment*, in the same volume).

Data integration

Glycomics data and knowledge now is spread around in several independent resources. Two main initiatives have attempted to link complementary information. The Consortium for Functional Glycomics (CFG, http://www.functionalglycomics.org) became, in 2001, the first large, well funded project to address the need for glycoinformatics and endorsed the management and automatic annotation of experimental data generated by glycomics research. In parallel, the integration of glyco-related biological pathways by the Kyoto Encyclopaedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg) and inclusion of an associated database of glycan structures, led to the establishment of connections between glycan

structures and enzymatic pathways (KEGG-Glycan, http://www.genome.jp/kegg/glycan) [20]. Several other database initiatives have been started in recent years due to this increasing interest in glycomics research (see [11]).

The use of mass spectrometry in analytical glycobiology has spread significantly as illustrated in a recent review [12] and UniCarb-DB (http://unicarb-db.biomedicine.gu.se/) has recently been initiated to collect and annotate MS/MS data on released glycoprotein oligosaccharides [13]. Previous efforts to integrate this type of data with sequence data as in the PeptideAtlas [4] are not directly transposable to glycomics.

Towards network glycobiology

The popularity of network biology has reached glycobiology and a topological study of the network of *N*-linked glycosylation was recently published [14]. The authors carried out a preliminary investigation into the large-scale organisation of glycosylation pathways in mammalian cells in order to propose rules for glycoprotein engineering.

A logical step would rely on the CAZy database (http://www.cazy.org) to build networks from the knowledge of the range of glycoenzymes featured in CAZy. Visualisation tools will be essential in highlighting the cooperative effects of glyco-enzymatic reactions in response to environmental changes.

Practical First Steps

Existing environment

In the past eighteen months, concerted action has been initiated with the prospect of collecting and linking glyco-related data and its annotation in a single virtual space. Knowledge of the structure and function of the glycosylation of proteins will be improved in much the same way as the structure and function of proteins was facilitated by the integration of proteomics knowledge into UniProtKB (http://www.uniprot.org). A nascent consortium [15] is collecting as much data on the glycosylation of proteins that is available to each participant in an attempt to create a UniCarbKB equivalent for the integration of glycomics knowledge (see contribution of M. P. Campbell *et al.*, *Linking Glycomics Repositories with Data Capture*, in the same volume). The consortium has reinforced its commitment to gathering and linking glycan information in a structure network of bioinformatics resources. A special effort is being directed toward facilitating the study of the role of sugars in infection.

Although it cannot compete with dedicated sites like GLYCOSCEINCES.de or the CFG portal, the ExPASy server (http://www.expasy.org) hosts several bioinformatics resources for glycobiologists. ExPASy was created in 1994 by members of PIG and of the Swiss-Prot

group of SIB [16]. Both groups co-developed and maintained the website until 2011 when the service originally known as a proteomics portal was extended to other bioinformatics applications. ExPASy provides a fair level of integration. Many hosted resources are cross-linked allowing the navigation across several databases and/or the combined use of software with databases. This is the case for instance for the first, and now ten years old, example of a glyco-bioinformatics tool, GlycoMod (http://web.expasy.org/glycomod/), that calculates the monosaccharide composition from mass spectrometry data [17] and predicts the structures possible with this composition by linking to GlycoSuiteDB, a curated database of glycan structures reported in the literature [21].

PIG has imported onto ExPASy in 2009 this key database, GlycosuiteDB (http://glycosuitedb.expasy.org/glycosuite/glycodb) originally developed in a commercial environment (Proteome Systems Ltd), and in 2010 brought in SugarBindDB (http://sugarbind.expasy.org) in preparation to stepping up their involvement in glyco-bioinformatics. The further development of SugarBind has become, in 2011, a collaborative project between the MITRE Corporation (http://www.mitre.org) who initiated the database, the Alimentary Glycoscience Research Cluster (AGRC) in Ireland and Prof. N. Packer's Glycomics Group in Australia.

The space shared by glyco-bioinformatics resources and many other –omics resources on the ExPASy server is a unique opportunity to integrate glycomics knowledge. Note that GlycosuiteDB is already cross-referenced in UniProtKB/Swiss-Prot and *vice-versa*.

Support for the study of host-pathogen interactions

As hinted in the introduction, the development of resources is best achieved in the context of a selection of biological questions. A point of focus is the mediation of host-pathogen interactions by sugars.

The scenario of bacterial infection involves molecular entities of both the host and the bacterium. Each has its own set of surface sugars and surface receptors that are susceptible to interaction. Roughly speaking, bacterial lectins recognise host sugars and host lectins recognise bacterial sugars [18]. The focus of research has remained so far mainly on one side of the story that is, elucidating recognition from the point of view of the host [19] since the adhesion of a microorganism to a surface sugar was identified as often the first step in human microbial pathogenesis. In some instances the host has evolved an innate protective mechanism, which actually mimics this initial adhesion interaction to trap invading organisms in their secreted fluids in order to subsequently clear them from the body. While various genetic tools such as microarrays have been applied for studies addressing gene regulation and immunological studies in pathogenesis (e. g., [22, 23]) the molecular basis for the initial cellular contact between the target cell and the invading pathogen, known to involve one or more protein-sugar interactions, is still poorly understood. Nevertheless,

the expression pattern of glycans combined with a better understanding of their functional roles is much needed. As the conserved glycan epitopes play an important role in infection, their prediction from mass spectrometry data appears like a reachable goal.

A breakdown of the specific glycan structures required for adhesion is necessary for mapping and comparing the precise sugar epitopes involved in the binding of glycoproteins to bacterial species. This data can advantageously be compared to previously published studies. Indeed, the investigation of the mediation of sugars in bacterial infection produced a wealth of knowledge that tends to be trapped in the literature. Information extraction is a long and tedious process undertaken by experienced readers. Only one database SugarBind (unpublished) collects information on host sugar epitopes and the bacterial proteins that recognise them as a result of manual curation. Note that a similar effort is undertaken in GlycoEpitope (http://www.glyco.is.ritsumei.ac.jp, unpublished) for eukaryotic proteins and that bacterial sugar structures are found in the Bacterial Carbohydrate Structure DataBase (BCSDB, http://www.glyco.ac.ru).

Table 1 summarises the range of resources available for each type of entity. It also highlights the current gaps in the knowledge that is stored in current resources and the imperative need for cross-linking them. It appears that the combination of GlycoSuiteDB (centred on the sugar structure (bacterial or eukaryotic) and the protein to which it is linked), and SugarBind, (presenting the recognition or presentation mechanism), spans the entities. The enhancement of these two resources will assist the determination of the structural specificity of glycoprotein oligosaccharides involved in adhesion.

Table 1. Summary of the range of resources available

	Glycosuite	SugarBind	GlycoEpitope	BCSDB	GRAL	Lectines
Sugars on host cell surface	yes	yes	yes			
Sugars on bacterial cell surface	yes			yes		
Sugar epitopes		yes	yes			
Host lectins	yes		yes		yes	yes
Bacterial lectins		yes				yes
Other glycoproteins	yes		yes			

Evolution of SugarBind

The SugarBind database collects protein-carbohydrate binding pairs associated with initial stages of infection by human pathogens. Further development is planned and on-going data curation is provided by a trained glycobiologist. SugarBind originally listed bacterial and viral pathogens, biotoxins, and proteins associated with adhesive structures including hemagglutinins, fimbriae, pili and adhesins with associated gene names, as well as their reported glycan binding ligands. Recent enhancements include the addition of recently

published material, pathogen strain designations, and new ligands. Furthermore, the cross-linking of SugarBind to GlycoSuiteDB (glycan structural data linked with protein source and tissue information) is in progress. This connects the pathogen adhesion information to the oligosaccharide structures attached to proteins and/or lipids through the IUPAC linear description of sugar structures, which is a common feature of both databases. The cross-linking of these two databases then will allow the integration of this protein-sugar interaction data into the overall curated knowledgebase of glycan information (UniCarbKB, http://unicarbkb.com, M.P. Campbell *et al. Linking Glycomics Repositories with Data Capture*, in the same volume) that is being assembled to facilitate the analysis of the structure and function of glycoconjugates.

The figure below describes with straight lines the cross-links already or soon to be implemented between resources. Planned links feature in dotted lines. Navigation between the different resources will support the functional interpretation of sugars.

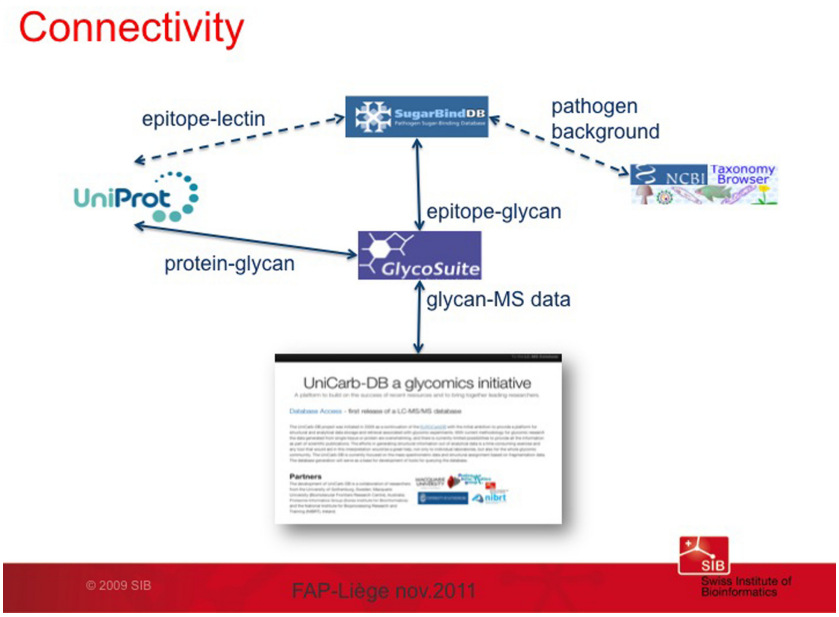

Figure 1. Cross-links between bioinformatics resources. Full lines show existing or upcoming cross-links. Dotted lines show planned work.

Conclusion

Modern day biology has proven unfeasible without the support of bioinformatics resources given the success of high throughput analytical methods. In the specific case of glycomics, this support is lagging behind mainly due to the complexity of sugar structures and their

carriers and the subsequent unknown extent of the functional role of these entities in cellular communication in general and in infection in particular. Current initiatives as described in this volume are intended to be a major step towards interconnecting isolated efforts and as such, are meant to boost faster progress.

As a first step, the molecular characterisation of microbial carbohydrate-binding proteins and oligosaccharide epitopes in infectious diseases will benefit from using bioinformatics tools that support the curation of mammalian glycan structures involved in infectious diseases and sets the basis for the prediction of antigenic sugar (sub-)structures.

Acknowledgements

We particularly thank Elaine Mullen, Leela Pavan Tadoori, Luc Mottin, Khaled Khatib and Julien Mariethoz for their contribution to the release of SugarBind in August 2011.

References

[1] Burks, C., Fickett, J.W., Goad, W.B., Kanehisa, M., Lewitter, F.I., Rindone, W.P., Swindell, C.D., Tung, C.S., Bilofsky, H.S. (1985) The GenBank nucleic acid sequence database. *Comput. Appl. Biosci*. **1**(4):225 – 33.

[2] Bairoch, A., Boeckmann, B. (1991) The SWISS-PROT protein sequence data bank. *Nucleic Acids Res.* **19** Suppl:2247 – 9.

[3] Frishman, D., Albermann, K., Hani, J., Heumann, K., Metanomski, A., Zollner, A., Mewes, H.W. (2001) Functional and structural genomics using PEDANT. *Bioinformatics* **17**(1):44 – 57.
doi: http://dx.doi.org/10.1093/bioinformatics/17.1.44.

[4] Desiere, F., Deutsch, E.W., King, N.L., Nesvizhskii, A.I., Mallick, P., Eng, J., Chen, S., Eddes, J., Loevenich, S.N., Aebersold, R. (2006) The PeptideAtlas project. *Nucleic Acids Res*. **34:**D 655 – 8.
doi: http://dx.doi.org/10.1093/nar/gkj040.

[5] Cohen-Boulakia, S., Biton, O., Davidson, S., Froidevaux, C. (2007) BioGuideSRS: querying multiple sources with a user-centric perspective. *Bioinformatics* **23**(10):1301 – 3.
doi: http://dx.doi.org/10.1093/bioinformatics/btm088.

[6] Ellrott, K., Zmasek, C.M., Weekes, D., Sri Krishna, S., Bakolitsa, C., Godzik, A., Wooley, J. (2011) TOPSAN: a dynamic web database for structural genomics. *Nucleic Acids Res*. **39:**D 494 – 6.
doi: http://dx.doi.org/10.1093/nar/gkq902.

[7] Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., Vingron, M. (2001) Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nat. Genet*. **29**(4):365 – 71.
doi: http://dx.doi.org/10.1038/ng1201-365.

[8] Achard, F., Vaysseix, G., Barillot, E. (2001) XML, bioinformatics and data integration. *Bioinformatics* **17**(1): 115 – 125.
doi: http://dx.doi.org/10.1093/bioinformatics/17.2.115.

[9] Lisacek, F., Cohen-Boulakia S., Appel R.D. (2006) Proteome informatics II: Bioinformatics for comparative proteomics. *Proteomics* **6:** 5445 – 5466.
doi: http://dx.doi.org/10.1002/pmic.200600275.

[10] Herget, S., Ranzinger, R., Maass, K., Lieth, C.W. (2008) GlycoCT-a unifying sequence format for carbohydrates. *Carbohydr. Res*. **343**(12): 2162 – 71.
doi: http://dx.doi.org/10.1016/j.carres.2008.03.011.

[11] Lütteke, T. (2008) Web resources for the glycoscientist. *Chembiochem* **9**(13):2155 – 60.
doi: http://dx.doi.org/10.1002/cbic.200800338.

[12] Zaia, J. (2008) Mass spectrometry and the emerging field of glycomics. *Chem. Biol*. **15**(9):881 – 92.
doi: http://dx.doi.org/10.1016/j.chembiol.2008.07.016.

[13] Hayes, C.M., Karlsson, N.G., Struwe, W., Rudd, P.M., Lisacek, F., Packer, N.H., Campbell, M.P. (2011) UniCarb-DB: A database resource for glycomic discovery *Bioinformatics* **27**(9):1343 – 4.
doi: http://dx.doi.org/10.1093/bioinformatics/btr137.

[14] Kim, P.J., Lee, D.Y., Jeong, H. (2009) Centralized modularity of N-linked glycosylation pathways in mammalian cells. *PLoS One* **4**(10):e7317.
doi: http://dx.doi.org/10.1371/journal.pone.0007317.

[15] Campbell, M.P., Hayes, C.A., Struwe, W.B., Wilkins, M.R., Aoki-Kinoshita, K.F., Harvey, D.J., Rudd, P.M., Kolarich, D., Lisacek, F., Karlsson, N.G., Packer, N.H. (2011) UniCarbKB: putting the pieces together for glycomics research. *Proteomics* **11**(21):4117 – 4121.
doi: http://dx.doi.org/10.1002/pmic.201100302.

[16] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31:**3784 – 3788.
doi: http://dx.doi.org/10.1093/nar/gkg563.

[17] Cooper, C.A., Gasteiger, E., Packer, N.H (2001) GlycoMod–a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* **1**(2):340 – 9.
doi: http://dx.doi.org/10.1002/1615-9861(200102)1:2<340::AID-PROT340>3.0.CO;2-B.

[18] Hitchen, P.G., Dell, A. (2006) Bacterial glycoproteomics. *Microbiology* **152:** 1575 – 1580.
doi: http://dx.doi.org/10.1099/mic.0.28859-0.

[19] Cummings, R. (2009) The repertoire of glycan determinants in the human glycome. *Mol. BioSyst* **5:**1087 – 1104.
doi: http://dx.doi.org/10.1039/b907931a.

[20] Hashimoto, K., Tokimatsu, T., Kawano, S., Yoshizawa, A.C., Okuda, S., Goto, S., Kanehisa, M. (2009) Comprehensive analysis of glycosyltransferases in eukaryotic genomes for structural and functional characterization of glycans. *Carbohydr. Res*. **344**(7):881 – 7.
doi: http://dx.doi.org/10.1016/j.carres.2009.03.001.

[21] Cooper, C.A., Joshi, H.J., Harrison, M.J., Wilkins, M.R., Packer, N.H. (2003) GlycoSuiteDB: a curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res*. **31:**511 – 513.
doi: http://dx.doi.org/10.1093/nar/gkg099.

[22] Joenvaara, S., Ritamo, I., Peltoniemi, H., Renkonen, R. (2008) N-glycoproteomics – an automated workflow approach. *Glycobiology* **18**(4): 339 – 49.
doi: http://dx.doi.org/10.1093/glycob/cwn013.

[23] Sharon, N. (2006) Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochim. Biophys. Acta* **1760** (4): 527 – 537.
doi: http://dx.doi.org/10.1016/j.bbagen.2005.12.008.