Using Databases and Web Resources for Glycomics Research

Kiyoko F. Aoki-Kinoshita*†

Many databases of carbohydrate structures and related information can be found on the World Wide Web. This review covers the major carbohydrate databases that have potential utility for glycoscientists and researchers entering the glycosciences. The first half provides a brief overview of carbohydrate databases and web resources (including a history of carbohydrate databases and carbohydrate notations used in these databases), and the second half provides a guide that can be used as an index to determine which resources provide the data of most interest to the user. *Molecular & Cellular Proteomics* 12: 10.1074/mcp.R112.026252, 1036–1045, 2013.

It was not until the 1990s, when CarbBank was developed, that a carbohydrate structure database was made available to the public. When the CarbBank project ended, many researchers found the need to somehow continue the development of carbohydrate-related databases, and so GLYCOSCIENCES.de, KEGG GLYCAN, and the Consortium for Functional Glycomics databases emerged. Since the development of these databases, many more carbohydrate-related web resources have been developed, to the point where it is becoming difficult to keep track of them all. Moreover, the use of different carbohydrate structure notations to represent the data in each database makes it difficult for users to utilize them all effectively. Therefore, in addition to a description of various carbohydrate-related databases and web resources, we also present briefly in this review each carbohydrate structure representation, along with suggestions of tools that can be used to convert one to another.

In the main section of this review, we summarize many of the well-known carbohydrate-related databases and web resources that are now publicly available. Because many of these resources have already been published in detail elsewhere, only a summary is presented here. The final section presents carbohydrate-related data from the researcher’s perspective; that is, various categories of data are listed (e.g., three-dimensional structures, taxonomy, experimental data, etc.), and the pertinent databases and web resources are listed within each category such that they can be assessed and compared with one another.

**History: CarbBank**—The very first carbohydrate database was called CCSD, for Complex Carbohydrate Structure Database (1, 2). However, because of the tool used to access this database, called CarbBank, it is more commonly known as CarbBank. This database was initially developed by the Complex Carbohydrate Research Center at the University of Georgia. It was the first attempt to accumulate such data from the research community, and it thus became a valuable resource for current carbohydrate databases. Unfortunately, funding for CCSD ceased in 1996, but the data, consisting of almost 45,000 entries of glycan structures, annotation, and literature information, were still made available to the public.

There has been some criticism of the quality of the data in CarbBank (3). This can be attributed to the fact that there was no curation of the accumulated data, which resulted in many inconsistencies in terms of duplicated carbohydrate structures and inconsistent notations to represent carbohydrates (or, more specifically, monosaccharides) in the database. However, the same can be said of GenBank (4), and users should be aware of the potential errors in all publicly available databases.

**History: Carbohydrate Structure Notation**—As many glycan structure databases were developed in the absence of a standard notation for representing glycan structures, many database providers developed their own notations. A list of these formats and corresponding examples for representing the N-linked glycan core structure are given in Table I. Detailed explanations of these formats have been presented elsewhere (5), but one can see from this table that there are a wide variety of carbohydrate structure formats now in use, each with its own advantages and disadvantages. Although many databases provide different ways to input glycan structures as queries using innovative user interfaces, not all databases return the results of these queries in multiple formats. That is, a query might be returned in a single format that will need to be converted to another format in order for it to be used as a query for another database, for example.

Therefore, it has become necessary to be able to convert a glycan structure from one format to another. The RINGS resource (6) of data mining and analytical tools for glycomics analysis provides a set of web-based utilities for structure...
| Name                     | Example                                                                 |
|--------------------------|--------------------------------------------------------------------------|
| CarbBank                 | a-D-Manp-(1→6)+                                                         |
|                          | b-D-Manp-(1→4)-b-D-GlcNAc-(1→4)-b-D-GlcNAc                               |
|                          | a-D-Manp-(1→3)+                                                         |
| IUPAC                    | Man a1–3(Man a1–6) Man b1–4 GlcNAc b1–4 GlcNAc b                           |
| LINUCS                   | (1b-D-GlcNAc)(1[(4 + 1)]b-D-GlcNAc)(1[(4 + 1)]b-D-Manp)(1[(3 + 1)]a-D-Manp)(1[(6 + 1)]a-D-Manp)) |
| GlycoMinds LinearCode    | Ma3(Ma6)Mb4GNb4GNb                                                       |
| KCF                      | NODE 5                                                                  |
|                          | | Glycan | 1 | GlcNAc | 10.0 | 7.0 |
|                          | |       | 2 | GlcNAc | 3.0  | 7.0 |
|                          | |       | 3 | Man    | −4.0 | 7.0 |
|                          | |       | 4 | Man    | −11.0| 12.0|
|                          | |       | 5 | Man    | −11.0| 2.0 |
| EDGE                    | 4                                                                        |
|                          | 1 | 2:b1 | 1:4 |
|                          | 2 | 3:b1 | 2:4 |
|                          | 3 | 5:a1 | 3:3 |
|                          | 4 | 4:a1 | 3:6 |
| GlycoCTcondensed         | RES                                                                     |
|                          | 1b:b-dglc-HEX-1:5                                                       |
|                          | 2s:n-acetyl                                                             |
|                          | 3b:b-dglc-HEX-1:5                                                       |
|                          | 4s:n-acetyl                                                             |
|                          | 5b:b-dman-HEX-1:5                                                       |
|                          | 6b:a-dman-HEX-1:5                                                       |
|                          | 7b:a-dman-HEX-1:5                                                       |
| LIN                      | 1:1d(2 + 1)2n                                                           |
|                          | 2:1o(4 + 1)3d                                                           |
|                          | 3:3d(2 + 1)4n                                                           |
|                          | 4:3o(4 + 1)5d                                                           |
|                          | 5:5o(3 + 1)6d                                                           |
|                          | 6:5o(6 + 1)7d                                                           |
| GlycoCTXML               | <?xml version = "1.0" encoding = "UTF-8"?>                                |
|                          | <sugar version = "1.0">                                                |
|                          | <residues>:                                                             |
|                          | <basetype id = "1" anomer = "b" superclass = "hex" ringStart = "1" ringEnd = "5" name = |
|                          | "b-dglc-HEX-1:5">                                                      |
|                          | <stemtype id = "1" type = "dglc" />                                    |
|                          | </basetype>                                                            |
|                          | <substituent id = "2" name = "n-acetyl" />                              |
|                          | <basetype id = "3" anomer = "b" superclass = "hex" ringStart = "1" ringEnd = "5" name = |
|                          | "b-dglc-HEX-1:5">                                                      |
|                          | <stemtype id = "1" type = "dglc" />                                    |
|                          | </basetype>                                                            |
|                          | <substituent id = "4" name = "n-acetyl" />                              |
|                          | <basetype id = "5" anomer = "b" superclass = "hex" ringStart = "1" ringEnd = "5" name = |
|                          | "b-dman-HEX-1:5">                                                     |
|                          | <stemtype id = "1" type = "dman" />                                   |
|                          | </basetype>                                                            |
|                          | <basetype id = "6" anomer = "a" superclass = "hex" ringStart = "1" ringEnd = "5" name = |
|                          | "a-dman-HEX-1:5">                                                     |
|                          | <stemtype id = "1" type = "dman" />                                   |
|                          | </basetype>                                                            |
|                          | <basetype id = "7" anomer = "a" superclass = "hex" ringStart = "1" ringEnd = "5" name = |
|                          | "a-dman-HEX-1:5">                                                     |
|                          | </basetype>                                                            |

Using Databases and Web Resources for Glycomics Research
Using Databases and Web Resources for Glycomics Research

| Name | Example |
|------|---------|
| <stemtype id = "1" type = "dman" /> | |
| <basetype> | |
| <residues> | |
| <linkages> | |
| <connection id = "1" parent = "1" child = "2"> | |
| <linkage id = "1" parentType = "d" childType = "n"> | |
| <parent pos = "2" /> | |
| <child pos = "1" /> | |
| </linkage> | |
| </connection> | |
| <connection id = "2" parent = "1" child = "3"> | |
| <linkage id = "2" parentType = "o" childType = "d"> | |
| <parent pos = "4" /> | |
| <child pos = "1" /> | |
| </linkage> | |
| </connection> | |
| <connection id = "3" parent = "3" child = "4"> | |
| <linkage id = "3" parentType = "d" childType = "n"> | |
| <parent pos = "2" /> | |
| <child pos = "1" /> | |
| </linkage> | |
| </connection> | |
| <connection id = "4" parent = "3" child = "5"> | |
| <linkage id = "4" parentType = "o" childType = "d"> | |
| <parent pos = "4" /> | |
| <child pos = "1" /> | |
| </linkage> | |
| </connection> | |
| <connection id = "5" parent = "5" child = "6"> | |
| <linkage id = "5" parentType = "o" childType = "d"> | |
| <parent pos = "3" /> | |
| <child pos = "1" /> | |
| </linkage> | |
| </connection> | |
| <connection id = "6" parent = "5" child = "7"> | |
| <linkage id = "6" parentType = "o" childType = "d"> | |
| <parent pos = "6" /> | |
| <child pos = "1" /> | |
| </linkage> | |
| </connection> | |
| <linkages> | |
| </sugar> | |

GLYDE-II

<?xml version = "1.0" encoding = "UTF-8"?>
< Glydell >
< molecule subtype = "glycan" id = "From_GlycoCT_Translation" >
< residue subtype = "base_type" partid = "1" ref = "http://www.monosaccharideDB.org/ GLYDE-II.jsp?G = b-dglc-HEX-1:5" /> |
< residue subtype = "substituent" partid = "2" ref = "http://www.monosaccharideDB.org/ GLYDE-II.jsp?G = n-acetyl" /> |
< residue subtype = "base_type" partid = "3" ref = "http://www.monosaccharideDB.org/ GLYDE-II.jsp?G = b-dglc-HEX-1:5" /> |
< residue subtype = "substituent" partid = "4" ref = "http://www.monosaccharideDB.org/ GLYDE-II.jsp?G = n-acetyl" /> |
< residue subtype = "base_type" partid = "5" ref = "http://www.monosaccharideDB.org/ GLYDE-II.jsp?G = b-dman-HEX-1:5" /> |
< residue subtype = "base_type" partid = "6" ref = "http://www.monosaccharideDB.org/ GLYDE-II.jsp?G = a-dman-HEX-1:5" /> |
< residue subtype = "base_type" partid = "7" ref = "http://www.monosaccharideDB.org/ GLYDE-II.jsp?G = a-dman-HEX-1:5" /> |
< residue_link from = "2" to = "1" >
conversion between many of these formats. Also, the GlycanBuilder tool (7, 8), developed as a part of the EuroCarbDB project (9), provides a graphical user interface through which carbohydrate structures can be imported in one format and then exported in another. In the future, however, it is expected that databases will be able to mutually exchange or link their data to one another, such that such format conversions will be unnecessary. Moreover, glycan structure analysis tools should provide ways to input glycan structures in any of the available formats, such that users need not take heed of such specifics.

Summary of Currently Available Glycomic Databases and Web Resources

GLYCOSCIENCES.de—GLYCOSCIENCES.de is one of the oldest web portals for glycomics research; it was originally developed at the German Cancer Research Center and now is maintained by Justus-Liebig University Giessen (10). It includes glycan structure references to CarbBank, automatically generated three-dimensional coordinates, 1H and 13C NMR spectra including 1H and 13C NMR-shift lists, masses of theoretically calculated glycan fragments, ligand data from the Protein Data Bank (PDB), and three-dimensional conformational maps for glyclosidic linkages. Recently, a Glyco-CD database has also been incorporated that consists of cluster-of-differentiation antigens to aid in the classification of various cell surface macromolecules.

In addition to databases, GLYCOSCIENCES.de also provides a number of tools to aid in carbohydrate structure analysis. The available tools are categorized into three major groups: three-dimensional-structure-related tools, tools related to structure representation, and mass spectrometry tools. The three-dimensional-structure-related tools are further categorized into the following three groups: detection/validation of carbohydrates in PDB files, statistical analysis of carbohydrates in PDB, and modeling. Table II provides references and brief descriptions of each tool.

KEGG GLYCAN—KEGG GLYCAN is a part of the Kyoto Encyclopedia of Genes and Genomes (KEGG) resource developed by Kanehisa Laboratories in Japan. Details regarding each of the data sets available in KEGG GLYCAN are available elsewhere (11); a brief summary is described in this section. KEGG GLYCAN contains glycan structures extracted from CarbBank and manually curated glycan structures from the literature. The Composite Structure Map (12) was built from these structures to provide an overview of the structures in the database and the relationships among them. This map provides links to the glyco-gene and to glycan structure data in KEGG.

Many of the structures in KEGG GLYCAN are linked to KEGG PATHWAY, which includes many reference pathways for glycan biosynthesis and metabolism. Genomic information regarding glycosyltransferases also has been summarized in the KEGG Orthology system and classified in KEGG BRITE. The KEGG Orthology system groups orthologous groups of genes according to sequence and function, and KEGG BRITE is a hierarchically organized database of information in KEGG. Moreover, a number of tools for glycan structure similarity searches (13), a two-dimensional glycan drawing and search
tool (KegDraw, available for download from the KEGG website), and tools for linking genomic/transcriptomic data of glycosyltransferases to glycan structures (14, 15) are available.

CFG—The Consortium for Functional Glycomics (CFG) was a large international research initiative started in 2001 with funding from a National Institute of General Medical Sciences Large-Scale Collaborative Project award. The CFG database (16) consists of glycan structures, glycan-binding proteins, and glycosyltransferase data (in their CFG Molecule Pages) that are supplemented by the CFG-derived information from their glycan arrays, glycan profiling data from MALDI-MS experiments, gene microarrays of glyco-gene expression, and mouse phenotyping data (CFG Data). The glycan structures in CFG have been accumulated from CarbBank, a database developed by GlycoMinds Ltd., and from their own CFG Data resources.

The Glycan Binding Protein data resource provides detailed information about glycan-binding proteins (GBPs), including DNA and protein sequences, binding specificities, biological functions, etc. Experts in various fields of GBP research have provided the summary information and references for each GBP. The Glycosyltransferase resource serves as a portal to relevant CFG data and other glycan-related databases. Each enzyme is categorized according to the carbohydrate structure with which it is involved. A graphic display of a combined glycan structure is also provided so that users can click directly on the glycosidic linkage to which the enzyme of interest is related.

The glycan array data resource provides all the results that have been obtained by the CFG from screening samples for glycan binding specificity (17). A large variety of GBPs such as lectins, antibodies, pathogens, and cells have been screened, and each experimental result can be viewed graphically in a web browser. The glycan profiling resource provides all the results that have been obtained by the CFG from MALDI-MS and other analyses to identify and characterize the glycans in human and mouse tissues and cells (18, 19). Each of the GBPs and glycans in the results from these experiments is also linked to other data resources and external databases.

The microarray data resource provides gene expression data from experiments using human and mouse samples on the glycosene microarray chip developed by the CFG (20), as well as some other commercial gene chips. Each experimental result has been processed to provide signals, present/absent calls, and p values for each gene. The mouse phenotyping resource (21) provides the data obtained from knockout gene experiments on a variety of mouse strains. The details regarding each experimental protocol, the raw data obtained, and processed data, including a summary of the experimental results, are all provided.

The CFG also provides Paradigm Pages, which is presented in a wiki format and describes detailed information about exemplary GBPs that are considered “paradigm GBPs.” CFG investigators volunteer to make contributions to these pages, which can also be updated as needed.

JCGGDB—The Japan Consortium for Glycobiology and Glycotechnology Database (JCGGDB) is a database of glycoscience data accumulated in Japan. It includes an integrated search function for many glycoscience databases throughout Japan. Table III lists the representative databases that are accessible from JCGGDB. In addition to a keyword search function that works across all the integrated databases, JCGGDB also provides many data resources of interest to glycobiologists, such as glycan-related diseases and experimental protocols. Many data have also been accumulated from the large amount of experimental data obtained at the National Institute of Advanced Industrial Science and Technology, which houses JCGGDB. These include mass spectrometry data, lectin affinity data, glycoprotein data, and glyco-gene information, as described in Table III.

BCSDB—The Bacterial Carbohydrate Structure Data Base (BCSDB) (22) is a database of carbohydrate structures found in bacteria or obtained via the modification of those found in almost 5000 bacterial organisms, curated from the literature. In addition to structural information, each entry provides information about references, biological sources, keywords, methods used to elucidate the structure, bioactivity, NMR assignment tables, etc.

GlycomeDB—GlycomeDB (23) is a web portal for all of the major glycan structure databases described above, as well as some others. Specifically, GlycomeDB has integrated glycan structures from the following databases: CarbBank, GLYCOSCIENCES.de, KEGG GLYCAN, CFG, JCGGDB, BCSDB, GlycoBase (Dublin), GlycoBase (Lille), GlyAffinity, and EuroCarbDB. Much effort went into ensuring that a consistent namespace was used in mapping structures between different databases. GlycoCT is used as the central carbohydrate representation format for integrating the structures in these databases and storing them in GlycomeDB. In addition to the glycan structures and links to their original database sources, taxonomic information is also included in each record.

GlycoSuiteDB—GlycoSuiteDB (24) is an annotated and curated database of glycan structures taken from the scientific literature published between 1999 and 2005. Originally developed commercially, it has since been made available publicly through the ExPASy server. For each glycan structure, detailed information is provided on native and recombinant sources (i.e., tissue and/or cell type, cell line, strain, and disease state), and links to Swiss-Prot/TrEMBL are provided when available. Search functions for the following items are also provided:

- Mass or mass range.
- Attached protein by name, keyword, or Swiss-Prot/TrEMBL accession number.
- Taxonomy selected from a list or by keyword.
- Composition of monosaccharides and substituents.
- Tissue or cell type selected from a list or by keyword.
### TABLE II
**Table of tools available in GLYCOSCIENCES.de**

| Category                              | Sub-category                              | Tool name  | Description                                                                                                                                                                                                 |
|---------------------------------------|-------------------------------------------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Three-dimensional-structure related   | Detection and validation of PDB files     | pdb-care (31) | Given a PDB ID or file containing carbohydrate information, this tool verifies the accuracy of the carbohydrate conformations, including bond length, valency, and carbohydrate residue nomenclature. |
|                                       |                                           | pdb2linucs (32) | Given a PDB ID or file containing carbohydrate information, this tool extracts the carbohydrate structures and returns them in LINUCS or IUPAC format.                                                     |
|                                       |                                           | CARP (33)    | Given a PDB ID or file containing carbohydrate information, this tool finds each glycosidic linkage and generates a Ramachandran plot of the backbone torsion angles that can be compared with existing knowledge as found in PDB or GlycoMapsDB. |
| Statistical analysis of carbohydrate properties in PDB | GlyVicinity (33) |              | Plots of the frequency of appearance of amino acids found surrounding particular carbohydrate residues can be generated from this tool, based on the latest PDB data.                                  |
|                                       | GlyTorsion (33) |              | Plots of the frequency of ranges of torsion angles of carbohydrate components, as found in the latest PDB data, can be generated.                                                                      |
|                                       | GlySeq (33)     |              | Plots of the frequency of amino acid compositions around glycosylation sites can be generated, based on the data derived from PDB or SwissProt.                                                      |
| Modeling                              | Sweet-II       |              | When the user inputs a glycan structure using the customized input form, three-dimensional structures of the glycan can be generated in a variety of formats, such as a PDB file, using JMol, VRML, Tinker, and Babel. |
|                                       | GlyProt (34)    |              | Given a PDB ID or file, potential glycosylation sites, their accessibility, and an in silico generation of the glycosylated protein in three dimensions can be generated.                                      |
|                                       | GlycoMapsDB (35) |            | This is a database of pre-calculated conformational maps of various oligosaccharides found in N- and O-linked glycans.                                                                                  |
| Carbohydrate-notation related         | LINUCS (36)     |              | Given a carbohydrate structure in CarbBank format, this tool generates the LINUCS code format.                                                                                                          |
|                                       | LiGraph        |              | Given a carbohydrate structure in CarbBank format, this tool generates two-dimensional images of the structure.                                                                                          |
|                                       | sumo           |              | Given a carbohydrate structure in LINUCS or IUPAC format, this tool extracts known structural motifs, such as Lewis antigens or core structures.                                                      |
| Mass spectrometry                     | Glycofragment  |              | Given a carbohydrate structure in CarbBank format, this tool attempts to find the fragments that can be expected to occur in MS spectra for the structure.                                               |

- Glycosidic linkage (*i.e.* N-linked, O-linked, or C-linked) and/or reducing-terminal sugar selected from a list.
- Disease selected from a list or by keyword.
- Literary reference by author name(s) or publication year(s).
- GlycoSuiteDB accession number.
- Structure using their own structure-drawing tool.
UniCarbKB—UniCarbKB (25) is an initiative that extends from the EuroCarbDB project (9), which was a European Union–funded initiative for the development of a framework for storing carbohydrate structures and their experimental evidence, namely, mass spectrometry, HPLC, and NMR data. UniCarbKB aims to further integrate annotations and adopt common standards to provide a knowledgebase for glycomics data. UniCarbKB includes UniCarb-DB (26), a platform for structural and analytical storage and the retrieval of data from glycomics experiments. The glycan structures in UniCarbKB are generally derived from GlycoSuiteDB, and all structures are linked to experimental evidence, which includes annotations regarding experimental conditions, biological sources, etc.

MonosaccharideDB—MonosaccharideDB was developed alongside GlycomeDB and the EuroCarbDB project as a central resource for monosaccharide information. This database is crucial in that it allows scientists to obtain mappings between different representations of monosaccharides, which might be defined differently depending on the research field. For example, N-acetylgalactosamine can be written as GlcNAc or β-D-GlcpNAc, or even as 2-acetamido-2-deoxy-β-D-glucopyranose. This database ensures that all of these representations refer to the same monosaccharide. Moreover, this database provides web services and on-the-fly web interfaces to make the data available as efficiently and accurately as possible. In addition to the major glycan structure representations listed in Table I, MonosaccharideDB also returns graphic notations, including three-dimensional representations, of monosaccharides in the database.

PolySac3DB—PolySac3DB is a database of three-dimensional structures of polysaccharides (27) accumulated from the scientific literature. The data are provided in a hierarchical manner in categories representing polysaccharide families, including agaroses, amyloses, celluloses, chitins, glycosaminoglycans, pectins, xylans, etc. Each record is annotated regarding the way in which the structure was obtained, references, the three-dimensional representation of the repeating unit, diffraction diagrams, etc. Information regarding polysaccharide structure determination techniques (x-ray crystallography, electron and neutron diffraction, and molecular modeling) is also provided for beginners in the field.

Lectin3D—Lectin3D, previously known as Lectines, is a database of three-dimensional structures of lectins categorized into different lectin families, including plant lectins, animal lectins, and viral lectins. Each record is annotated with the PDB identification, resolution, species, references, etc. Detailed structural data can also be viewed as images using Jmol or downloaded as PDB files.

Table III: The representative data resources that are accessible from JCGGDB

| Name                                      | Description                                                                 |
|-------------------------------------------|-----------------------------------------------------------------------------|
| GGDB (GlycoGene DataBase)                 | GlycoGene is a database that includes genes associated with glycan synthesis such as glycosyltransferases, sugar nucleotide synthases, sugar-nucleotide transporters, and sulfotransferases. All of the nearly 200 human glycoenzymes have been identified, cloned, and characterized (28). It also includes substrate specificity information. |
| LfDB (Lectin Frontier DataBase)            | LfDB provides basic information on lectins, as well as interaction data obtained via the frontal affinity chromatography–fluorescence detection system (37, 38). |
| GlycoProtDB (GlycoProtein Database)       | GlycoProtDB is a database of N-glycoproteins that have been experimentally identified in C. elegans N2 and mouse tissues (strain C52BL/6J, male). |
| GMDB (Glycan Mass Spectral DataBase)       | GMDB currently stores MS2, MS3, and MS4 spectra of N- and O-linked glycans and glycolipid glycans and their fragments (40). |
| LipidBank                                 | LipidBank (41) is a freely available database of natural lipids including fatty acids, glycerolipids, sphingolipids, steroids, and various vitamins. It is the official database of the Japanese Conference on the Biochemistry of Lipids. |
| GlycoEpitope                              | The GlycoEpitope database provides information on carbohydrate antigens such as glycoproteins that express carbohydrate antigens, glycolipids of which the partial structure is a carbohydrate epitope, enzymes that take part in the synthesis and degradation of glyceoptopes, the time and site of expression of carbohydrate epitopes, diseases to which carbohydrate epitopes are related, etc. |
| GALAXY (Glycoanalysis by the three axes of MS and chromatography) | GALAXY contains data on approximately 500 different pyridylamino-glycans, including the structures, HPLC elution positions expressed in glucose units on ODS and amide-silica columns, relative molecular mass, code numbers, sources of samples, and references (42). |
| GlycoPOD (GlycoScience Protocol Online Database) | GlycoPOD is a collection of experimental protocols used for glycomics research. Each protocol has been written by experts in glycoscience. |
Using Databases and Web Resources for Glycomics Research

(a) JCGGDB provides a list of all experimentally verified glyco-genes in humans (28). Each record is annotated with literary references and provides a graphical view of substrate specificity.

(b) KEGG provides a list of glycosyltransferases and other glyco-enzymes in their KEGG ORTHOLOGY and KEGG BRITE data resources (29).

(c) The CFG provides not only genomic information of glyco-genes, but also expression data as obtained from their microarray experiments.

Pathways—
(a) The KEGG resource provides the basic biosynthesis pathways for glycan structures, including N-glycans, O-glycans, glycosaminoglycans, glycosylphosphatidylinositol anchors, glycolipids, lipopolysaccharides, and peptidoglycans. KEGG PATHWAY also provides information on glycan-relevant pathways such as signaling and interactions, cell communication, and the immune system.

Proteomics Data—
GBPs—
(a) The CFG provides Molecule Pages for GBPs that include genomic, proteomic, and glycomic details for each record and are annotated by experts in the field. Links to CFG resource data (glycan arrays and profiling data) are also provided where available.

(b) The Lectin Frontier DataBase, a part of JCGGDB, provides structural, taxonomic, literary, protein sequence, and glycan-binding affinity data for a variety of lectins.

(c) Lectin3D provides structural, taxonomic, and literary information regarding lectins for which structures are known (as found in PDB). PDB files can be downloaded and structures can be visually modified using Jmol at this web site.

Glycoproteins—
(a) UniProt is a major protein database that also includes glycosylation-site annotations. These annotations are marked as predicted from computer simulations or experimentally verified.

(b) GlycoProtDB, a part of JCGGDB, provides N-linked glycoprotein data from C. elegans and mouse tissues. Both predicted and experimentally confirmed glycosylation-site information is provided, along with links to external databases such as GenBank and Swiss-Prot where available.

(c) GlycoSuiteDB provides information about proteins known to be glycosylated as confirmed in the literature. Experimental methods, taxonomic and biological source information, references, and glycan structures are provided for each record.

Glycan Structure Profiling Data—
(a) UniCarb-DB, a part of UniCarbKB, provides mass spectrometric data and structural assignments based on fragmentation data. Glycan structures can be queried individually, and the results include all experimental evidence for the selected structure. Biological source information, experimental conditions, and annotated peak lists are also provided.

(b) GlycoSuiteDB provides a search function for queries by species or class by listing all available species or classes in the database, from which the user can select his or her item of interest. It is also possible to enter a (partial) keyword, which can be a species name (e.g. Homo sapiens), a common name (e.g. human), or a class.

Experimental Data—
Protocols—GlycoPOD, a part of JCGGDB, provides an organized classification of experimental protocols for glycomics experiments. Experts selected by the steering committee of this database provide the details for each protocol, which are also checked by the committee for consistency.

Raw Experimental Data—
(a) UniCarb-DB, a part of UniCarbKB, provides mass spectrometric data and structural assignments based on fragmentation data. Glycan structures can be queried individually, and the results include all experimental evidence for the selected structure. Biological source information, experimental conditions, and annotated peak lists are also provided.

(b) GlycoSuiteDB provides the data from glycan array, glycan profiling, glyco-gene microarray experiments, and knockout mouse experiments are all provided in the CFG Data section. Processed and summarized data, as well as the original raw data files, can be downloaded.

Pathological data—
(a) KEGG includes a resource called KEGG DISEASE and DRUG that might contain links, albeit indirect, to KEGG GLYCAN records.

(b) JCGGDB provides a listing of glycan-related diseases in their newly developed Glyco-Disease Genes Database and Tumor Markers Reference Database. Each database has been curated from the literature and provides references for each record.

Structural Data—
Three-dimensional Data—
Using Databases and Web Resources for Glycomics Research

(a) GLYCOSCIENCES.de provides a variety of modeling tools for carbohydrates, and with them computationally predicted three-dimensional structures of glycoproteins and glycans can be generated.

(b) MonosaccharideDB contains three-dimensional structures of monosaccharides that can be downloaded in PDB format.

(c) PolySac3DB is a database of three-dimensional polysaccharide structures extracted from the literature. Educational information regarding polysaccharides and experimental procedures is also provided.

(d) Lectin3D is a database of three-dimensional lectin structures as found in PDB.

Epitopes—
(a) GlycoSuiteDB provides a search function for glycan epitopes.

(b) GlycoEpitopeDB is a database of carbohydrate epitopes and antibodies that has been manually curated from the literature.

Mass Data—
(a) GLYCOSCIENCES.de provides a search function that allows the user to search for carbohydrates using a list of mass peaks.

(b) KEGG GLYCAN includes mass information for each glycan structure, computed from the composition.

(c) GlycoSuiteDB provides a search function allowing the user to search for glycans with a particular isotopic mass, a list of masses, or a range of masses.

(d) UniCarb-DB provides a search function allowing the user to search the experimental data by mass.

NMR Signal Data—
(a) GLYCOSCIENCES.de and BCSDB both provide search functions allowing the user to search NMR data based on atoms, peaks, and chemical shifts.

Summary—In addition to the databases and web resources described in this review, new resources are constantly being developed, and it would be impossible to comprehensively cover them all. This review focused on those that are well known in the glycoscience community and provide sufficient annotation information for researchers to find their data of interest. Because of the variety of tools and interfaces that are provided by each resource, this review also summarized these resources by data type in the hopes of helping the reader determine which is the most pertinent resource for his or her research. It is expected that in the future, these resources will become better integrated with one another so that a consistent interface can be provided, allowing users to most efficiently take advantage of the valuable knowledge that can be gained from them.

‡ To whom correspondence should be addressed: Tel./Fax: +81-42-691-4116; E-mail: kklyoko@soka.ac.jp.

REFERENCES
1. Doubet, S., Bock, K., Smith, D., Darvill, A., and Albersheim, P. (1989) The Complex Carbohydrate Structure Database. Trends Biochem. Sci. 14, 475–477
2. Doubet, S., and Albersheim, P. (1992) CarbBank. Glycobiology 2, 505
3. Egorova, K. V., and Toukach, P. V. (2012) Critical analysis of CCSD data quality. J. Chem. Inf. Model 52, 2812–2814
4. Harris, D. J. (2003) Can you bank on GenBank? Trends Ecol. Evol. 18, 317–319
5. Aoki-Kinoshita, K. F. (2010) Glycome Informatics: Methods and Applications. CRC Press, Boca Raton, FL
6. Akune, Y., Hosoda, M., Kalya, 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
7. Damerell, D., Ceroni, A., Maass, K., Ranzheimer, D., Dell, A., and Haslam, S. M. (2012) The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. Source Code Biol. Med. 7, 3
8. von der Lieth, C. W., Freire, A. A., Blank, D., Campbell, M. P., Ceroni, A., Damerell, D. R., Dell, A., Dwok, R. A., Ernst, B., Fogg, R., Frank, M., Geyer, H., Geyer, R., Harrison, M. J., Henrick, K., Herget, S., Hull, W. E., Ionides, J., Joshi, H. J., Kamerling, J. P., Leeflang, B. R., Lütteke, T., Lundborg, M., Maass, K., Merry, A., Ranzheimer, R., Rosen, J., Royle, L., Rudd, P. M., Schlossing, S., Stenutz, R., Vranken, W. F., Widmalm, G., and Haslam, S. M. (2011) EUROCARB: an open-access platform for glycominformatics. Glycobiology 21, 493–502
9. Lütteke, T., Bohne-Lang, A., Loss, A., Goet, T., Frank, M., and von der Lieth, C. (2006) GLYCOSCIENCES.de: an internet portal to support glycomics and glycomics research. Glycobiology 16, 719–811
10. Hashimoto, K., Kawano, S., Goto, S., Aoki-Kinoshita, K., Kawashima, M., and Kanehisa, M. (2005) A global representation of the carbohydrate structures: a tool for the analysis of glycans. Genome Inform. 16, 214–222
11. Aoki-Kinoshita, K. F., and Kanehisa, M. (2010) Bioinformatics analysis of glycan structures from a genomic perspective. In Bioinformatics for Glycobiology and Glycomics: An Introduction (von der Lieth, C.-W., Lütteke, T., and Frank, M., eds.) Wiley, Chichester, West Sussex, UK., pp. 125–141
12. Hashimoto, K., Yamaguchi, A., Ueda, N., Akutsu, T., Mamitsuka, H., Goto, S., and Kanehisa, M. (2004) KCaM (KEGG Carbohydrate Matcher): a software tool for analyzing the structures of carbohydrate sugar chains. Nucleic Acids Res. 32, W267–W272
13. Sugaya, A., Yamanishi, Y., Hashimoto, K., Goto, S., and Kanehisa, M. (2007) An improved scoring scheme for predicting glycan structures from gene expression data. Genome Inform. 18, 237–246
14. Kawano, S., Hashimoto, K., Miyama, T., Goto, S., and Kanehisa, M. (2005) Prediction of glycan structures from gene expression data based on glycosyltransferase reactions. Bioinformatics 21, 3976–3982
15. Raman, R., Venkataraman, M., Ramakrishnan, S., Lang, W., Kaguram, S., and Sasisekharan, R. (2006) Advancing glycomics: implementation strategies at the Consortium for Functional Glycomics. Glycobiology 16, 829–90R
16. Bleix, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., Bryan, M. C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D. J., Skelgel, J. J., van Dier, I., Burton, D. R., Wilson, I. A., Cummings, R., Bovin, N., Wong, C. H., and Paulson, J. C. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. Proc. Natl. Acad. Sci. U.S.A. 101, 17033–17038
17. North, S. J., Hitchon, P. G., Haslam, S. M., and Dell, A. (2009) Mass spectrometry in the analysis of N-linked and O-linked glycans. Curr. Opin. Struct. Biol. 19, 498–506
18. Jang-Gee, J., North, S. J., Sutton-Smith, M., Goldberg, D., Panico, M., Morris, H., Haslam, S., and Dell, A. (2006) Glycomic profiling of cells and tissues by mass spectrometry: fingerprinting and sequencing methodologies. Methods Enzymol. 415, 59–86
19. Comelli, E. M., Head, S. R., Gilmartin, T., Whisnant, T., Haslam, S. M., North, S. J., Wong, N. K., Kudo, T., Narimatsu, H., Esko, J. D., Drickamer, K., Dell, A., and Paulson, J. C. (2006) A focused microarray approach to functional glycomics: transcriptional regulation of the glycome. Glycobiology 16, 117–131
Orr, S. L., Le, D., Long, J. M., Sobiesczuk, P., Ma, B., Tian, H., Fang, X., Paulson, J. C., Marth, J. D., and Varki, N. (2012) A phenotype survey of thirty-six mutant mouse strains with gene targeted defects in glycosyltransferases or glycan-binding proteins. *Glycobiology** 23, 363–380

22. Toukach, P., Joshi, H., Ranzinger, R., Knirel, Y., and von der Lieth, C. (2007) Sharing of worldwide distributed carbohydrate-related digital resources: online connection of the Bacterial Carbohydrate Structure Database and GLYCOSCIENCES.de. *Nucleic Acids Res.* 35, D280-D286

23. Ranzinger, R., Herget, S., von der Lieth, C. W., and Frank, M. (2011) GlycomeDB—a unified database for carbohydrate structures. *Nucleic Acids Res.* 39, D373-D376

24. 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., and Packer, N. H. (2011) GlycoSuiteDB: a curated relational database of glycoprotein glycan structures and their biological sources. 2011 update. *Nucleic Acids Res.* 31, 511–513

25. Sarkar, A., and Pérez, S. (2012) PolySac3DB: an annotated data base of 3 dimensional structures of polysaccharides. *BMC Bioinformatics* 13, 302

26. Narimatsu, H. (2004) Construction of a human glycogene library and comprehensive functional analysis. *Glycoconj. J.* 21, 17–24

27. Kameyama, A., Kikuchi, N., Nakaya, S., Ito, H., Sato, T., Shikanai, T., Takahashi, Y., Takahashi, K., and Narimatsu, H. (2005) A strategy for identification of oligosaccharide structures using observational multi-stage mass spectral library. *Anal. Chem.* 77, 4719–4725

28. Tomiya, N., Awaya, J., Kurono, M., Endo, S., Arata, Y., and Takahashi, N. (1988) Analyses of N-linked oligosaccharides using a two-dimensional mapping technique. *Anal. Biochem.* 171, 73–90

Using Databases and Web Resources for Glycomics Research