



Article scientifique

Article

2014

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

UniCarbKB: building a knowledge platform for glycoproteomics

Campbell, Matthew P.; Peterson, Robyn; Mariethoz, Julien; Gasteiger, Elisabeth; Akune, Yukie; Aoki-Kinoshita, Kiyoko F.; Lisacek, Frédérique; Packer, Nicolle H.

How to cite

CAMPBELL, Matthew P. et al. UniCarbKB: building a knowledge platform for glycoproteomics. In: Nucleic acids research, 2014, vol. 42, n° D1, p. D215–D221. doi: 10.1093/nar/gkt1128

This publication URL: <https://archive-ouverte.unige.ch/unige:159521>

Publication DOI: [10.1093/nar/gkt1128](https://doi.org/10.1093/nar/gkt1128)

UniCarbKB: building a knowledge platform for glycoproteomics

Matthew P. Campbell¹, Robyn Peterson¹, Julien Mariethoz², Elisabeth Gasteiger³, Yukie Akune⁴, Kiyoko F. Aoki-Kinoshita⁴, Frederique Lisacek^{2,5} and Nicolle H. Packer^{1,*}

¹Biomolecular Frontiers Research Centre, Macquarie University, North Ryde, NSW 2109, Australia, ²Proteome Informatics Group, Swiss Institute of Bioinformatics, Geneva, Switzerland, ³Swiss-Prot Group, Swiss Institute of Bioinformatics, Geneva, Switzerland, ⁴Department of Bioinformatics, Faculty of Engineering, Soka University, 1-236 Tangi-machi, Hachioji, Tokyo, Japan and ⁵Section of Biology, Faculty of Sciences, University of Geneva, Switzerland

Received September 14, 2013; Revised October 18, 2013; Accepted October 24, 2013

ABSTRACT

The UniCarb KnowledgeBase (UniCarbKB; <http://unicarbk.org>) offers public access to a growing, curated database of information on the glycan structures of glycoproteins. UniCarbKB is an international effort that aims to further our understanding of structures, pathways and networks involved in glycosylation and glyco-mediated processes by integrating structural, experimental and functional glycoscience information. This initiative builds upon the success of the glycan structure database GlycoSuiteDB, together with the informatic standards introduced by EUROCarbDB, to provide a high-quality and updated resource to support glycomics and glycoproteomics research. UniCarbKB provides comprehensive information concerning glycan structures, and published glycoprotein information including global and site-specific attachment information. For the first release over 890 references, 3740 glycan structure entries and 400 glycoproteins have been curated. Further, 598 protein glycosylation sites have been annotated with experimentally confirmed glycan structures from the literature. Among these are 35 glycoproteins, 502 structures and 60 publications previously not included in GlycoSuiteDB. This article provides an update on the transformation of GlycoSuiteDB (featured in previous NAR Database issues and hosted by ExPASy since 2009) to UniCarbKB and its integration with UniProtKB and GlycoMod. Here, we introduce a refactored database, supported by substantial new curated data collections and intuitive user-interfaces that improve database searching.

INTRODUCTION

Protein glycosylation is an important and universal post-translational modification that is estimated to occur on between 20% and 50% (1,2) of all secreted and cellular proteins. Glycoproteins are characterized by the presence of oligosaccharides linked to the peptide backbone through *N*- or *O*-glycosidic bonds at asparagine or serine/threonine residues, respectively. For both *N*- and *O*-glycosylation, there can be considerable diversity of glycan structures associated with each glycosylation site. Such micro-heterogeneity is governed by an elaborate process carried out by numerous intricate and competitive steps, which result in the generation of tissue and cell type-specific glycan expression patterns.

Given that protein glycosylation is involved in numerous cellular processes and is implicated in disease progression (3–5), the ability to accurately characterize glycan structures (at a global and site-specific manner) and the identification of the modified proteins is increasingly important in functional glycomics (6–8). The molecular and functional complexity of glycoproteins is challenging and requires sustainable bioinformatic resources aimed at capturing, integrating and maintaining the available knowledge. The more complete our understanding of glycosylation is the better equipped we will be to understand the functional and structural roles of both glycoproteins and the attached glycans at the molecular level. Unfortunately and despite the success of several international initiatives the glycosciences still lack a managed infrastructure that contributes to the advancement of research through the provision of comprehensive structural and experimental glycan data collections. As described by the US National Academy of Sciences report ‘Transforming Glycoscience: A Roadmap for the Future’ (9) an important factor in broadening the appreciation of glycomics is the necessity to develop robust, scalable and standardized bioinformatic platforms to

*To whom correspondence should be addressed. Tel: +61 2 9850 8176; Fax: +61 2 9850 8313; Email: nicki.packer@mq.edu.au

acquire and disseminate the information-rich data collections that are becoming increasingly available.

OVERVIEW OF UniCarbKB

UniCarbKB is an initiative that aims to promote the creation of a curated, glycan structure based, online information storage and search platform for glycoscience research (10). This initiative builds upon the previously successful databases, the Australian developed GlycoSuiteDB (11) and the EU-funded EUROCarbDB (12), to offer a freely accessible updated platform built with modern front and back-end technologies. The re-engineered framework offers an intuitive user-interface with enhanced features and greater support for on-going international efforts to establish common data standards that will better integrate structural, experimental and functional data collections.

GlycoSuiteDB

GlycoSuiteDB is acknowledged as the first effort to provide detailed non-redundant, curated structural information derived from the published literature on conjugated glycans. The database connects glycan structure and biological origin with protein specific information where known. For each glycan structure (e.g. glycan type, mass and composition), detailed information is provided on the native and recombinant sources (i.e. tissue and/or cell type, cell line, strain and disease state), with appropriate links to Swiss-Prot/TrEMBL entries, a record of the methods used to determine the structure, and the PubMed ID of the cited publication. The design objectives and functionality of GlycoSuiteDB has been published in previous NAR database issues (11,13). Originally developed commercially, it has since been made available publicly through the ExPASy server (14). Access to the content has not only preserved the efforts of the curation team, but is now helping to seed the UniCarbKB effort to build an up-to-date high-quality resource for glycoscience.

EUROCarbDB

The EUROCarbDB project was a collaborative European design study that focused on building the foundations of a technical framework to support glycobioinformatic activities. This resulted in the provision of sophisticated open-source tools and structure encoding formats and databases that, to date, continue to support several facets of analytical glycomics. The architecture of the EUROCarbDB database started to address stumbling blocks impeding progress in glycomics by providing the glycobiology community with (i) universal standards for the representation of monosaccharides and complex glycans, (ii) a freely accessible database of known glycan structures and experimental evidence, (iii) freely accessible analytical tools for researchers and (iv) a technical framework of open-source code.

UniCarbKB: building upon the foundations laid by GlycoSuiteDB and EUROCarbDB

UniCarbKB is focused on enhancing existing tools, standards and applications to be more accessible and amenable to modern research workflows. In particular we have leveraged previous experiences to build a modern and scalable framework, which uses technologies and web frameworks that are more familiar to developers. As the first step we have merged the glycan structural information from the no longer supported GlycoSuiteDB and EUROCarbDB initiatives into a high-quality updated framework. Several libraries developed by the EUROCarbDB initiative including GlycanBuilder (15), MonosaccharideDB (<http://www.monosaccharidedb.org>) and GlycoCT (16) have also been incorporated.

DESIGN AND IMPLEMENTATION

UniCarbKB is built with the open-source framework Play (Release 2.1.3) written in Java and Scala, which follows the model-view-controller architecture. The views (user-interface) are predominantly written in Scala and include JQuery and Bootstrap Javascript libraries. The model and controller layers are written in Java and the Ebean object-relational mapping (ORM) library is used to query the underlying database model. UniCarbKB uses PostgreSQL (Version 9.2) as the underlying database system that consists of multiple schemas to ensure data integrity by managing structural, literature and experimental data collections.

We have updated the front-end of GlycoSuiteDB with a refreshed interface that is easier to navigate and focuses on displaying content the researcher wants to access. The new update is more visual, includes new, simpler content layout and improved accessibility options. All pages now display a menu bar to make it easier to access features and navigate around the website in a consistent way. The inclusion of JavaScript libraries and Bootstrap has allowed us to enhance the user experience. For example, we have changed the search functionality by including an auto completion feature in addition the use of pagination and drop-down lists improve the handling and display of large data collections. During the design process a series of sketched user interface wireframes and mockups were created, which allowed developers to consult with the end-user to test and refine navigation, evaluate the effectiveness of page layouts and determine web development/programming requirements.

DATABASE CONTENT

Primarily, UniCarbKB is a eukaryotic glycoprotein-centric resource built on the corpus of curated information originating from GlycoSuiteDB and a select few datasets from EUROCarbDB. We have expanded the content by manually curating over 60 more recent publications that contain (partial or completely characterized) glycan structures with supporting experimental data that substantially extend the content coverage of GlycoSuiteDB. A majority of the newly sourced data are derived from a literature

study by Thaysen-Andersen and Packer, which sought to correlate *N*-glycan structures characterized from purified glycoproteins with protein structure (17). This comprehensive dataset will contribute over 470 glycosylation sites from over 160 mammalian glycoproteins, from different tissues and body fluids to the over 3700 glycan structure entries, 400 glycoproteins and 598 protein glycosylation sites already curated by UniCarbKB. For each glycoprotein record two levels of database annotations are provided: (i) site-specific data for individual glycan structures that are associated to an amino acid sequence position and (ii) where a single purified glycoprotein has been analysed, all characterized glycan structures are linked to the glycoprotein accession number in UniProtKB (14). Also, new structural and experimental glycan data have been contributed from the integration of the final public release of GlycoBase (18) developed in conjunction with EUROCarbDB.

SEARCH FEATURES

As part of our effort to improve the overall user experience a series of new interfaces have been implemented

(Figure 1). Many build upon and retain the features available in GlycoSuiteDB but with improved functionality. For example, we have (i) enhanced native selects by including a multi-select interface (ii) made greater use of Javascript to load items via Ajax supporting the ability to partially load dataset and (iii) paginated for better support of the presentation and navigation of large datasets. Supported queries include (sub)structure, monosaccharide composition, glycan mass, taxonomy, tissue, disease, glycoprotein (accession number or Swiss-Prot name) and literature publication.

Structure searching with GlycanBuilder

Previously, GlycoSuiteDB provided a structure interface that consisted of textual and form based input, however, many researchers prefer to graphically visualize glycan structures due to their inherent complexity. We have incorporated GlycanBuilder (Vaadin Release) (19) into the search functionality of UniCarbKB that supports the exact or partial matching of structures in the database. The user may (i) build a new structure, (ii) extend a structure from a predefined list or (iii) build a substructure/epitope; in all instances the anomeric configuration of a

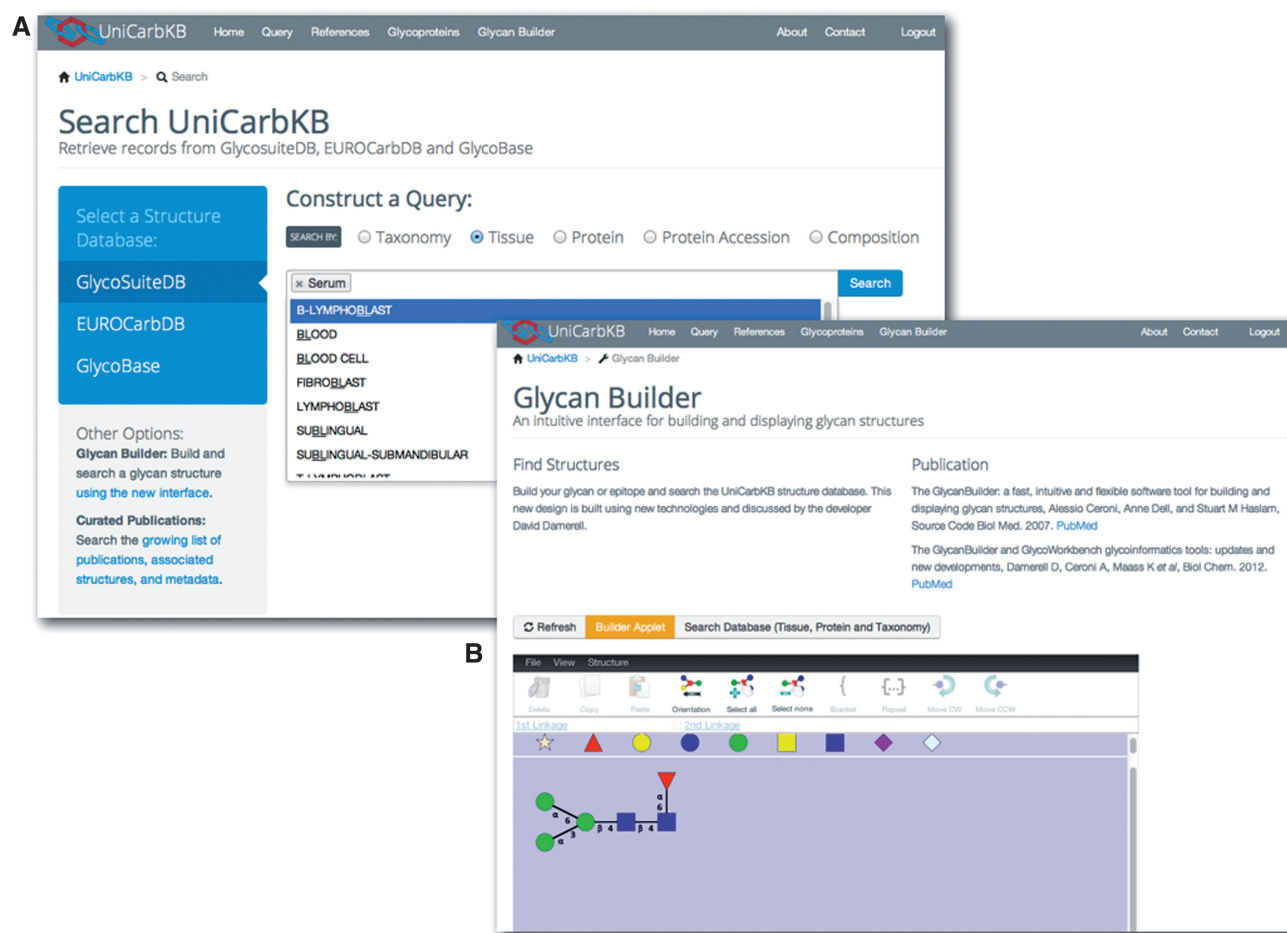


Figure 1. UniCarbKB offers a number of improved query interfaces. (A) Users can search the database content by monosaccharide composition, attached protein, taxonomy or tissue by using an auto completion feature. (B) The integration of latest version of GlycanBuilder allows users to query UniCarbKB by (sub)structure searching.

monosaccharide residue and linkage type can be defined. By default an exact search will only retrieve those database structures that perfectly match the topology, linkage and anomeric configuration submitted. In the case of partial searching, a level of fuzziness is introduced, whereby unknown information is handled as wildcards by the search algorithm. For substructure searching only those structures that have the (extended) epitope or motif built will be returned.

Glycan structure encoding

A plethora of graphical and textual formats are available for the depiction of glycan structures including: the Consortium for Functional Glycomics/Essentials, the Oxford nomenclature and IUPAC formats. Historically, GlycoSuiteDB encoded glycan structures in an IUPAC style format, however, recent databases have adopted connection table approaches exemplified by GlycoCT and KCF (KEGG Chemical Function) (20) to describe oligosaccharide sequences with a controlled vocabulary. By extension of these efforts UniCarbKB supports the storage of GlycoCT and IUPAC formats, and to further extend database interoperability an IUPAC to KCF and a modified IUPAC to GlycoCT translator have recently

been developed to complement existing translators. Similar to GlycomeDB and EUROCarbDB we have implemented a feature that enables users to switch between supported graphical formats. This feature is made possible by integrating the GlycanBuilder API, which produces high-quality representations of glycan structures.

METADATA EXTENSIONS

To enable users of UniCarbKB to assess the reliability of the contained information, provenance metadata must be recorded. Provenance metadata relates to the origin of the data and deals less with the finer details and more with the process of how the data came to be.

Biological source

The biological context module, developed by EUROCarbDB, handles the association of structure to biological source that amalgamates taxonomy and tissue, together with a varying number of disease and perturbation associations (Figure 2). The library adopts the controlled vocabularies derived from the NCBI Taxonomy and the MeSH (Medical Subject Headings) databases. Its inclusion reduces data redundancy by providing a hierarchical controlled vocabulary that links

The screenshot displays the UniCarbKB interface for the Coagulation Factor IX entry. The top navigation bar includes links for Home, Query, References, Glycoproteins, and Glycan Builder. The main content area is titled 'Associated Structures' and shows the UniProtKB/Swiss-Prot PTM Description for Coagulation Factor IX. Below this, there is a table for Glycosylation Sites with columns for Position, Structures, Description, and Evidence. The table lists two sites: SER-99, SER-107 and SER-107. The SER-107 site is further detailed in the Site Specific Information section. The Glycan Structures section shows three glycan structures. The Structure Format section allows users to select between ESSENTIALS, TRADITIONAL, and OXFORD formats. The Sequence section displays the protein sequence. The Biological Associations section shows links to TAXONOMY, PROTEIN, and SOURCE. The References section lists two references related to the structure of the protein.

Figure 2. Screenshot for the Coagulation Factor IX entry in UniCarbKB. The database provides a description of the glycan structures characterized for this glycoprotein and the number of structures associated with experimentally confirmed glycosylation sites. A general description of the glycoprotein is provided that has been derived from the relevant UniProtKB entry in addition to the protein sequence. Users have the Structure Format option to select a preferred graphical format to display the glycan structures that includes the three most commonly used notations. Supporting biological metadata and publication information is also provided. The Biological Associations include details pertinent to the species, the individual protein and tissue/secretory fluid source. Finally, those References that have been manually curated to obtain this information are summarized with appropriate links to PubMed.

specific taxonomic descriptions with more generalized terms e.g. specific tumours or cancer of the lung are grouped under the more general term 'Lung Neoplasms'. This approach improves upon the disconnected terms used in GlycoSuiteDB, by providing a more robust interface to searching and grouping together glycan structures based on taxonomic or disease terms.

Identification of methods

The reporting of descriptive metadata that is representative of the reported literature poses many challenges, but is essential for the development of a well-documented glycan and glycoprotein database. Efforts led by the Minimum Information for A Glycomics Experiment (MIRAGE) (21) project and the ontology work of GlycoRDF aim to alleviate this situation by providing standardized data entry terms, therefore fulfilling one of the recommendations of the NAS Committee on Assessing the Importance and Impact of Glycomics and Glycosciences. UniCarbKB has started to address those standardization guidelines proposed by MIRAGE, by establishing a high-level vocabulary that captures (i) the sample preparation procedures; encompassing glycan release techniques and/or methods that alter glycan structure, including exoglycosidase treatment and derivatization, (ii) the general analytical approach and (iii) the use of complementary validation methods such as lectin studies and monosaccharide analysis. This information is provided for all published references that have been curated and the vocabulary is continually expanded to reflect database content. By listing the methods used by the authors of the publication to determine the structure, users can determine their own level of confidence in the reported structures; in particular, by assessing the suitability of orthogonal methods such as array platforms, capillary electrophoresis, gas chromatography, lectin-binding, liquid chromatography, mass spectrometry and nuclear magnetic resonance.

N-glycan pathways

GlycanSynth is a new feature in UniCarbKB that integrates known genes and enzymes involved in the biosynthesis of *N*-glycans. A list of enzymes was manually curated from the Kyoto Encyclopedia of Genes and Genomes (22) and GlycoGene (23) databases. Data related to enzyme activity, including but not limited to glycosylation-related processes were also catalogued from the BRENDA (24) and UniProt databases. In addition, the Consortium for Functional Glycomics (25) and Carbohydrate-Active enzymes (26) were used as valuable resources for extracting glycosyltransferase genes and related downstream targets information. Furthermore, we aggregated appropriate gene information from the National Center for Biotechnology Information (NCBI).

For each catalogued protein *N*-glycosylation-related gene name we constructed a broad set of disaccharide reactions that match gene against a particular donor and acceptor substrate. In total, 37 glycosyltransferases have been documented that are involved in the synthesis of

N-glycan structures in humans stemming from the Man5 structure. A list of these gene names, enzymes and reactions is available at <http://unicarbkb.org/enzymes>. By using these reaction rules it is possible to (i) connect gene function with glycan structure and (ii) validate the accuracy of structures in a database based on implicit knowledge of the glycosylation machinery. This will be achieved by encoding the disaccharide sequences in the GlycoCT condensed format or IUPAC form, and using a tree traversal technique to assign linkage information.

INTERFACING UNICARBKB WITH EXTERNAL RESOURCES

Following the release of GlycoSuiteDB in 2002 several international initiatives have developed structural and experimental glycan databases notably the CFG, EUROCarbDB, BCSDb (27), RINGS (28) and JCGGDB. A key component of UniCarbKB is to forge relationships with these valuable resources. In the first instance we have worked with the glycan MS/MS data repository, UniCarb-DB (29) and liquid chromatography retention data collection, GlycoBase (18) (projects that stemmed from EUROCarbDB) to cross-reference these databases of experimental data together through structure-based URL links in UniCarbKB. In partnership with Australian National Data Service we have integrated UniCarbKB curated data collections with Research Data Australia—a discovery platform that enhances connections between data projects, researchers and institutions aimed at promoting the visibility of research. Also, the GlycoMod tool (14) (hosted at ExPASy <http://web.expasy.org/glycomod>) designed to predict oligosaccharides structures from experimentally determined masses is now directly linked to UniCarbKB; connecting theoretically possible compositions with curated glycan structures. Finally, we have also made use of the UniProtJAPI Java web service (30), which facilitates the integration of UniProtKB data into our web application. Here, we extract the glycoprotein description from UniProtKB for all glycan structure entries that have an assigned protein accession number; such information is displayed to the user in each protein summary page (Figure 3).

FUTURE DEVELOPMENTS

We envisage that this resource will be extended in the future to encompass knowledge and information on all glycoconjugates, however, due to limited resources the emphasis initially will be placed on publications containing well characterized *N*- and *O*-linked structures and the associated experimental data on proteins derived from eukaryotic organisms. UniCarbKB will be updated on a regular basis with newly curated data collections. In the short term, we will also enhance the functional information of glycans by cross-linking the SugarBind database (31) to UniCarbKB and target sub-structures recognized by lectins.

We plan to make available a web service API this year to support access to UniCarbKB data. By using the API

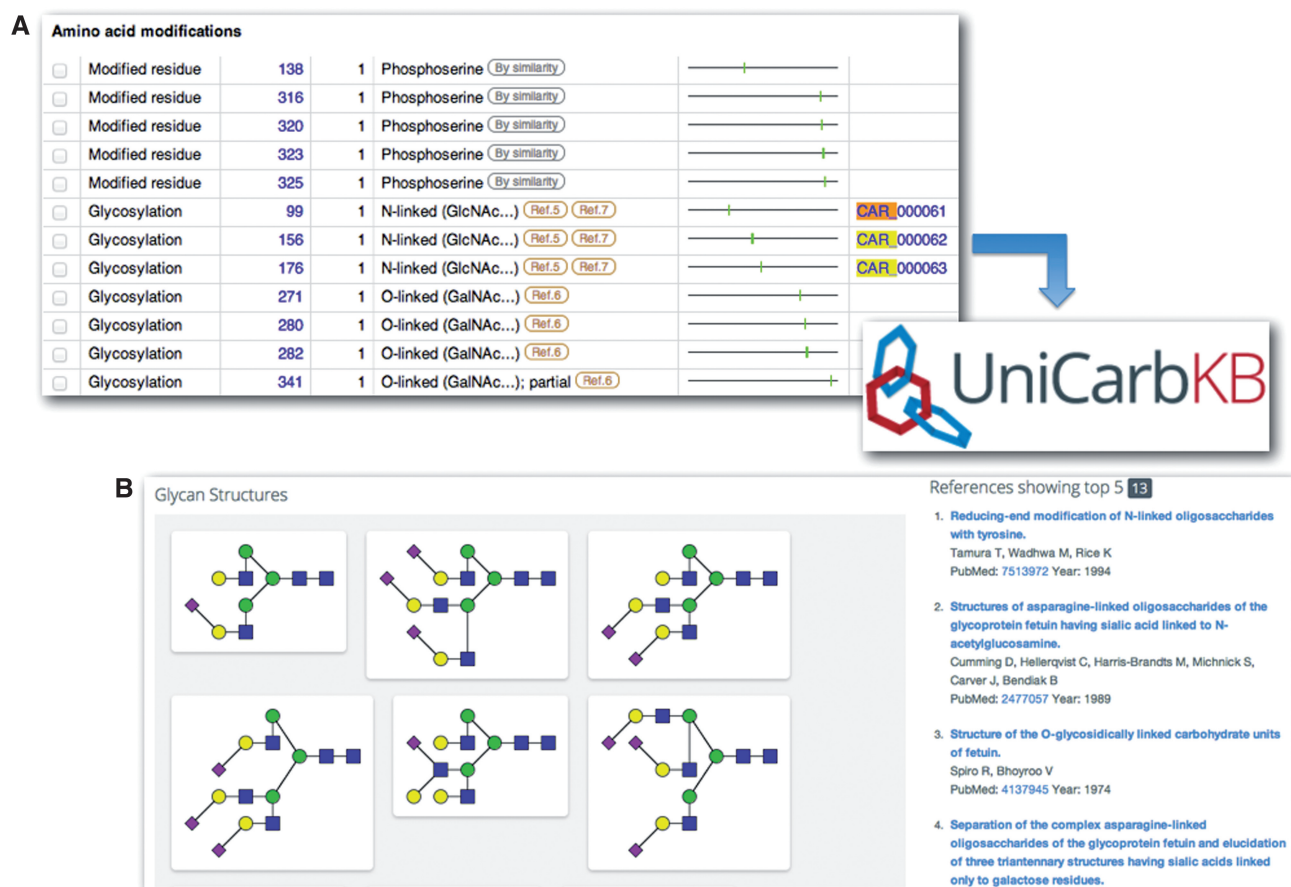


Figure 3. Navigating from UniProtKB to UniCarbKB. An example is shown for the (A) UniProtKB alpha-2-HS-glycoprotein entry that is linked to (B) the glycan structures on the individual amino acid sites as curated in UniCarbKB, by CAR identifiers. The curated records in the first release of UniCarbKB are being updated and included in the sequence annotation feature records of UniProtKB.

developers will be able to search against UniCarbKB and its affiliated mass spectral-based project UniCarb-DB. In conjunction with the GlycoRDF project we have started to represent our data in a standardized Resource Description Framework (RDF) format that will tackle the problems of disparate and decentralized databases by using Semantic Web technologies to unify content. We also plan to implement support for new tools that utilize the growing information stored in UniCarbKB e.g. ‘GlycoDigest’ (an exoglycosidase digestion prediction tool in development at SIB) and glycan translators that will support commonly used encoding formats including WURCS (Web3.0 Unique Representation of Carbohydrate Structures). To the best of our abilities, our development effort guarantees data exchange and tool compatibility (32). In the longer term we plan to establish UniCarbKB as a structure-centric, high-quality glycan database from which all available information on each glycan structure is easily accessible.

ACKNOWLEDGEMENTS

The authors thank the support provided by many developers and collaborators who have contributed

considerable effort to provide tools and resources for the glycosciences. In particular we thank teams involved in UniCarb-DB, GlycoBase, EUROCarbDB, GlycomeDB and PubChem. Finally, the authors acknowledge the efforts of MIRAGE and support from the Beilstein Institut to develop guidelines and standards, and the GlycoRDF project.

FUNDING

The Australian National eResearch Collaboration Tools and Resources project [NeCTAR RT016 to M.P.C and N.H.P]; Swiss National Science Foundation [SNSF 31003A_141215 J.M.]; Swiss Federal Government through the State Secretariat for Education, Research and Innovation SERI [F.L. and E.G.]; ExPASy is maintained by the web team of the Swiss Institute of Bioinformatics and hosted at the Vital-IT Competency Center; UniCarbKB was also supported by Agilent’s University Relations program to M.P.C and N.H.P; GlycoSuiteDB was developed by Proteome Systems Ltd [N.H.P] and transferred to SIB in 2009; EUROCarbDB was originally funded by European Union as a Research Infrastructure Design Study implemented as a Specific

Support Action under the FP6 Research Framework Program (RIDS Contract number 011952). Funding for open access charge: Australian National eResearch Collaboration Tools and Resources [NeCTAR RT016].

Conflict of interest statement. None declared.

REFERENCES

1. Khoury, G.A., Baliban, R.C. and Floudas, C.A. (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.*, **1**, 90.
2. Apweiler, R., Hermjakob, H. and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta.*, **1473**, 4–8.
3. Fuster, M.M. and Esko, J.D. (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat. Rev. Cancer*, **5**, 526–542.
4. Dube, D.H. and Bertozzi, C.R. (2005) Glycans in cancer and inflammation—potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.*, **4**, 477–488.
5. Raman, R., Raguram, S., Venkataraman, G., Paulson, J.C. and Sasisekharan, R. (2005) Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat. Methods*, **2**, 817–824.
6. Kolarich, D., Jensen, P.H., Altmann, F. and Packer, N.H. (2012) Determination of site-specific glycan heterogeneity on glycoproteins. *Nat. Protoc.*, **7**, 1285–1298.
7. An, H.J., Froehlich, J.W. and Lebrilla, C.B. (2009) Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr. Opin. Chem. Biol.*, **13**, 421–426.
8. Dwek, R.A. (1996) Glycobiology: toward understanding the function of sugars. *Chem. Rev.*, **96**, 683–720.
9. *Transforming Glycoscience: A Roadmap for the Future* (2012), Washington (DC).
10. 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. *et al.* (2011) UniCarbKB: putting the pieces together for glycomics research. *Proteomics*, **11**, 4117–4121.
11. Cooper, C.A., Harrison, M.J., Wilkins, M.R. and Packer, N.H. (2001) GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res.*, **29**, 332–335.
12. von der Lieth, C.W., Freire, A.A., Blank, D., Campbell, M.P., Ceroni, A., Damerell, D.R., Dell, A., Dwek, R.A., Ernst, B., Fogh, R. *et al.* (2011) EUROCarbDB: an open-access platform for glycoinformatics. *Glycobiology*, **21**, 493–502.
13. Cooper, C.A., Joshi, H.J., Harrison, M.J., Wilkins, M.R. and Packer, N.H. (2003) GlycoSuiteDB: a curated relational database of glycoprotein glycan structures and their biological sources. 2003 update. *Nucleic Acids Res.*, **31**, 511–513.
14. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E. *et al.* (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.*, **40**, W597–W603.
15. Ceroni, A., Dell, A. and Haslam, S.M. (2007) The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. *Source Code Biol. Med.*, **2**, 3.
16. Herget, S., Ranzinger, R., Maass, K. and von der Lieth, C.-W. (2008) GlycoCT—a unifying sequence format for carbohydrates. *Carbohydr. Res.*, **343**, 2162–2171.
17. Thaysen-Andersen, M. and Packer, N.H. (2012) Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology*, **22**, 1440–1452.
18. Campbell, M.P., Royle, L., Radcliffe, C.M., Dwek, R.A. and Rudd, P.M. (2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics*, **24**, 1214–1216.
19. Damerell, D., Ceroni, A., Maass, K., Ranzinger, R., Dell, A. and Haslam, S.M. (2012) The GlycanBuilder and GlycoWorkbench glycoinformatics tools: updates and new developments. *Biol. Chem.*, **393**, 1357–1362.
20. Aoki-Kinoshita, K.F. (2010) *Glycome Informatics: Methods and Applications*. Chapman & Hall, London.
21. Kolarich, D., Rapp, E., Struwe, W.B., Haslam, S.M., Zaia, J., McBride, R., Agravat, S., Campbell, M.P., Kato, M., Ranzinger, R. *et al.* (2013) The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting mass-spectrometry-based glycoanalytic data. *Mol. Cell. Proteomics*, **12**, 991–995.
22. Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
23. Narimatsu, H. (2004) Construction of a human glycogene library and comprehensive functional analysis. *Glycoconj. J.*, **21**, 17–24.
24. Schomburg, I., Chang, A., Placzek, S., Sohngen, C., Rother, M., Lang, M., Munnaretto, C., Ulas, S., Stelzer, M., Grote, A. *et al.* (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.*, **41**, D764–D772.
25. Raman, R., Venkataraman, M., Ramakrishnan, S., Lang, W., Raguram, S. and Sasisekharan, R. (2006) Advancing glycomics: implementation strategies at the consortium for functional glycomics. *Glycobiology*, **16**, 82R–90R.
26. Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.*, **37**, D233–D238.
27. Toukach, F.V. and Knirel, Y.A. (2005) New database of bacterial carbohydrate structures. *Glycoconj. J.*, **22**, 216–217.
28. Akune, Y., Hosoda, M., Kaiya, S., Shinmachi, D. and Aoki-Kinoshita, K.F. (2010) The RINGS resource for glycome informatics analysis and data mining on the Web. *OmicS*, **14**, 475–486.
29. Hayes, C.A., Karlsson, N.G., Struwe, W.B., Lisacek, F., Rudd, P.M., Packer, N.H. and Campbell, M.P. (2011) UniCarb-DB: a database resource for glycomics discovery. *Bioinformatics*, **27**, 1343–1344.
30. Patient, S., Wieser, D., Kleen, M., Kretschmann, E., Jesus Martin, M. and Apweiler, R. (2008) UniProtJAPI: a remote API for accessing UniProt data. *Bioinformatics*, **24**, 1321–1322.
31. Shakhshere, B., Anderson, M., Khatib, K., Tadoori, L., Joshi, L., Lisacek, F., Hirschman, L. and Mullen, E. (2013) SugarBind database (SugarBindDB): a resource of pathogen lectins and corresponding glycan targets. *J. Mol. Recognit.*, **26**, 426–431.
32. Campbell, M.P., Ranzinger, R., Lutteke, T., Mariethoz, J., Hayes, C.A., Zhang, J., Akune, Y., Aoki-Kinoshita, K.F., Damerell, D., Carta, G. *et al.* Toolboxes for a standardised and systematic study of glycans. *BMC Bioinformatics*.