

## Glyco3D: A Portal for Structural Glycosciences

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

### Abstract

The present work describes, in a detailed way, a family of databases covering the three-dimensional features of monosaccharides, disaccharides, oligosaccharides, polysaccharides, glycosyltransferases, lectins, monoclonal antibodies against carbohydrates, and glycosaminoglycan-binding proteins. These databases have been developed with non-proprietary software, and they are open freely to the scientific community. They are accessible through the common portal called “Glyco3D” <http://www.glyco3d.cermav.cnrs.fr>. The databases are accompanied by a user-friendly graphical user interface (GUI) which offers several search options. All three-dimensional structures are available for visual consultations (with basic measurements possibilities) and can be downloaded in commonly used formats for further uses.

**Key words** Three-dimensional structures, Bioactive oligosaccharides, Polysaccharides, Lectins, Glycosyl-transferases, Monoclonal antibodies against carbohydrates, Glycosaminoglycans interacting proteins

---

## 1 Introduction

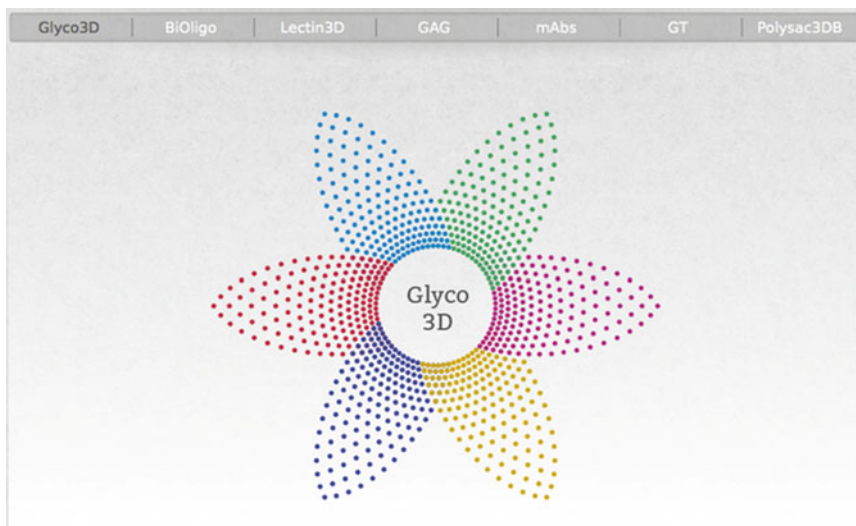
Structural glycobiology is a rapidly progressing field of research where the diverse structural and functional roles of carbohydrates (in the form of oligosaccharides and polysaccharides and glycoconjugates) are investigated and established throughout a wide diversity of experimental and theoretical methods. A large number of carbohydrate sequences have been determined through extensive work in areas of chemical and biochemical fragmentations followed by analysis using mass spectroscopy and nuclear magnetic resonance [1, 2]. The primary impetus behind the growth of glycoinformatics [3–7] has been the construction of large-scale repositories to store, organize, and disseminate the data that was rapidly being generated through experiments and theoretical calculations in relation to glycan sequence and structure [8–15].

Various algorithms and tools have been developed to query these repositories, interlink them, and provide useful calculations and analyses of the existing data.

Though there have been major advances and efforts in the sequence determination of carbohydrates, the three-dimensional structures of the complex glycans lagged behind considerably due to their inherent complexity and variability [16]. The knowledge of the three-dimensional structures of these molecules that result from complex biosynthetic events is needed for understanding the biological processes involving their interactions at the molecular level.

Experimentally determined three-dimensional structures of carbohydrates are available in two distinct repositories. The Cambridge Structural Database (CSD, <http://www.cdc.cam.ac.uk>), which is not an open source resource, contains entries in the form of structural data related to geometry, configuration, conformation, and packing of molecular crystals and organometallic structures. Although the amount of data related to carbohydrate is over 5,000, only a small fraction appears to be relevant to the field either because many deal with synthetic compounds or because of the limited size of the oligosaccharides that have been crystallized [17, 18]. Many crystal structures of small oligosaccharides are available through the Glyco3D Web portal (<http://glyco3d.cermav.cnrs.fr/>). The reluctance of carbohydrates to crystallize in a form suitable for X-ray diffraction studies is more pronounced for compounds having molecular weight ranging from 1,000 to 5,000. Over the last two decades, an increasing number of crystal structures have been reported for glycoproteins and protein-carbohydrate complexes. Whereas the structural information can be assessed in the Protein Data Bank (PDB, <http://www.rcsb.org/pdb>), it must be recognized that the quality of the data does not always meet the high quality standards and the structures relevant to the field need to be curated and annotated [19]. Unfortunately, due to the lack of a consistently used nomenclature for carbohydrates in PDB files, it may be difficult to find the structure of interest. The GLYCOSCIENCES.de Web portal [12] and the Glycoconjugate Data Bank Structure (<http://www.glycostructures.jp>) [20] provide convenient ways for searching carbohydrate structures in the PDB. As for the polysaccharide structures, although a large amount of 3D information regarding their structures has accumulated over time, the effort to collect, curate, and disseminate this data electronically and freely to the scientific community has been feeble when compared to similar initiatives in the field of proteomics or genomics.

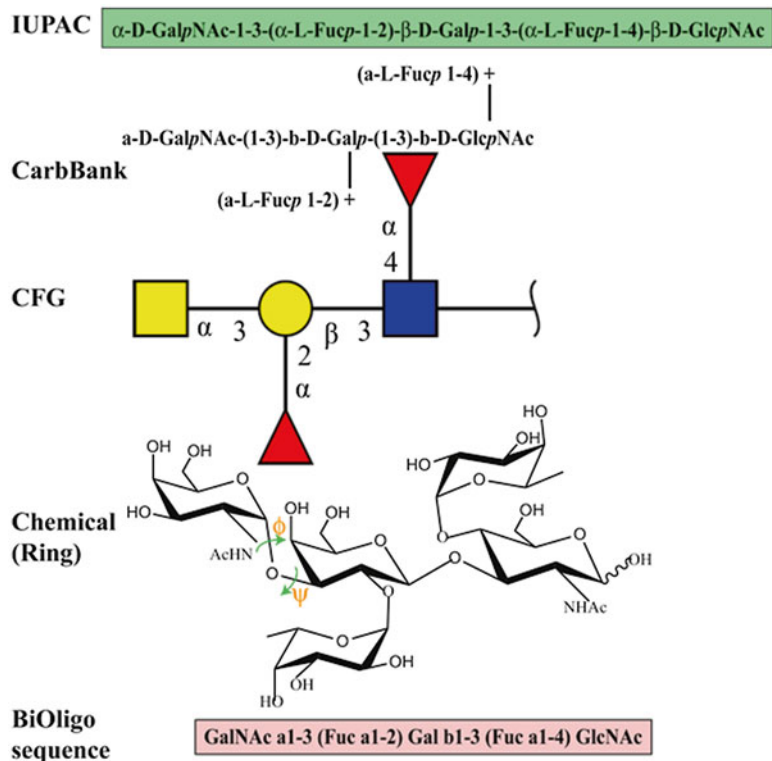
The present work describes, in a detailed way, a family of databases covering the three-dimensional features of monosaccharides, disaccharides, oligosaccharides, polysaccharides, glycosyltransferases, lectins, monoclonal antibodies against carbohydrates, and



**Fig. 1** The Glyco3D portal provides access to several databases covering three-dimensional structures of monosaccharides, disaccharides, oligosaccharides, polysaccharides; and protein–carbohydrate interactions

glycosaminoglycan-binding proteins (Fig. 1). These databases have been developed with non-proprietary software and they are opened freely to the scientific community. They are accessible throughout the 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 utilization of the databases offers a unique opportunity to characterize the three-dimensional features that a given oligosaccharide molecule can take in different environments, i.e., in vacuum, crystalline state, interacting with different proteins having different biological functions. To this aim, a common nomenclature has been adopted for the structural encoding of the carbohydrates.

The variety in nomenclature and structural representations of glycans makes it complex to decide the best form of illustrating the approach of the scientific investigation. The choice of notation is frequently based on whether the study is focused on the chemistry or has a more biological approach. Moreover, the information content of each representation may vary or highlight a particular aspect compared to others. For example, while representing a complex glycan structure, chemists privilege the representation 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. The most popular and distinct ways of representing complex carbohydrates are shown on Fig. 2.



**Fig. 2** The different levels of glycan encodings. The example used to illustrate the variety of notations in this figure is blood group A Lewis B antigen

## 2 Materials

### 2.1 Hardware

The databases can be accessed from any computer with access to Internet.

### 2.2 Software

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 for further independent use is an option provided for all the databases. The GUI has been designed to retrieve, interpret, and display the related information about each entry stored in the back end on the four tables of the relational database and display it interactively to the user.

## 3 Methods

### 3.1 *BiOligo: A 3D Structural Database of Bioactive Oligosaccharides*

#### 3.1.1 Database Content

BiOligo is an annotated database that contains the 3D structural information of more than 250 entries of bioactive oligosaccharides (referred to as “glycan determinants”) along with their constituting disaccharide segments (about 120) and monosaccharide segments (about 80). The glycan determinants are complex carbohydrates with their associated substitutions and aglycones that are recognized by glycan-binding proteins; these include lectins, receptors, toxins, microbial adhesins, antibodies, and enzymes [2]. They belong to the widely occurring families like the blood group antigens, core structures, fucosylated oligosaccharides, sialylated oligosaccharides, Lewis antigens, GPI-anchors, N-linked oligosaccharides, globosides, ... (Table 1). For establishing the 3D

**Table 1**  
**The classification of 3D structures of glycan determinants in BiOligo**

BiOligo category	Number of entries
Blood group A antigens	11
Blood group B antigens	11
Blood group H antigens (Blood group O)	12
Blood group H antigens (Blood group O) and Globo H tetraose	1
Core structures	1
Core structures (Type 1 and Type 2)	4
Core structures (Type 1)	4
Core structures (Type 2)	16
Core structures (Type 4)	1
Fucosylated oligosaccharides	4
Fucosylated oligosaccharides (3 Fucosyllactose core)	4
Fucosylated oligosaccharides (Lacto-Series)	13
GAGs	14
Gal $\alpha$ -3Gal oligosaccharides (Galili and xeno antigens)	6
Gal $\alpha$ -3Gal oligosaccharides (Isogloboseries)	3
Ganglioside sugars	17
Globoside sugars (P antigens) (Forssman antigens)	3
Globoside sugars (P antigens) (Globo series—core structure type 4)	3

(continued)

**Table 1**  
**(continued)**

BiOligo category	Number of entries
Globoside sugars (P antigens) (P blood group antigens and analogues)	6
Globoside sugars (P antigens) (Stage-specific Embryonic antigens: SSEA-3 and SSEA-4)	4
Glucuronylated oligosaccharides	2
Glycosphingolipid	2
Lewis antigens	29
Miscellaneous	22
Miscellaneous (Blood group-related oligosaccharides)	2
Miscellaneous (Chitin oligosaccharides)	4
Miscellaneous (Fibrinogen related oligosaccharides)	3
Miscellaneous (LDN-related oligosaccharides)	6
Miscellaneous (Lewis X-related oligosaccharides)	2
Miscellaneous (TF-related oligosaccharides)	4
Miscellaneous (TN-related oligosaccharides)	4
Miscellaneous (Trehalose-like sugars)	2
N-linked oligos	18
Sialylated oligosaccharide (Type 1)	11
Sialylated oligosaccharide (Type 2)	12
Disaccharides	130
Monosaccharides	70

database, they all have been subjected to systematic conformational sampling to determine their conformational preferences, using the Shape software [21]. Several low energy conformations [1–5] are available for each entry. The details concerning the construction of the BiOligo database are given in **Note 1**.

### 3.1.2 Data Query

The database is available from Glyco3D portal or directly from <http://bioligo.cermav.cnrs.fr>. Upon reaching the search page, two buttons to query the database appear on the left hand panel: Simple search and Advanced search.

1. *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 glycan name, category and molecular weight to user to make an informed choice.

2. *Advanced Search.* Four search boxes appear each of them offering the choice between criteria to select: trivial name; type of constituent, category, 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.

Both the Simple Search and the Advanced Search options are equipped with an “auto-complete” function, which guides the user while querying the database. It comprises two parts; (1) a single field of entered text; (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.

### 3.1.3 Results

The detailed results are organized under two tabs: “Molecule Information” and “View and Download”.

1. *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 (<http://www.functionalglycomics.org/>), 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.
2. *View and Download.* This tab incorporates the best representatives of the families of the most-probable low energy conformations. The molecules are displayed under Jmol applet windows that enable basic viewing and measurement options, under the right-click options. Each of the conformation can be downloaded from this section, the coordinate files being at the PDB format.

### 3.2 *PolySac3DB: An Annotated Database of Three-Dimensional Structures of Polysaccharides*

#### 3.2.1 *Database Content*

PolySac3DB [22] is an annotated database that contains the 3D structural information of about 160 polysaccharide entries that have been collected from an extensive screening of scientific literature (for review *see* [17]). This yielded about 90 publications that supplied records of the atomic coordinates of polysaccharides (unit) structures established using various structure determination techniques (fiber X-ray and neutron diffraction, electron diffraction on single crystals, molecular modelling, and high resolution NMR spectroscopy, ...). A total of 157 polysaccharide structures have been incorporated into PolySac3DB (Table 2). The information was manually extracted and curated before incorporation into the repository. Space group information was retrieved from the publications together with atomic coordinates of the asymmetric unit contents in various formats (fractional coordinates ...). These atomic coordinates have been extracted from each publication and converted to standardized PDB format [23]. The symmetry operators of the space group were applied to generate the atomic content

**Table 2**  
**The classification of polysaccharide structures in PolySac3DB**

Polysaccharide family	Number of entries
Agarose	3
Alginates	3
Amyloses and starches	10
Bacterial polysaccharides	16
Carrageenans	3
Celluloses	10
Chitins and chitosans	4
Curdans	3
Glycosaminoglycans	14
Galactoglucans	1
Galactomannans	1
Glucomannans	1
Mannans	4
Pectins	9
Scleroglucans	1
Xylans	2
Nigeran	1
Others	4



of the unit cell and extend them to larger structures, such as simple, double and triple helices. The entries have been systematically organized using standard names into 18 categories representing polysaccharide families. The details concerning the construction of the PolySac3DB database are given in **Note 1**.

Upon accessing the entry page of the Web page from Glyco3D or directly from <http://polysac3db.cermav.cnrs.fr/>, various utilities and search engine are provided on the left panel via which the data content of the repository can be browsed and retrieved by the user.

1. *User Guide*. This page describes each search parameter and its output with detailed examples.
2. *Search*. This option gives access to background information about the entry/family in which the polysaccharides have been categorized. Detailed description of the Search option is given below.
3. *Build* (with POLYS). In the future, a Web version of the POLYS software, which is a molecular builder for polysaccharides and complex oligosaccharides [24], will be incorporated in this database. This shall empower the user to build his/her own polysaccharide three-dimensional structure, by simply specifying the composition (constituent monosaccharides) and sequence (glycosidic linkages) information.
4. *Methods*. These are informative pages about theoretical and experimental methods that are specifically employed to determine the three-dimensional structures of polysaccharides.
5. *Reference*. The reference page contains the list of all the publications from which the atomic coordinates of polysaccharide structures have been extracted to be inserted in the database.

### 3.2.2 Data Query

The polysaccharide data organized in the database can be browsed starting from the *search* page. The data can be accessed by two ways.

1. The “*Search by Name*” 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.
2. The “*Search by Family*” groups all entries in the database into 18 groups/families to clearly categorize the overall properties displayed by these polysaccharides, their occurrence in nature, and eventually their biosynthesis.

### 3.2.3 Results

The detailed results are organized under two tabs, depending upon the desired level of information, “Discover Mode” or “Expert Mode”.

1. *Discover Mode*. This mode provides well-annotated information that presents to the user an introduction to the respective polysaccharide family and components; its constituent members,

their major functions, their occurrences, the advances made in the structure determination of the members(s) of the family interspersed with appropriate illustrations and references.

2. *Expert Mode*. This is directed towards expert users who already know about the polysaccharide and the background information, and are looking to gather structural data available for it from one repository. Structure-related information includes the saccharides making up the repeat unit(s) and their glycosidic linkages, the expanded 3D representation of the repeat unit, unit cell dimensions and space group, helix type, diffraction diagram(s) (where applicable), experimental and/or simulation methods used for structure description, link to the abstract of the publication, and other relevant structure-associated information.
3. *View and download*. Up to three levels of structural information are displayed: the atomic content of the asymmetric unit, the polysaccharide chain, and the content of the unit-cell. The molecules are displayed under Jmol applet windows that enable basic viewing and measurement options, under the right-click options. Each of the conformation can be downloaded from this section, the coordinate files being at the PDB format.

### **3.3 Lectin3D: An Annotated Database of Three-Dimensional Structures of Lectins**

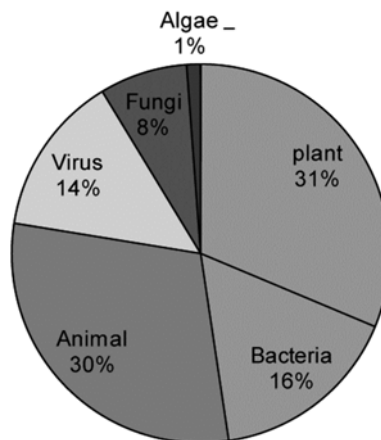
#### **3.3.1 Database Content**

Lectins are oligomeric proteins that can specifically recognize carbohydrates, which as per present knowledge act as macromolecular tools to decipher sugar-encoded messages [25–28]. Among the proteins that interact non-covalently with carbohydrates, lectins bind monosaccharides and oligosaccharides reversibly and specifically while displaying no catalytic or immunological activity. More than 1,000 lectin three-dimensional structures are available in the database. Most of them have been determined by x-ray diffraction, although some neutron diffraction structures are available as well as NMR solution structures or theoretical models. About 70 % of these structures have been determined in complex with a carbohydrate ligand, ranging from monosaccharides to oligosaccharides or glycoproteins. This consists therefore in a very large amount of information about the molecular bases of carbohydrate recognition by lectins (Fig. 3).

#### **3.3.2 Data Query**

Upon reaching the search page form Glyco3D or directly from <http://lectin3d.cermav.cnrs.fr/search.php>, two buttons to query the database appear on the left hand panel: Simple search and Advanced Search.

1. *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, (6) virus. Upon selecting one family, a right click opens a new menu that prompts the user to choose among a sub-classification based on the fold family, then on



**Fig. 3** Origin of lectin structures present in Lectin-3D database

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 going to “Lectin Information”.

2. *Advanced Search*. Under the name “Select Criteria”, a search box offers to select among the following items: (1) species, (2) family, (3) sugars, (4) PDB, (5) ligands, (6) authors, (7) sequence. 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. A more complex search can be made by combining criteria from up to four search boxes.

### 3.3.3 Results

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

1. *Lectin Information*. Under the button “Lectin 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 sugar is indicated along with its sequence. Provision is also given to view an image of the source of the protein, along with a three-dimensional ribbon-type representation, together with access to original

3D information at the Protein Database. Links to NIH sites for references and taxonomy are also provided, as well as to the glycan array data when available at the Consortium for Functional Glycomics.

2. *Display and Download.* On this page are given graphical representations of the lectin as well as of the binding site with carbohydrate ligands (for complexes) that have been constructed from the reported atomic coordinates with the help of PyMol software ([www.pymol.org](http://www.pymol.org)). A particular emphasis is given to indicate the location and conformation of the bound carbohydrate. The three-dimensional structure can be displayed under Jmol applet windows that enable basic viewing and measurements options. The atomic coordinates at the PDB can be downloaded for further use.

### **3.4 GAG3D: GlycosAminoGlycan Binding Proteins**

#### **3.4.1 Database Content**

The glycosaminoglycans (abbreviated to as GAGs) comprise a family of complex anionic polysaccharides including heparin, heparan sulfate, chondroitin sulfate, dermatan sulfate, keratan sulfate, and hyaluronic acid. In addition to their participation in the physico-chemical 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 [29–31]. Due to difficulty of obtaining crystals of complexes between protein and glycosaminoglycans, a limited number of structures are available in the Protein Data Bank. Different proteins have been co-crystallized with heparin oligosaccharides. Most of them are of animal origin with the exception of one bacterial enzyme and two viral proteins.

#### **3.4.2 Data Query**

Upon reaching the search page from Glyco-3D, two buttons to query the database appear on the left hand panel: Simple search and Advanced Search.

1. *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 sub-classification. A further right click opens a window on “GAG information”.
2. *Advanced Search.* Under the name “Select Criteria” a search box offers to select among the following items: (1) protein, (2) nature of GAG, (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 prompt to guide the user in selecting the “hits” found in the database, by a simple search engine. A more complex search can be made by

selecting criteria which can be combined 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 information provided allows the user to make an informed choice prior going to “GAG Information”.

### 3.4.3 Results

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

1. *GAG Information*. Under the button “GAG Information” is given: protein, classification, GAG type, species, PDB code, resolution, and length of oligosaccharide, comments, and reference. 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 Swissprot (<http://www.uniprot.org/uniprot>) are also provided.
2. *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 with the help of the PyMol molecular visualization system (<http://www.pymol.org>). In the case of protein–GAG crystalline complexes, a particular emphasis is given to indicate the location and conformation of the bound carbohydrate. The three-dimensional structure can be displayed under Jmol applet windows that enable basic viewing and measurements options. The atomic coordinates at the PDB can be downloaded for further use.

## 3.5 MAbs: An Annotated Database of Monoclonal Antibodies Recognizing Carbohydrate Antigens

### 3.5.1 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/> [32]. Anti-carbohydrate antibodies with specificity to oligosaccharides and polysaccharides are of a high importance in immunology and vaccine development [33]. The present database is concerned with the limited set of high resolution structures of carbohydrate–antibody complexes. 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.

### 3.5.2 Data Query

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

1. *Simple Search.* The first level offers the choice between the human or murine nature of the antibody. A further right click opens a window on “Antibody information”.
2. *Advanced Search.* Under the name “Select Criteria” a search box offers to select among the following items: (1) nomenclature, (2) antibody, (3) origin, (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 prompt to guide the user in selecting the “hits” found in the database, by a simple search engine. More complex search 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”.

### 3.5.3 Results

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

1. *Antibody Information.* Under the button: “Antibody Information” are given: origin, name of the antibody, nomenclature of the bound carbohydrate, PDB code, resolution, comment, immunoglobulin class, 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.
2. *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 with the help of the PyMol molecular visualization system ([www.pymol.org](http://www.pymol.org)). 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. The two three-dimensional structures can be displayed under Jmol applet windows that enable basic viewing and measurements options. The atomic coordinates at the PDB can be downloaded for further use.

## 3.6 GT3D: An Annotated Database of Three-Dimensional Structures of Glycosyltransferases

### 3.6.1 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. Acceptor substrates are carbohydrates, proteins, lipids, DNA, and numerous small molecules such as antibiotics, flavonol, and steroids. The majority of GTs utilizes nucleotide-sugars as donors, although lipid phosphate sugars and phosphate sugars may be used. 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. A classification system has been adopted that groups GTs into families based on amino-acid sequence similarities (CAZY database: <http://www.cazy.org>) [34]. At the time of writing, the database contained ~100,000 entries divided into over 90 GT families (designated GT<sub>x</sub>, x corresponding to the family number), the vast majority of these sequences (more than 90 %) being uncharacterized open-reading frames. To date, X-ray crystal structures are available for over 100 GTs in 38 GT families [35]. A limited number of three-dimensional architectures, noted GT-A and GT-B, have been observed for nucleotide-sugar dependent GTs [36].

### 3.6.2 Data Query

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

1. *Simple Search.* The classification of the GTs proteins is made based on their origin: (1) animal, (2) archaea, (3) bacteria, (4) plant, (5) virus, (6) yeast and fungi. Upon selecting one family, a right click opens a new menu that prompts the user to choose among a sub-classification based either on the function, or the fold (i.e., GT-A or GT-B folds). As an example, upon selection of a fold type, a further right click opens a menu where the corresponding GTs are numbered according to the CAZY classification; that allows the user to select the GT family and then the requested protein and being brought to the “GT information” page.
2. *Advanced Search.* Under the name “Select Criteria”, a search box offers to select among the following items: (1) organism, (2) family, (3) PDB, (4) authors, (5) fold, (6) resulting linkage, (7) enzyme name, (8) Abbreviation. A search box is provided in which appears a drop-down button enlisting all the entries corresponding to the selected item. The result is prompt to guide the user in selecting the “hits” found in the database, by a simple search engine. A more complex search 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 classification, the protein name, the GT type. The amount of information provided allows the user to make an informed choice prior going to “GT Information”.

### 3.6.3 Results

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

1. *GT information.* Under the button: “GT Information” are given: origin, organism, CAZY family, type of fold, enzyme name (and EC number), abbreviation, resulting linkage, mechanism

(inverting or retaining), PDB code, bound ligands, general comments about protein complexes, resolution, UniProtKB accession number, and reference. Links to PDB and Medline are also provided.

2. *Display and Download.* On this page, is represented one or more graphical representation of the glucosyltransferase, that have been constructed from the reported atomic coordinates with the help of PyMol software ([www.pymol.org](http://www.pymol.org)). The three-dimensional structure can be displayed under Jmol applet windows that enable basic viewing and measurements options. The atomic coordinates at the PDB can be downloaded for further use.

### **3.7 Conclusions and Outlook**

The present Glyco3D portal offers a single entry to access three-dimensional features of carbohydrate and carbohydrate polymers in different physical and biological conditions, as well as protein databases that interact with carbohydrates (Lectins, GAG, MAb, GTs). Whereas the databases are mainly populated by structural data arising from diffraction experiments, some of them have provisions to integrate three-dimensional models resulting from theoretical calculations. In all the databases, a common nomenclature has been adopted for the structural encoding of the carbohydrates. The next step will be the development of a unique search engine that will scan the full content of all the databases for queries related to sequential information of the carbohydrates or other related descriptors. 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.

---

## **4 Notes**

1. The databases, which have been manually curated, are Web-based, platform independent. They 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. An intermediate layer is an Apache-PHP application [Apache 2.x, PHP 5.3.1] that received the query from the use 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 the Java Script (with JQuery plugin) 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).

---

## Acknowledgments

The research leading to this publication has received funding from the European Commission's Seventh Framework Programme FP7/2007–2013 under grant agreement no. 215536 (ITN: EuroGlycoArras) and no. 213592 (ITN: CARMUSYS) 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. We acknowledge the help of Cyril Bras for testing the Web pages for the search engine. The implementation of the design of the databases benefited from the excellent contribution of Alexandre Finet and Hervé Valentin, respectively.

## References

- Cummings RD (2009) The repertoire of glycan determinants in the human glycome. *Mol Biosyst* 5:1087–1104
- Varki A, Cummings RD, Esko JD, Freeze HH, Hart GW, Etzler ME (2008) *Essentials of glycobiology*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Frank M, Schloissnig S (2010) Bioinformatics and molecular modeling in glycobiology. *Cell Mol Life Sci* 67:2749–2772
- Lutheke T (2012) The use of glycoinformatics in glycochemistry. *Beilstein J Org Chem* 8: 915–929
- Nakahara T, Nishimura S, Shirai T (2007) Current aspects of carbohydrate structural bioinformatics. *Curr Chem Biol* 1:571–578
- Perez S, Mulloy B (2005) Prospects for glycoinformatics. *Curr Opin Struct Biol* 15: 517–524
- Woods RJ, Tessier MB (2010) Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Curr Opin Struct Biol* 20:575–583
- Cooper C, Joshi H, Harrison M, Wilkins M, Packer N (2003) GlycoSuiteDB: a curated relational database of glycoprotein glycan structures and their biological sources. 2003 update. *Nucleic Acids Res* 31:511–513
- Doubet S, Bock K, Smith D, Darvill A, Albersheim P (1989) The complex carbohydrate structure database. *Trends Biochem Sci* 14:475–477
- Hashimoto K, Goto S, Kawano S, Aoki-Kinoshita K, Ueda N, Hamajima M, Kawasaki T, Kanehisa M (2006) KEGG as a glycome informatics resource. *Glycobiology* 16:63R–70R
- Herget S, Toukach PV, Ranzinger R, Hull WE, Knirel YA, von der Lieth CW (2008) Statistical analysis of the Bacterial Carbohydrate Structure Data Base (BCSDB): characteristics and diversity of bacterial carbohydrates in

- comparison with mammalian glycans. *BMC Struct Biol* 8:35
12. Lutteke T, Bohne-Lang A, Loss A, Goetz T, Frank M, von der Lieth CW (2006) GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiochemistry research. *Glycobiology* 16:71R–81R
  13. Raman R, Raguram S, Venkataraman G, Paulson JC, Sasisekharan R (2005) Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Methods* 2:817–824
  14. Ranzinger R, Herget S, von der Lieth CW, Frank M (2011) GlycomeDB – a unified database for carbohydrate structures. *Nucleic Acids Res* 39:D373–D376
  15. von der Lieth CW, Freire AA, Blank D, Campbell MP, Ceroni A, Damerell DR, Dell A, Dwek RA, Ernst B, Fogh R, Frank M, Geyer H, Geyer R, Harrison MJ, Henrick K, Herget S, Hull WE, Ionides J, Joshi HJ, Kamerling JP, Leeftang BR, Lutteke T, Lundborg M, Maass K, Merry A, Ranzinger R, Rosen J, Royle L, Rudd PM, Schloissnig S, Stenutz R, Vranken WF, Widmalm G, Haslam SM (2011) EUROCarbDB: an open-access platform for glycoinformatics. *Glycobiology* 21:493–502
  16. Imberty A, Perez S (2000) Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. *Chem Rev* 100:4567–4588
  17. Pérez S (2007) Oligosaccharide and polysaccharide conformations by diffraction methods. In: Kamerling JP (ed) *Comprehensive glycosciences: analysis of glycans*, vol 2. Elsevier, Oxford
  18. Pérez 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, Weinheim, pp 969–1001
  19. Lutteke T (2009) Analysis and validation of carbohydrate three-dimensional structures. *Acta Crystallogr D Biol Crystallogr* 65:156–168
  20. 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
  21. Rosen J, Miguet L, Perez S (2009) Shape: automatic conformation prediction of carbohydrates using a genetic algorithm. *J Cheminform* 1:16
  22. Sarkar A, Perez S (2012) PolySac3DB: an annotated data base of 3 dimensional structures of polysaccharides. *BMC Bioinform* 13:302
  23. Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol* 10:980
  24. Engelsen SB, Cros S, Mackie W, Perez S (1996) A molecular builder for carbohydrates: application to polysaccharides and complex carbohydrates. *Biopolymers* 39:417–433
  25. Ambrosi M, Cameron NR, Davis BG (2005) Lectins: tools for the molecular understanding of the glycode. *Org Biomol Chem* 3: 1593–1608
  26. Gabius HJ, Andre S, Jimenez-Barbero J, Romero A, Solis D (2011) From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem Sci* 36:298–313
  27. Imberty A, Mitchell EP, Wimmerová M (2005) Structural basis for high affinity glycan recognition by bacterial and fungal lectins. *Curr Opin Struct Biol* 15:525–534
  28. Sharon N, Lis H (2003) *Lectins*, 2nd edn. Kluwer Academic Publishers, Dordrecht
  29. Gandhi NS, Mancera RL (2008) The structure of glycosaminoglycans and their interactions with proteins. *Chem Biol Drug Des* 72: 455–482
  30. Imberty A, Lortat-Jacob H, Pérez S (2007) Structural view of glycosaminoglycan-protein interaction. *Carbohydr Res* 342:430–439
  31. Raman R, Sasisekharan V, Sasisekharan R (2005) Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chem Biol* 12:267–277
  32. Allcorn LC, Martin AC (2002) SACS – self-maintaining database of antibody crystal structure information. *Bioinformatics* 18:175–181
  33. Pazur JH (1998) Anti-carbohydrate antibodies with specificity for monosaccharide and oligosaccharide units of antigens. *Adv Carbohydr Chem Biochem* 53:201–261
  34. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37:D233–D238
  35. Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* 328:307–317
  36. Breton C, Fournel-Gigleux S, Palcic MM (2012) Recent structures, evolution and mechanisms of glycosyltransferases. *Curr Opin Struct Biol* 22:540–549