Chapter 7
Glyco3D: A Suite of Interlinked Databases of 3D Structures of Complex Carbohydrates, Lectins, Antibodies, and Glycosyltransferases

Serge Pérez, Anita Sarkar, Alain Rivet, Sophie Drouillard, Christelle Breton, and Anne Imberty

Abstract Glyco3D is a portal for structural glycobiology of several interlinked databases that is covering the three-dimensional features of monosaccharides, disaccharides, oligosaccharides, polysaccharides, glycosyltransferases, lectins, monoclonal antibodies, and glycosaminoglycan-binding proteins. Collection of annotated NMR data of bioactive oligosaccharides is also available. A common nomenclature has been adopted for the structural encoding of the carbohydrates. Each individual database stands by itself as it covers a particular family of either complex carbohydrates or carbohydrate-binding proteins. A unique search engine is available that scans the full content of all the databases for queries related to sequential information of the carbohydrates. The interconnection of these databases provides a unique opportunity to characterize the three-dimensional features that a given oligosaccharide molecule can take in different environments, i.e., vacuum, crystalline state, or interacting with different proteins having different biological function. The databases, which have been manually curated, were developed with nonproprietary software. They are web-based platform and are freely available to the scientific community at http://glyco3d.cermav.cnrs.fr.

Keywords Monosaccharides • Disaccharides • Oligosaccharides • Polysaccharides • Glycosyltransferases • Lectins • Antibodies • Glycosaminoglycan-binding proteins

S. Pérez (✉)
Département de Pharmacochimie, UMR5063 CNRS-Université Grenoble Alpes, BP53, 38041 Grenoble cédex 09, France
e-mail: spsergeperez@gmail.com

A. Sarkar • A. Rivet • S. Drouillard • C. Breton • A. Imberty (✉)
Centre de Recherches sur les Macromolécules Végétales, UPR5301 CNRS (affiliated with Université Joseph Fourier and ICMG), BP53, 38041 Grenoble cédex 09, France
e-mail: imberty@cermav.cnrs.fr

© Springer Japan 2017
K.F. Aoki-Kinoshita (ed.), A Practical Guide to Using Glycomics Databases,
DOI 10.1007/978-4-431-56454-6_7
7.1 Introduction

Major advances in structural elucidation methods are benefitting glycobiology at large. Progress arising from the use of synchrotron radiation sources along with major advances in high-resolution NMR spectrometry and electron microscopy contributes strongly to these advances. In conjunction with these experimental sources of structural investigations, computational and molecular modeling methods are providing complementary information (Demarco and Woods 2008; Perez and Tvaroška 2014). Several distinct repositories hold 3D structural information which has been experimentally and theoretically determined for carbohydrates and carbohydrate-containing molecules (Perez and Mulloy 2005).

Among the 5000 entries related to carbohydrates in the Cambridge Structural Database (Allen 2002), only a small fraction is relevant to the field of glycobiology since most glyco-related entries relate to monosaccharides or to substituted intermediates for synthetic pathways (Perez 2007; Perez et al. 2000). As for polysaccharide structures, although a large amount of 3D structural models has accumulated over time, the effort to collect, curate, and disseminate this data electronically and freely to the scientific community has been limited when compared to other initiatives dealing with biomacromolecules.

In the past 45 years, more than 100,000 atomic-resolution structures have been deposited into the Protein Data Bank (PDB) (Berman et al. 2003). An increasing number of crystal structures have been reported for glycoproteins and protein-carbohydrate complexes. More than 5000 entries for glycoproteins or protein-carbohydrate complexes have been deposited, forming a valuable resource for glycoscientists (Jo and Im 2013). They occur, in their vast majority, as N-glycans or non-covalently bound ligands, O-Glycan chains forming a minority. In total, about 3.5% of the proteins in the PDB carry covalently bound glycan chains and thus can be classified as glycoproteins. The quality of the data does not always meet the high-quality standards; the structure needs to be curated and annotated. The glycosciences.de web portal (Lutteke et al. 2006) is a valuable utility for searching carbohydrate structures in the PDB.

As regards to theoretical characterizations, the complexity of oligosaccharide and polysaccharide topologies has required the design of dedicated molecular building procedures. These procedures can convert sequence information into reliable 3D models prior to any optimization through molecular mechanics or molecular dynamics methods, either as isolated molecules or in their interactions with proteins. All these constructions are based on the linking of preconstructed 3D molecular templates of monosaccharides (Engelsen et al. 2013; Lutteke et al. 2006; Woods 2014).

Despite their availability, these structural information have not yet gained full utilization in the rational design and engineering of glycan-based multivalent vehicles or glycan-grafted materials in view of their potential implications. These approaches, along with those provided by novel methodologies in the field of click chemistry or in the understanding of multivalency, are now part of the tool box that
glycoscientists have in hand to address and develop smart constructions that would exploit the full repertoire of the informational power of glycans. Indeed, a vast majority of structures deal with those carbohydrate molecules that are referred to as “glycan determinants,” i.e., those which are recognized by glycan-binding proteins. Those proteins are lectins, receptors, toxins, antibodies, microbial adhesins, carbohydrate-binding modules, transporters, but also enzymes involved in their synthesis, modification, and degradation. In the present work, only some of these glycan-binding proteins have been selected; they are lectins, monoclonal antibodies, glycosaminoglycan-binding proteins, and glycosyltransferases. As regards to the carbohydrates, the three-dimensional features of monosaccharides, disaccharides, oligosaccharides, and polysaccharides are covered in the form of databases, without claiming exhaustivity. A collection of annotated NMR data of bioactive oligosaccharides is also available. These databases have been developed with nonproprietary software, and they are opened freely to the scientific community. They are accessible throughout a common portal called “Glyco3D” http://glyco3d.cermav.cnrs.fr. Each individual database stands by itself as it covers a particular field of structural glycosciences. Nevertheless, the interconnection of these databases provides a unique opportunity to characterize the three-dimensional features that a given oligosaccharide molecule can take in different environments, i.e., vacuum, crystalline state, or interacting with different proteins having different biological function. To this aim, a common nomenclature has been adopted for the structural encoding of the carbohydrates.

7.2 Representing and Encoding Complex Carbohydrates

Representation in text of the primary structure, or sequence, of complex carbohydrates was first described following the IUPAC-IUBMB terminology in its extended and condensed forms (Mcnaught 1997). These forms are used within the carbohydrate community and are adequate for describing complex sugar sequences. Recommendations apply to the description of polysaccharides and glycoproteins (Mcnaught 1997). Other types of representations have been developed in glycobiology, favoring pictorial representations that facilitate the visualization of the monosaccharides. This is adequate as the number of basic carbohydrate units found in mammals is limited. Extension to the constituents found in bacterial and plant polysaccharides has also been developed and adopted (Varki et al. 2015). Figure 7.1 presents in a non-exhaustive fashion the results of such an extension, which allows the description of some structural descriptors which were not taken into account previously.

From the standpoint of bioinformatics, it is impractical to encode glycans (composed of more than 100 monosaccharide units) into distinct graphical symbols. To establish effective databases that can intercommunicate, a simple representation in a common/standard format is essential. This would facilitate computational
processing and ensure that the data content is non-redundant. Two approaches can
be followed to encode a carbohydrate molecule:

Connecting atom sets through chemical bonds is commonly used in chemoinfor-
matics. Chemical file formats like InChi (Mcnaught 1997) and SMILES (Weininger
1988) have been developed to aid storing of molecule information in chemical
databases like PubChem (Wang et al. 2010) or ChEBI (Degtyarenko et al. 2008).
IUPAC (extended), InChi, and SMILES encoding are computed from the chemical
drawing (ring structure), thereby allowing auto-generation of these encodings. There
are severe limitations that do not make this type of encoding the favored choice,
but InChi and SMILES are the proper formats to exchange data between distinct
databases.

Connecting building blocks (monosaccharides) through glycosidic linkages is far
more efficient to encode carbohydrates using a residue-based approach (Frank and
Schloissnig 2010). As compared to nucleic acids or proteins, there are a far greater
numbers of monosaccharides. In addition, carbohydrates are frequently found to
have branched structures; most of them are tree-like molecules. The prerequisite for
a residue-based encoding format is a controlled vocabulary of its residue names. It
makes sense to restrict the number of residues to as low as possible. The lack of
clear rules to subscribe atoms of a molecule to one particular monosaccharide, and
not to its substituent(s), is the main hurdle in encoding monosaccharide names. The

---

**Fig. 7.1** Pictorial representation of some important monosaccharide units
variety in nomenclature and structural representation of glycans makes it complex to decide the best form to use. The choice of notation is frequently based on whether the study is focused on the chemistry or on biology. The information content of each representation may vary or highlight a particular aspect as compared to others. While representing a complex glycan structure, chemists prefer to elucidate the structure that includes information about the anomeric carbon, the chirality of the glycan, the monosaccharides present, and the glycosidic linkages that connect them. For others, it is more interesting to visualize the monosaccharides present, and hence a symbolic/diagrammatic notation is favored (Fig. 7.2).

Due to the independent development of glycan databases in various geographical locations, several formats for representing glycan structures have been created. Chapter 2 provides an overview of the most common representations used in the field.

The three-dimensional depiction of glycans, polysaccharides, and glycoconjugates, coping with the accepted nomenclature and pictorial representation used in carbohydrate chemistry, biochemistry, and glycobiology, is made possible throughout the molecular visualization program Sweet Unity Mol (Perez et al. 2015).
7.3 Glyco3D Portal: An Ensemble of 3D Databases for Glycosciences

Glyco3D encompasses a family of databases covering the 3D features of monosaccharides, disaccharides, oligosaccharides, polysaccharides, glycosyltransferases, lectins, monoclonal antibodies, and glycosaminoglycan-binding proteins (Fig. 7.3). These databases, which have been manually curated, are web based and platform independent.

Whereas each of the databases has been set to account for the specific features of the class of molecules covered, the set of databases in Glyco3D share common facilities. Data retrieval and usability are the primary goals set by the developers of an effective database. An interactive front end was designed for each database with HTML pages and server side scripts that extract data from the tables on the relational database for user queries on Data Query and display the retrieved information in a coherent manner on Results.

The Data Query page comprises primarily two levels of increased complexity to query the database, i.e., Simple and Advanced searches. In the Simple Search option, a text is typed in the box provided, based upon which a result prompt appears to guide the user in selecting from the “hits” found in the database. An accordion function eases preview of the results. This can be used to expand or minimize the preview of the listed results of the user query for a first glance into the entries matching the request to the database. The Advanced Search is a multi-criteria search that can be used together for querying in various combinations as best suits the user’s requirement. Both the Simple Search and the Advanced Search options are equipped with an “auto-complete” function, which guides the user while querying.

Fig. 7.3 The Glyco3D home page
the database. It comprises two parts: (1) a single field of entered text and (2) the auto-prompt when the data is entered, through which the desired hit in the database can be selected either by scrolling down with the mouse or by using the arrow keys on the keyboard.

The Results page details the results which are organized under two tabs, namely, Molecule Information and Display and Download. The Molecule Information page provides a detailed description of the molecule (or macromolecule) as described throughout the several nomenclature schemes, along with key elements representative of the type of databases (e.g., trivial name, molecular weight, some literature references, etc.). The illustrative representations of the glycans or the glycan-interacting proteins can be viewed through the “Zoombox” feature that was developed by modifying an existing JQuery plug-in that allows the selected image to be zoomed and highlighted. The Display and Download tab incorporates the best representatives of the most probable low-energy conformational families. As regards to the display of the results, 3D structures can be viewed over the website via the Jmol application http://jmol.sourceforge.net/. Jmol is an interactive web browser applet that is an open-source, cross-platform 3D Java visualizing tool for viewing chemical and molecular structures. It provides high-performance 3D rendering with standard available hardware. Downloading the atomic coordinates (in pdb format) for further independent use is an option provided for all the databases. Additionally, a GUI has been designed to retrieve, interpret, and display the related information about each entry stored in the back end in the tables of the relational database and to display it interactively to the user. Finally, in an effort to assimilate other relevant resources for sugars, “External Links” are provided that empowers the user to explore more online glycoinformatics resources.

The databases run on an Apache web server (http://www.apache.org/) with the application program Hypertext Preprocessor (PHP) (http://www.php.net/). It has been implemented using the open source MySQL database (http://www.mysql.com/). They have been developed based on a combination of three layers. The underlying layer is the MySQL database system, a relational database management system that stores all the structure-related information in the back end and provides the facility to link two or more tables in the database. The intermediate layer is an Apache-PHP application [Apache 2, PHP] that receives the query from the user and connects to the database to fetch data from the upper layer, which comprises populated HTML pages, to the web browser client. The PHP and Java scripts are embedded in the HTML web pages for this effect and are used as application programs for integrating the back end (MySQL database) to the web pages (HTML). Apache has been used as the web server for building the interface between the web browser and the application programs. PHP was used for writing scripts to query the database, and Javascript (with JQuery plug-in) was used to design the auto-complete function for the user interface. The graphical user interface was developed with HTML (version 5) and CSS (version 3).
7.3.1 Monosaccharides

**Database Content** This is an annotated database that contains the 3D structural information of about 100 entries of monosaccharides. These monosaccharides constitute the building blocks of the vast majority of oligosaccharides, complex carbohydrates such as “glycan determinants” (blood group antigens, core structures, fucosylated oligosaccharides, sialylated oligosaccharides, Lewis antigens, GPI anchors, N-linked oligosaccharides, globosides, etc.), glycosaminoglycans, plant and algal polysaccharides, as well as some bacterial polysaccharides. For establishing the 3D database, they all have been subjected to systematic conformational sampling to determine their conformational preferences, using molecular mechanics optimization. Whereas most of the monosaccharides exhibit a fairly rigid ring conformation, some cases exist such as in the case of iduronic acid, idose, and all furanosides where several ring shapes can occur. In these cases, the low-energy conformations are available for each entry.

**Data Query** Upon reaching the search page, two buttons to query the database appear on the left hand panel.

**Simple Search** A search box is provided, in which the user inputs textual information related to the search. The result is a prompt to guide the user in selecting from the “hits” found in the database, by a simple search engine. A preview of the results is displayed in an accordion fashion. This can be used to expand or minimize the preview of the listed results of the user query for a first glance into the entries matching the request to the database. The preview provides the monosaccharide name, its absolute configuration (D or L), its anomeric configuration, category, and molecular weight to the user to make an informed choice.

**Advanced Search** Four search boxes appear, each of them offering the choice between criteria to select: trivial name, type of constituent, category, and molecular weight. A slider is provided for assigning a range of values to be queried in the molecular weight of the database entries. It consists of two cursors that can navigate on a bar for specifying the minimum and maximum limit of the search. Two text fields display the values of the current position on the slider bar. The slider cursors auto-adjust themselves when values are entered directly in the text boxes.

**Results** The detailed results are organized under two tabs: “Molecule Information” and “Display and Download” (Fig. 7.4).

**Molecule Information** This includes the trivial name of the monosaccharide, the graphical representation of the stereochromical configuration (when available, the symbol notation for carbohydrates of the Consortium for Functional Glycomics (http://www.functionalglycomics.org/)), and the molecular weight. Additional comments and literature references are present if available. The illustrative representations of the monosaccharide can be viewed through the “Zoombox” feature that allows the selected image to be zoomed and highlighted.
**Fig. 7.4** The monosaccharide page

*Display and Download* This tab incorporates the low-energy conformation of the monosaccharide and its methyl glycoside.

### 7.3.2 Disaccharides

**Database Content** This annotated database contains the 3D structural information of about 120 entries of disaccharides. These disaccharides constitute molecules in their own rights, and they constitute the building blocks of the vast majority of oligosaccharides, complex carbohydrates such as “glycan determinants” (blood group antigens, core structures, fucosylated oligosaccharides, sialylated oligosaccharides, Lewis antigens, GPI anchors, N-linked oligosaccharides, globosides, etc.), glycosaminoglycans, plants and algal polysaccharides, and some bacterial polysaccharides.

The relative orientation of two contiguous monosaccharides linked by a glycosidic bond in a disaccharide is characterized by the $\Phi$ and $\Psi$ torsion angles. In the so-called Heavy Atom Definition commonly used in crystallography, $\Phi$ is the torsion angle $\Phi = \text{O5-C1-O-Cx}$, and $\Psi$ is the torsion angle $\Psi = \text{C1-O1-Cx-Cx+1}$, where $x$ is the number of the carbon atom of the second monosaccharide with which the $1\rightarrow x$ glycosidic bond is formed. An alternate definition of use in NMR spectroscopy refers to the hydrogen atoms about the glycosidic bond in a way such as $\Phi^H = \text{H1-C1-O-Cx}$ and $\Psi^H = \text{C1-O-Cx-Hx}$. For two monosaccharides linked by a $1\rightarrow 6$ linkage, another parameter ($\omega$) is required describing the orientation
about the exocyclic bond C5-C6. Its orientation is customarily described by the
torsion angles O5-C5-C6-O6 and C4-C5-C6-O6, which combination defines the so-called
gauche-trans (gt), gauche-gauche (gg), and trans-gauche (tg) conformations
(Marchessault and Perez 1979). For each disaccharide, an exhaustive search was
performed using the MM3 molecular mechanics force field. This gave a complete
sampling of the conformational space, yielding the construction of a relaxed
adiabatic energy map, which is represented as a function of Φ and Ψ torsion angles.
In the case of 1→6 linkages, relaxed adiabatic maps can be established for the
three low-energy orientations of the torsion angle ω. Typically, the exploration of
each such energy maps indicates the occurrence of 2–4 energy minima.

Data Query  Upon reaching the search page, two buttons to query the database
appear on the left hand panel.

Simple Search  A search box is provided, in which the user inputs textual infor-
mation related to the search. A preview of the results is displayed in an accordion
fashion. The preview provides the disaccharide name, its absolute configuration, the
axial-equatorial nature of the glycosidic linkage, and molecular weight to the user
to make an informed choice.

Advanced Search  Four search boxes appear, each of them offering the choice
between criteria to select: trivial name, type of constituent, category, and molecular
weight. A slider is provided for assigning a range of values to be queried in the
molecular weight of the database entries. It consists of two cursors that can navigate
on a bar for specifying the minimum and maximum limit of the search. Two text
fields display the values of the current position on the slider bar. The slider cursors
auto-adjust themselves when values are entered directly in the text boxes.

Results  The detailed results are organized under two tabs: “Molecule Information”
and “Display and Download” (Fig. 7.5).

Molecule Information  This includes the trivial name of the disaccharide, its
sequence, the graphical representation of the stereochemical configuration (when
available, the symbol notation for carbohydrates of the Consortium for Functional
Glycomics), and the molecular weight. Additional comments and literature refer-
ences are present if available. The illustrative representations of the disaccharide
can be viewed through the “Zoombox” feature that allows the selected image to be
zoomed and highlighted.

Display and Download  This tab incorporates the best representatives of the
families of the most probable low-energy conformation(s) from 1 to 4.

7.3.3 BiOligo

Database Content  More than 250 entries of bioactive oligosaccharides are listed
in the BiOligo-annotated database, with details about 3D structural information.
Fig. 7.5 The disaccharide page

The glycan epitopes are complex carbohydrates with their associated substitutions and aglycones, most of them being targets for glycan-binding proteins. The glycan epitopes belong to widely occurring families like the blood group antigens, core structures, fucosylated oligosaccharides, sialylated oligosaccharides, Lewis antigens, GPI anchors, N-linked oligosaccharides, globosides, etc. Table 7.1 gives the classification of glycan determinants in BiOligo.

For establishing the database, the three-dimensional structures of each constituent were generated using a combination of the available carbohydrate molecular builders (Engelsen et al. 2013; Lutteke et al. 2006; Woods 2014) or the building facilities offered by Sybyl (Tripos Inc.) and Chimera (Pettersen et al. 2004). Once constructed, the glycans were subjected to systematic conformational sampling to determine their conformational preferences, using the Shape software (Rosen et al. 2009). In such cases, several low-energy conformations (1–5) are available for each entry. At the present time, the monumental work required to complete the computation work is still ongoing, and the results are being implemented in the database on a regular basis.

**Data Query** The database is available from Glyco3D portal. Upon reaching the search page, two buttons to query the database appear on the left hand panel.

**Simple Search** A search box is provided, in which the user inputs textual information related to the search. The result is a prompt to guide the user in selecting...
Table 7.1 Classification of glycan determinants in BiOligo

| Index | BiOligo category                                                                 |
|-------|----------------------------------------------------------------------------------|
| 1     | Blood group A antigens                                                            |
| 2     | Blood group B antigens                                                            |
| 3     | Blood group H antigens (blood group O)                                            |
| 4     | Blood group H antigens (blood group O) and Globo H tetraose                       |
| 5     | Core structures                                                                   |
| 6     | Core structures (type 1 and type 2)                                               |
| 7     | Core structures (type 1)                                                           |
| 8     | Core structures (type 2)                                                           |
| 9     | Core structures (type 4)                                                           |
| 10    | Fucosylated oligosaccharides                                                      |
| 11    | Fucosylated oligosaccharides (3-fucosyllactose core)                               |
| 12    | Fucosylated oligosaccharides (lacto series)                                        |
| 13    | GAGs                                                                             |
| 14    | Galα-3Gal oligosaccharides (Galili and xeno antigens)                              |
| 15    | Galα-3Gal oligosaccharides (isogloboseries)                                       |
| 16    | Ganglioside sugars                                                                |
| 17    | Globoside sugars (P antigens) (Forsman antigens)                                  |
| 18    | Globoside sugars (P antigens) (Globo series: core structure type 4)                |
| 19    | Globoside sugars (P antigens) (P blood group antigens and analogues)              |
| 20    | Globoside sugars (P antigens) (stage-specific embryonic antigens: SSEA-3 and SSEA-4) |
| 21    | Glucuronylated oligosaccharides                                                   |
| 22    | Glycosphingolipid                                                                 |
| 23    | Lewis antigens                                                                    |
| 24    | Miscellaneous                                                                     |
| 25    | Miscellaneous (blood group-related oligosaccharides)                              |
| 26    | Miscellaneous (chitin oligosaccharides)                                           |
| 27    | Miscellaneous (fibrinogen-related oligosaccharides)                               |
| 28    | Miscellaneous (LDN-related oligosaccharides)                                      |
| 29    | Miscellaneous (Lewis X-related oligosaccharides)                                  |
| 30    | Miscellaneous (TF-related oligosaccharides)                                       |
| 31    | Miscellaneous (TN-related oligosaccharides)                                       |
| 32    | Miscellaneous (Trehalose-like sugars)                                            |
| 33    | N-linked oligos                                                                    |
| 34    | Sialylated oligosaccharide (type 1)                                               |
| 35    | Sialylated oligosaccharide (type 2)                                               |
| 36    | Disaccharides (GlycoLego)                                                         |
| 37    | Monosaccharides (GlycoLego)                                                       |

from the “hits” found in the database, by a simple search engine. A preview of the results is displayed in an accordion fashion. This can be used to expand or minimize the preview of the listed results of the user query for a first glance into
the entries matching the request to the database. The preview provides the glycan name, category, and molecular weight to the user to make an informed choice.

**Advanced Search** Four search boxes appear each of them offering the choice between criteria to select: trivial name, type of constituent, category, and molecular weight. A slider is provided for assigning a range of values to be queried in the molecular weight of the database entries. It consists of two cursors that can navigate on a bar for specifying the minimum and maximum limit of the search. Two text fields display the values of the current position on the slider bar. The slider cursors auto-adjust themselves when values are entered directly in the text boxes.

**Results** The detailed results are organized under two tabs: “Molecule Information” and “Display and Download” (Fig. 7.6).

**Molecule Information** This includes the trivial name of the glycan, its sequence, the graphical representation of the stereochemical configuration, the symbol notation for carbohydrates of the Consortium for Functional Glycomics, the molecular weight, the glycan category or family in which it has been classified in the BiOligo Database, the glycan composition (i.e., the comprising glycan type and number of each such glycan), and the glycosidic linkages present in it. Additional comments and literature references are present if available. The illustrative representations of the glycan can be viewed through the “Zoombox” feature that allows the selected image to be zoomed and highlighted.
View and Download  This tab incorporates the best representatives of the families of the most probable low-energy conformation(s) from 1 to 4.

7.3.4  NMR: A NMR Database of Bioactive Oligosaccharides

This database contains the NMR structural information of more than 150 entries of bioactive oligosaccharides. This set of glycans determinants is a subset of the group of bioactive oligosaccharides which constitute the core of the BiOligo Database (Sarkar et al. 2015). They have been systematically organized using standard names in the field of glycobiology, into 31 categories and subcategories. The glycan determinants in the NMR Database constitute a subset of the entries of BiOligo. Prior to the establishment of the database, these glycan were synthesized in pure form and in sufficient quantity to be investigated throughout NMR spectroscopy. The synthetic work was conducted using recombinant methodology (Priem et al. 2002). For each of these glycans, the experimental work encompassed the recording and interpretation of $^1$H and $^{13}$C spectra, along with COSY, TOCSY, HMQC, and HMBC correlation spectra.

Data Query  The database is available from Glyco3D portal. Upon reaching the search page, two buttons to query the database appear on the left hand panel.

Simple Search A search box is provided, in which the user inputs textual information related to the sequence of the glycan. The result is a prompt to guide the user in selecting from the “hits” found in the database, by a simple search engine. A preview of the results is displayed in an accordion fashion. This can be used to expand or minimize the preview of the listed results of the user query for a first glance into the entries matching the request to the database.

Advanced Search  can be performed on different criteria: trivial name, sequence, category, and type of constituents. More complex searches can be made by combining criteria which can be interlaced from up to four search boxes.

Results  The detailed results are organized under two tabs: “Molecule Information” and “Display and Download” (Fig. 7.7).

Molecule Information  provides the trivial name, the sequence, the graphical representation of the stereochemical configuration, the symbol notation for carbohydrates of the Consortium for Functional Glycomics, the type of constituent, and the glycan category. The experimental conditions used to record the NMR spectra are given, i.e., temperature, solvent, frequency, and concentration.

Display and Download  This tab incorporates the representations of the chemical repeat and, in many cases, the $^1$H and $^{13}$C spectra, along with COSY, TOCSY, HMQC, and HMBC correlation spectra.
7.3.5 PolySac: A 3D Structural Database of Polysaccharides

**Database Content** The database contains the 3D structural information of about 140 polysaccharide entries that have been collected from an extensive screening of scientific literature (for review, see Perez (2007)). These were established using various structure determination techniques (fiber X-ray and neutron diffraction, electron diffraction on single crystals, molecular modeling, and high-resolution NMR spectroscopy). The details concerning the construction of the atomic coordinates of polysaccharides have been published previously (Sarkar and Perez 2012). The classification of polysaccharide families present in PolySac3DB is shown in Fig. 7.8.

**Data Query** The database is available from Glyco3D portal. The polysaccharide data organized in the database can be browsed starting from the search page. The data can be accessed in two ways.

**Simple Search** This option searches the database by just entering the name of the polysaccharide of interest. This is available through a drop-down button that enlists all the polysaccharides present in the database and groups all the entries in the database into 18 groups/families to clearly categorize the overall properties
displayed by these polysaccharides. Upon selection of a family, a further drop-down menu offers a list of the polysaccharides belonging to this family for which 3D structural information is available.

**Advanced Search** offers the choice among two criteria, either the chemical structure of the repeat unit or the method of resolution used to establish the structure.

**Results** The detailed results are organized under two tabs: “Molecule Information” and “Display and Download” (Fig. 7.8).

**Molecule Information** This includes the trivial name of the polysaccharide, its family and subfamily, its origin, the sequence of the repeating unit, the graphical representation of the stereochemical configuration, and the symbolic representation. Additional comments and literature references are present if available. The illustrative representations of the glycan can be viewed through the “Zoombox” feature that allows the selected image to be zoomed and highlighted.

**View and Download** This tab incorporates the molecular representations of the repeat unit, the macromolecular chain, and in some instances some packing features.
7.3.6  GT: A 3D Structural Database of Glycosyltransferases

**Database Content** Glycosyltransferases (GTs) constitute a ubiquitous group of enzymes that catalyze the synthesis of glycosidic linkages by the transfer of a sugar residue from a donor to an acceptor (Breton et al. 2012). Acceptor substrates are carbohydrates, proteins, lipids, DNA, and numerous small molecules such as antibiotics, flavonol, steroids, etc. Glycosyl donor substrates are mostly sugar nucleotides, such as UDP-GlcNAc, UDP-Gal, and GDP-Man. However, lipid-linked sugars, e.g., dolichol phosphate saccharides and unsubstituted phosphates, are also utilized. Acceptor substrates are carbohydrates, proteins, lipids, DNA, antibiotic, or other small molecules. Sugar-nucleotide-dependent GTs are often referred to as Leloir enzymes. The transfer of saccharides by GTs is regiospecific and stereospecific with two possible stereochemical outcomes resulting in either inversion or retention of the anomeric configuration of the transferred sugar. Glycosyltransferases display low sequence homology, and they have recently been classified into 90 families (CAZY database: http://www.cazy.org). At the present time, more than 140 GT crystal structures are available that have been grouped in 40 families. Surprisingly, the three-dimensional architectures of Leloir-type GTs are remarkably conserved, and their X-ray structures exhibit mostly two general types of folds, termed GT-A and GT-B. As for the GTs that utilize lipid-phosphate donor substrates, different folds have been observed.

**Data Query** Upon reaching the search page, two buttons to query the database appear on the left hand panel: Simple Search and Advanced Search.

**Simple Search** The classification of the GT proteins is made based on their origin: (1) animal, (2) archaea, (3) bacteria, (4) plant, (5) virus, and (6) yeast and fungi. Upon selecting one organism, a click opens a new menu that prompts the user to choose among a subclassification based on the fold, i.e., GT-A or GT-B. A further click opens a menu where the GTs are numbered according to the CAZY classification, and upon clicking on a CAZY family, the user can select the requested protein and be brought to the “GT information” page by selecting one PDB access code.

**Advanced Search** Under the name “Select Criteria,” a search box offers to select among the following items: (1) organism, (2) family, (3) PDB, and (4) authors. A search box is provided in which appears either a drop-down button enlisting all the entries corresponding to the selected item, or the user can directly enter a query (i.e., a PDB code). The result is a prompt to guide the user in selecting the “hits” found in the database, by a simple search engine. More complex searches can be made by combining criteria which can be combined from up to four search boxes. A preview of the results is displayed in an accordion fashion, whereby the enzyme name, the organism and type of complex if any, are given. The amount of information provided allows the user to make an informed choice prior going to more information on the selected GT.
Results  The detailed results are available under two tabs: “Molecule Information” and “Display and Download” (Fig. 7.9).

Molecule Information  Under this button, the following information are provided: enzyme name, short name, origin, organism, resulting linkage, fold, CAZY family, and mechanism. As regards to the crystal structure, the PDB code, the resolution, the nature of the complex (if any), comments, and references are given.

Display and Download  On this page will be represented one or more graphical representations of the glycosyltransferase.

7.3.7 mABS: A 3D Structural Database of Monoclonal Antibodies Against Carbohydrates

Database Content  Antibodies are glycoproteins belonging to the immunoglobulin superfamily. Three-dimensional structures have been established from X-ray crystallography as listed in http://www.bioinf.org.uk/abs/sacs/ (Allcorn and Martin 2002). Carbohydrate determinants recognized by antibodies are expressed on the cell surface as glycolipids and glycoproteins. In many instances, the minimum carbohydrate epitopes are located at the terminal end of more complex carbohydrate chains, experiencing a wide range of contexts, surface densities, and
surroundings. Therefore, antibodies with similar specificities for individual carbohydrate epitopes can exhibit different selective cell profiling depending upon the unique presentation of the carbohydrate on the target cells. As a consequence, anti-carbohydrate antibodies with specificity to oligosaccharides and polysaccharides are of a high importance in immunology and are attractive targets for vaccine design (Pazur 1998). In the present database, the set of high-resolution structures of carbohydrate-antibody complexes is somehow limited. These studies are typically limited to systems involving antibody fragments, such as the antigen-binding fragment (Fab) or variable fragment (Fv), and to small oligosaccharides. Analysis of these complexes reveals general trends about how antibodies recognize different types of carbohydrates. Antibodies which recognize a terminal carbohydrate motif generally feature cavity-like binding sites, where one or more carbohydrate residues are anchored in the cavity by “end-on” extension. Antibodies which recognize an internal carbohydrate motif, as a single repeat of a bacterial polysaccharide, for example, generally exhibit groove-like binding sites or very large cavities which are open at both ends of the site, allowing for “side-on” entry of the antigen.

**Data Query** Upon reaching the search page, two buttons to query the database appear on the left hand panel: Simple Search and Advanced Search.

**Simple Search** The first level offers the choice between the natures of the antibody, i.e., human, humanized, or mouse. After selection, a further right click opens a window on “Antibody Information.”

**Advanced Search** Under the name “Select Criteria,” a search box offers to select among the following items: (1) nomenclature, (2) antibody, (3) origin, and (4) immunoglobulin type. A search box is provided in which appears a drop-down button enlisting all the entries corresponding to the selected item. The result is a prompt to guide the user in selecting the “hits” found in the database, by a simple search engine. More complex searches can be made by combining criteria which can be combined from up to four search boxes. A preview of the results is displayed. The amount of information provided allows the user to make an informed choice prior going to “Antibody Information.”

**Results** The detailed results are available under two tabs: “Antibody Information” and “Display and Download” (Fig. 7.10).

**Molecule Information** lists the origin, name of the antibody, nomenclature of the bound carbohydrate, PDB code, resolution, comment, immunoglobulin class, and reference to the original article. Provision is also given to view a still three-dimensional ribbon-type representation of the three-dimensional structure. Links to Medline (http://www.ncbi.nlm.nih.gov/pubmed/) and Protein Data Bank (http://www.rcsb.org/pdb) are also provided.

**Display and Download** On this page, is given a three-dimensional representation of the three-dimensional structure of the complex which has been constructed from the reported atomic coordinates. In the case of mAbs–carbohydrate crystalline
complexes, a particular emphasis is given to indicate the conformation of the bound carbohydrate which can be viewed.

### 7.3.8 GAG: 3D Structural Database of Glycosaminoglycan-Binding Proteins

**Database Content** The glycosaminoglycans (GAGs) comprise a class of complex anionic polysaccharides which, through their linkage to a core protein, are part of more complex macromolecules (proteoglycans). There are several GAG families: (1) glycosaminoglycans (heparin and heparan sulfate), (2) galactosylaminoglycans (chondroitin sulfate and dermatan sulfate), and (3) hyaluronic acid and keratan sulfate. GAGs are assembled from repeating disaccharides, and they exhibit diverse patterns of sulfation. Among them, hyaluronic acid is unique as it is not attached covalently to a core protein and it lacks sulfation. In addition to their participation in the physicochemical properties of the extracellular matrix, glycosaminoglycan fragments are specifically recognized by protein receptors, and they play a role in the regulation of many processes, such as hemostasis, growth factor control, anticoagulation, and cell adhesion (Gandhi and Mancera 2008; Imberty et al. 2007; Raman et al. 2005).
Given the importance of protein–GAG interactions, oligosaccharide fragments are prime targets for drug design. The classes of proteins interacting with GAGs are chemokines, complement proteins, components of the extracellular matrix, enzymes, growth factors, lectins, toxins, and viruses. There is a small number of crystal structures of complexes available in the Protein Data Bank. Most of the reported structures deal with proteins which have been co-crystallized with heparin oligosaccharides. Enzymes are limited to only two cases: one heparinase and one sulfotransferase illustrating that a large number of protein interacting with heparin sulfate are receptors.

**Data Query**  Upon reaching the search page from Glyco3D, two buttons to query the database appear on the left hand panel: Simple Search and Advanced Search.

**Simple Search** The classification of the GAG-binding proteins is made based on their biological function: chemokine, complement protein, extracellular matrix (ECM) protein, enzyme, growth factor, lectin, toxin, and virus. Upon selecting one family, a right click opens a new menu that prompts the user to choose among a subclassification. A further right click opens a window on “GAG information.”

**Advanced Search** Under the name “Select Criteria,” a search box offers to select among the following items: (1) protein, (2) nature of GAG, and (3) PDB. A search box is provided in which appears a drop-down button enlisting all the entries corresponding to the selected item. The result is a prompt to guide the user in selecting the “hits” found in the database, by a simple search engine. More complex searches can be made by combining criteria which can be interlaced from up to four search boxes. A preview of the results is displayed in an accordion fashion, whereby the classification, the protein name, the GAG type, and the size of the oligosaccharide are given. The amount of information provided allows the user to make an informed choice prior to going to “GAG Information.”

**Results** The detailed results are available under two tabs: “GAG Information” and “Display and Download” (Fig. 7.10).

**Molecule Information** Under the button, “GAG Information” is given: protein, classification, GAG type, species, PDB code, resolution, length of oligosaccharide, comments, and references. Provision is also given to view an image of the source of the protein, along with a graphical representation of the three-dimensional structure. Links to Medline (http://www.ncbi.nlm.nih.gov/pubmed/), Protein Data Bank (http://www.rcsb.org/pdb), and Uniprot (http://www.uniprot.org/uniprot) are also provided.

**Display and Download** On this page, is represented a still three-dimensional ribbon-type representation of the three-dimensional structure which has been constructed from the reported atomic coordinates. In the case of protein–GAG crystalline complexes, a particular emphasis is given to indicate the location and conformation of the bound carbohydrate.
7.3.9 LECTIN3D: A 3D Structural Database of Lectins

**Database Content** Lectins are proteins of nonimmune origin that bind to specific carbohydrates without modifying them. Per current knowledge, they act like molecular readers to decipher sugar-encoded information. They play biologically important roles in recognition processes involved in fertilization, embryogenesis, inflammation, metastasis, and parasite–symbiote recognition in microbes, invertebrates, plants, and vertebrates (Ambrosi et al. 2005; Arnaud et al. 2013, Sharon 2007). In the plant kingdom, lectins have been demonstrated to play a role in defense against pathogens or predators and hypothesized to be involved in establishing symbiosis with mushrooms and bacteria of the Rhizobia species. Among the proteins that interact non-covalently with carbohydrates, lectins bind mono- and oligosaccharides reversibly and specifically.

More than 1400 three-dimensional structures of lectins have been solved and are available in the database as of January 2016. They have been determined by X-ray diffraction, although some neutron diffraction structures are available as well as NMR solution structures or theoretical models. This covers almost 250 different proteins. Most 3D structures have been for the time being, obtained for plant and animal lectins. Nevertheless, the number of structural investigations dealing with viral and bacterial materials increases rapidly. Among these structures, 64% are complexed with a carbohydrate ligand which can be multiforms; i.e., they are monosaccharides, oligosaccharides, glycoproteins, or synthetic glyco-compounds from organic chemistry (Table 7.2). Accurate determination of carbohydrate–lectin complexes remains a nontrivial problem due to the shallow and multichambered binding sites of many lectins. This is nevertheless a requirement to access information about their binding mechanisms in biologically relevant conditions.

Structures of plant and animal lectins are the most abundant in the database (Fig. 7.11). The number of structures of lectins from bacteria, fungi, and viruses is growing steadily as their biological functions are being recognized (Imberty and Varrot 2008; Varrot et al. 2013). Only one structure of algal lectin is available at the present time (Ziolkowska et al. 2006). As for structural motifs, there is a strong predominance of β-sheets. Two such β-sheets can assemble to form a β-sandwich, an architecture commonly found in more than half of the entries. These β-sandwiches exhibit dissimilarities and the location of binding sites shows many variations. For example, the immunoglobulin-like fold of animal sialo-adhesins is very different from the jelly-roll fold of legume lectins. More complex combinations of β-sheets occur, giving rise to β-propellers and β-prisms. Despite the fact that topologies are very different between families, some interesting structural convergences are nevertheless observed. Intracellular animal lectins which are involved in the quality control of glycoprotein synthesis share the same protein fold with legume lectins, now referred to as L-lectin (Loris 2002).

**Data Query** Upon reaching the search page form, two buttons to query the database appear on the left hand panel: Simple Search and Advanced Search.
Table 7.2 Classification and distribution of apo/complex 3D structures of lectins

| Plant              | Lectin            | Free | Complex | Lectin            | Free | Complex |
|--------------------|-------------------|------|---------|-------------------|------|---------|
| L-type (legume)    | 75                | 152  | Hevein type | 13                | 19   |
| β-prism I (jacalin)| 16                | 35   | β-prism II (monocot) | 7        | 12   |
| R-type lectin (β-trefoil) | 15 | 20 | Cyanovirin-N family | 1      | 0    |
| β-prism II (monocot) | 7           | 12   |          |                    |      |         |
| Bacteria           | Pili adhesin      | 9    | 42      | Cytolysin         | 2    | 7       |
| Neurotoxin         | 21                | 20   | Staphylococcal toxin | 2      | 5    |
| AB5 toxin          | 15                | 20   | β-trefoil | 1      | 4    |
| 2-Ca β-sandwich    | 3                 | 30   | Scytovirin | 3        | 0    |
| Cyanovirin-N family| 17                | 8    | Serine-rich repeat | 2      | 1    |
| 1-Ca β-sandwich    | 2                 | 16   | Toxin repetitive domain | 1      | 1    |
| β-propeller        | 0                 | 11   | TNFα-like | 0      | 1    |
| Oscillatoria agglutinin | 7 | 3   |          |                    |      |         |
| Animal             | C-type            | 44   | 86      | Calnexin-calreticulin | 5 | 2    |
| Galectin           | 33                | 103  | TIM-lectin | 2      | 6    |
| Fibrinogen-like    | 6                 | 15   | L-rhamnose binding | 4      | 3    |
| I-type             | 5                 | 11   | C-type lectin-like | 3      | 1    |
| P-type             | 6                 | 9    | Malectine | 1      | 2    |
| R-type (β-trefoil) | 3                 | 10   | F-type | 0      | 2    |
| H-type             | 4                 | 11   | Cys-knot | 1      | 0    |
| Pentraxin          | 7                 | 3    | β-propeller | 2      | 2    |
| Microneme MAR      | 2                 | 7    | Chitin binding | 1      | 0    |
| L-type             | 5                 | 3    |          |                    |      |         |
| Virus              | Hemagglutinin     | 44   | 23      | Rotavirus spike   | 4    | 11     |
| Norovirus capsid   | 18                | 38   | Fiber knob | 2      | 12   |
| Polyomavirus capsid| 12              | 15   | Coat protein | 0      | 3    |
| Phage tailspike    | 8                 | 8    | Coronavirus spike | 1      | 0    |
| Fungi and yeast    | Galectin          | 13   | 21      | L-type 1          | 7    | 0      |
| Actinoporin-like   | 7                 | 14   | 7-blades β-propeller | 2      | 6    |
| β-trefoil          | 8                 | 20   | 6-blades β-propeller | 1      | 9    |
| Yeast adhesin      | 2                 | 11   | Ig-like | 1      | 0    |
| Cyanovirin-N hom.  | 7                 | 1    |          |                    |      |         |
| Algae              | Griffithsin       | 5    | 8       |          |      |         |

**Simple Search** The classification of the lectins is made based on their origin: (1) algae, (2) animal, (3) bacteria, (4) fungi and yeast, (5) plant, and (6) virus. Upon selecting one family, a right click opens a new menu that prompts the user to choose among a subclassification based on the fold family and then on the species of organisms. A further right click opens a new menu that contains all the three-dimensional structures of the selected lectin either in the apo state or complexed with ligand. A preview of the results is displayed in an accordion fashion, whereby the PDB code, the species, the resolution at which the structure has been solved, and the
reference to the original publication are given. The amount of information provided allows the user to make an informed choice prior to going to “Lectin Information.”

**Advanced Search** Under the name “Select Criteria,” a search box offers to select among the following items: (1) species, (2) family, (3) sugars, and (4) PDB. A search box is provided in which appears a drop-down button enlisting all the entries corresponding to the selected item. For other items, the menu guides the user in selecting the “hits” found in the database, by a simple search engine. More complex searches can be made by combining criteria from up to four search boxes.

**Results** The detailed results are available under two tabs: “Molecule Information” and “Display and Download.”

**Molecule Information** Under the button, “Molecule Information” is given: origin, class, family, species, PDB code, resolution, comment, and reference. The comment section indicates whether the lectin has been solved in the form of a protein-carbohydrate complex. In that case, the nature of the sugar is indicated along with its sequence. Provision is also given to view an image of the source of the protein, along with a still three-dimensional ribbon-type representation of the three-dimensional structure, together with access to original 3D information at the Protein Database. Links to NIH sites for references and taxonomy are also provided. If the lectin has been submitted to the Consortium for Functional Glycomics for analysis of specificity through glycan arrays (Agravat et al. 2014), a link to the data page is also provided.

**Display and Download** On this page, is represented one or more graphical representations of the lectin or of the binding site with carbohydrate ligands that have
been constructed from the reported atomic coordinates. A particular emphasis is given to indicate the location and conformation of the bound carbohydrate (Fig. 7.12 provides an overall presentation of the “Lectin page”).

### 7.4 Sequence Search in the Whole Glyco-3D Portal

Glyco3D encompasses a family of databases covering the 3D features of monosaccharides, disaccharides, oligosaccharides, polysaccharides, glycosyltransferases, lectins, monoclonal antibodies, and glycosaminoglycan-binding proteins. Each of these databases has been set to account for the specific features of the class of molecules covered. Nevertheless, a logical network has been established that links all these databases together (Fig. 7.13). A search engine has been developed that scans the full content of all the databases for queries related to sequential information of the carbohydrates or other related descriptors. This is performed under the “search sequence” command.

For example, when looking for all information related to the H-type 1 trisaccharide, one needs to insert the request “Fuc a1-2 Gal b1-3 GlcNAc” in the “search sequence.” This results in 22 hits in the “BiOligo” Database, 23 NMR spectra, and
Fig. 7.13 Example of oligosaccharide search in the Glyco-3D portal

11 crystal structures of lectins. The large number of hits is due to the fact that the H-type 1 epitope is embedded not only in type 1 ABO oligosaccharides but also in several Lewis epitopes such as Lewis a, sialyl Lewis a, and Lewis b. The resulting structures for which NMR data are available are shown, whereas those lectin structures which have been co-crystallized with this glycan are given. They include lectins from virus, plants, and animals, as well as the bacterial lectin BambL from *Burkholderia ambifaria* complexed with H-type 1 tetrasaccharide depicted in Fig. 7.13 (Audfray et al. 2012).

### 7.5 Conclusions

The present Glyco3D portal offers a single entry to access three-dimensional features of glycans, polysaccharides, and glycan-recognizing proteins. A particular emphasis was given to the use of a common nomenclature for the structural encoding of the carbohydrates. Yet every glycan molecule is described by four different types of representations in order to cope with the different usages in chemistry and biology. As the information content of each representation may vary or highlight a particular aspect compared to others, this offers the most complete description of the many features that characterize glycans. If the use of GlycoCT
code is meant to facilitate future exchanges with other databases, this may not be sufficient, and other coding such as InChi and SMILES may be required.

As with any other databases, the value of the repository lies very much on the quality of the annotation and the curation of data. Now that the overall architecture of the databases and their interconnection are established, there remains the daily intervention to maintain the content updated and validated by expert scientists. To this end, an administration interface has been designed to facilitate greatly this task. Access to this interface can be open to extend the size of the curators.

The scientific value of the Glyco3D portal and its constituting databases has proven to be of great help at the level of structural biology. Obviously, connections with other 3D structural databases would increase the number of the glycans covered, as well as the scope of the applications toward rational design of complex glycan-containing architectures. Another expected enhanced value would result from mutual online access between other databases dealing with the functional aspects of complex carbohydrates. Glyco3D should be an asset to the community for probing further into the behavior of the very important class of glycomolecules and would open the way to establish a closer collaboration with bioinformatics groups in proteomics and genomics.

Acknowledgments The research leading to this publication has received funding from the European Commission’s Seventh Framework Programme FP7/2007–2013 under grant agreement n° 215536 and from the Agence Nationale de la Recherche under the “Genomic and Plant Biotechnology” Action throughout the “Wall-Array” project. We are grateful to the Marie Curie Initial Training Network as part of the FP7 People Programme for training and funding.

References

Agravat SB, Saltz JH, Cummings RD, Smith DF (2014) GlycoPattern: a web platform for glycan array mining. Bioinformatics 30:3417–3418

Allcorn LC, Martin AC (2002) SACS—self-maintaining database of antibody crystal structure information. Bioinformatics 18:175–181

Allen FH (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. Acta Crystallogr Sect B Struct Sci 58:380–388

Ambrosi M, Cameron NR, Davis BG (2005) Lectins: tools for the molecular understanding of the glycocode. Org Biomol Chem 3:1593–1608

Arnaud J, Audfray A, Imberty A (2013) Binding sugars: from natural lectins to synthetic receptors and engineered neolecists. Chem Soc Rev 42:4798–4813

Audfray A, Claudinon J, Abounit S, Ruvoën-Clouet N, Larson G, Smith DF, Wimmerová M, Le Pendu J, Römer W, Varrot A, Imberty A (2012) The fucose-binding lectin from opportunistic pathogen Burkholderia ambifaria binds to both plant and human oligosaccharidic epitopes. J Biol Chem 287:4335–4347

Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide protein data bank. Nat Struct Biol 10:980

Breton C, Fournel-Gigleux S, Palcic MM (2012) Recent structures, evolution and mechanisms of glycosyltransferases. Curr Opin Struct Biol 22:540–549
Degtyarenko K, De Matos P, Ennis M, Hastings J, Zbinden M, Mcnaught A, Alcantara R, Darsow M, Guedj M, Ashburner M (2008) ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res 36:D344–D350

Demarco ML, Woods RJ (2008) Structural glycobiology: a game of snakes and ladders. Glycobiology 18:426–440

Engelsen SB, Hansen P, Perez S (2013) POLYS: an open source software package for building three-dimensional structures of polysaccharides. Biopolymers 101:733–743

Frank M, Schloissnig S (2010) Bioinformatics and molecular modeling in glycobiology. Cell Mol Life Sci 67:2749–2772

Gandhi NS, Mancera RL (2008) The structure of glycosaminoglycans and their interactions with proteins. Chem Biol Drug Des 72:455–482

Imberty A, Varrot A (2008) Microbial recognition of human cell surface glycoconjugates. Curr Opin Struct Biol 18:567–576

Imberty A, Lortat-Jacob H, Perez S (2007) Structural view of glycosaminoglycan-protein interaction. Carbohydr Res 342:430–439

Jo S, Im W (2013) Glycan fragment database: a database of PDB-based glycan 3D structures. Nucleic Acids Res 41:D470–D474

Loris R (2002) Principles of structures of animal and plant lectins. Biochim Biophys Acta 1572:198–208

Lutteke T, Bohne-Lang A, Loss A, Goetz T, Frank M, Von Der Lieth CW (2006) GLYCO-SCIENCES.de: an internet portal to support glycomics and glycobiology research. Glycobiology 16:71R–81R

Marchessault RH, Perez S (1979) Conformations of the hydroxymethyl group in crystalline aldohexopyranoses. Biopolymers 18:2369–2374

Menaught AD (1997) Nomenclature of carbohydrates (Recommendations 1996). Adv Carbohydr Chem Biochem 52:43–177

Nakahara T, Hashimoto R, Nakagawa H, Monde K, Miura N, Nishimura S (2008) Glycoconjugate data bank: structures—an annotated glycan structure database and N-glycan primary structure verification service. Nucleic Acids Res 36:D368–D371

Pazur JH (1998) Anti-carbohydrate antibodies with specificity for monosaccharide and oligosaccharide units of antigens. Adv Carbohydr Chem Biochem 53:201–261

Perez S (2007) Oligosaccharide and polysaccharide conformations by diffraction methods. In: Kamerling IP (ed) Comprehensive glycosciences: analysis of glycans. Elsevier, Oxford, pp 193–220

Perez S, Mulloy B (2005) Prospects for glycoinformatics. Curr Opin Struct Biol 15:517–524

Perez S, Tvaroška I (2014) Carbohydrate-protein interactions: molecular modeling insights. Adv Carbohydr Chem Biochem 71:9–136

Perez S, Gautier C, Imberty A (2000) Oligosaccharide conformations by diffraction methods. In: Ernst B, Hart G, Sinay P (eds) Oligosaccharides in chemistry and biology: a comprehensive handbook. Wiley/VCH, Weinheim, pp 969–1001

Perez S, Tubiana T, Imberty A, Baaden M (2015) Three-dimensional representations of complex carbohydrates and polysaccharides: a video game based computer graphic software. Glycobiology 25:483–491

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612

Priet B, Gilbert M, Wakarchuk WW, Heyraud A, Samain E (2002) A new fermentation process allows large-scale production of human milk oligosaccharides by metabolically engineered bacteria. Glycobiology 12:235–240

Raman R, Sasisekharan V, Sasisekharan R (2005) Structural insights into biological roles of protein-glycosaminoglycan interactions. Chem Biol 12:267–277

Rosen J, Miguet L, Perez S (2009) Shape: automatic conformation prediction of carbohydrates using a genetic algorithm. J Cheminf 1:16
Sarkar A, Perez S (2012) PolySac3DB: an annotated data base of 3 dimensional structures of polysaccharides. BMC Bioinf 13:302
Sarkar A, Drouillard S, Rivet A, Perez S (2015) Databases of conformations and NMR structures of glycan determinants. Glycobiology 25:1480–1490
Sharon N (2007) Lectins: carbohydrate-specific reagents and biological recognition molecules. J Biol Chem 282:2753–2764
Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, Prestegard JJ, Schmaar RL, Freeze HH, Marth JD, Bertozzi CR, Etzler ME, Frank M, Vliegenthart FG, Lütteke T, Perez S, Bolton E, Rudd P, Paulson J, Kanehisa M, Toukach P, Aoki-Kinoshita KF, Dell A, Narimatsu H, York W, Taniguchi N, Kornfeld S (2015) Symbol nomenclature for graphical representations of glycans. Glycobiology 25:1323–1324
Varrot A, Basheer SM, Imberty A (2013) Fungal lectins: structure, function and potential applications. Curr Opin Struct Biol 23:678–685
Wang Y, Bolton E, Dracheva S, Karapetyan K, Shoemaker BA, Suzek TO, Wang J, Xiao J, Zhang J, Bryant SH (2010) An overview of the PubChem BioAssay resource. Nucleic Acids Res 38:D255–D266
Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36
Woods RJ (2014) GLYCAM Web (http://www.glycam.com). Complex Carbohydrate Research Center, University of Georgia, Athens
Ziolkowska NE, O’keefe BR, Mori T, Zhu C, Giomarelli B, Vojdani F, Palmer KE, McMahon JB, Wlodawer A (2006) Domain-swapped structure of the potent antiviral protein griffithsin and its mode of carbohydrate binding. Structure 14:1127–1135